International Series in
Operations Research & Management Science

Raimund M. Kovacevic
Georg Ch. Pflug
Maria Teresa Vespucci *Editors*

Handbook of Risk Management in Energy Production and Trading

Springer

International Series in Operations Research & Management Science

Volume 199

For further volumes:
http://www.springer.com/series/6161

Raimund M. Kovacevic • Georg Ch. Pflug
Maria Teresa Vespucci
Editors

Handbook of Risk Management in Energy Production and Trading

Springer

Editors
Raimund M. Kovacevic
Department of Statistics & OR
University of Vienna
Wien, Austria

Georg Ch. Pflug
Department of Statistics & OR
University of Vienna
Wien, Austria

Maria Teresa Vespucci
Department of IT & Mathematical Methods
University of Bergamo
Dalmine, Bergamo, Italy

ISSN 0884-8289
ISBN 978-1-4614-9034-0 ISBN 978-1-4614-9035-7 (eBook)
DOI 10.1007/978-1-4614-9035-7
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2013953340

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

Energy production, delivery, and trading developed from simple supply chains with one producer, owning the entire delivery network, at the beginning of the twentieth century to the very complex structures of today. Moden energy systems consist of interconnected producers, huge transportation networks, managed by independent system operators, and exchange markets on which spot products as well as futures and other types of derivatives are traded. By complexity, such modern systems are subject to many different risks, such as technical risk in production, transportation and delivery, operational risk for the system operators as well as market risks for traders and political and other long-term risks in strategical management.

This book attempts to give an overview over these types of risk, and many of its chapters describe how modern risk management methods may be applied. All management decisions have to be made in situations, where not all relevant data are precisely known. Therefore *decision making under uncertainty* is the methodological background and many papers of this book use multistage stochastic optimization as a basic tool for analysis.

The book is divided into four parts. Part I is devoted to energy markets, in particular electricity markets. Chapter 1 by P. Gross, R. Kovacevic, and G. Pflug gives a first, nontechnical, overview of energy markets and their main properties. Both physical and financial products are discussed. R. D'Ecclesia gives an introduction to basic price models for energy commodity prices in Chap. 2. In Chap. 3, F. Paraschiv reviews modeling approaches for electricity price processes and also applies extreme value theory to the tails of the price changes.

In energy risk management it is important to keep in mind the whole production, storing, and distribution process with all related economic and physical restrictions. This makes a big difference to purely financial markets. Part II therefore deals with optimal decisions in managing energy systems. Because the resulting optimization problems are typically difficult to deal with, algorithms are an important issue. In Chap. 4, M. Densing gives a review of hydropower dispatch models and discusses two models for pumped storage plants and related case studies in deep detail. A. Philpott, A. Dallagi, and E. Gallet then discuss two classes of algorithms, heuristic decomposition and cutting planes, in Chap. 5. These algorithms are again

applied to hydroelectric generation planning. R. E. C. Gonçalves, M. Gendreau, and E. C. Finardi widen the scope and analyzes combined hydrothermal production in Chap. 6. This chapter discusses decomposition techniques for multistage stochastic programs as far as algorithms are concerned. A. Eichhorn then extends the spectrum of production sources in Chap. 7 and includes renewables as well as dedicated energy storages. In Chap. 8 by A. S. Werner, A. Pichler, K. T. Midthun, L. Hellemo, and A. Tomasgard the focus changes to stochastic investment and operational optimization models for natural gas transport systems. Convenient risk measures and a new tree structure for modeling the related stochastic processes are discussed in detail. Decision making in the operation of electricity networks is treated in Chap. 9 by A. Grothey, W. Bukhsh, K. I. M. McKinnon, and P. A. Trodden, who consider good islanding decisions in unstable network situations. The issue of natural gas transmission and distribution is again widened by J. P. Luna, C. Sagastizabal, and M. Solodov in Chap. 10. They take into account market equilibria for profit-maximizing agents both in deterministic and stochastic settings. In particular the impact of market power on equilibrium prices is analyzed. Investment in the extension of energy production systems is discussed in Chap. 11 by M. T. Vespucci et al.

While many chapters deal with renewable energy up to some extent, Chaps. 12–14 are completely devoted to this important issue. In Chap. 12, D. Wozabal, C. Graf, and D. Hirschmann give an overview of renewable energy and analyze its impact on power markets. In particular the impacts of technological development, wind and solar output, and subsidies for renewable energies are considered. In Chap. 13, A. Nordveit, K. T. Watle, and S. E. Fleten use a copula-based Monte Carlo model for hedging the risk of renewable energy sources by forward and future contracts. Finally, L. Baringo and A. J. Conejo analyze investments in highly stochastic sources of power production, e.g., wind production, in Chap. 14. A multistage stochastic complementarity model is used to determine the optimal investment capacity.

Part III is devoted to pricing: R. Kovacevic and G. Pflug describe several pricing principles and especially the pricing of electricity swing options (Chap. 15) in a unified framework. The pricing of derivatives with volume control is treated in a classical financial setup by F. Espen-Benth and M. Erikson in Chap. 16.

The final part widens the scope of risks to long-term and political risks. V. Krey and K. Riahi study energy systems under aspects of climate change in Chap. 17, while P. Burgherr, S. Hirschberg, and M. Spada deal with operational risks such as catastrophic risks, in particular risks from terrorist attacks in Chap. 18.

The editors thank all contributors for their work. Furthermore we thank F. Hillier for having accepted this book in the "The International Series in Operations Research and Management Science" and the publisher for supporting the final production steps. P. Gross and B. Analui helped us in proofreading the manuscript.

Vienna, Austria — Raimund M. Kovacevic
Vienna, Austria — Georg Ch. Pflug
Bergamo, Italy — Maria Teresa Vespucci

Contents

Part I
Energy Markets

Chapter 1
Energy Markets

Peter Gross, Raimund M. Kovacevic, and Georg Ch. Pflug

Abstract Most of the risks in energy production and trading are related to market prices. As a consequence, this first chapter provides a short introduction to energy markets. Products (or more precisely contracts) which are traded in energy markets can concern either the physical delivery of energy (physical settlement) or only the payment of the financial value of such a delivery (financial settlement). In the case of a physical settlement, the traded quantities directly influence the whole system; if the settlement is financial, trades are basically bets on prices. Motivated by this distinction, we separate this chapter into two major parts: The first part considers the physical side of markets, focusing on the physical spot markets for natural gas and electric power. The second part serves as an introduction to the financial aspects of the markets, describing derivatives on physical spot contracts.
In both sections our geographical focus will be on European markets. Due to the inhomogeneities of market designs, we will focus on stylized market characteristics rather than details. We mainly consider natural gas and electricity due to their distinctively different behavior to financial markets.

P. Gross
IK Computational Optimization, University of Vienna, Vienna, Austria
e-mail: petergross@univie.ac.at

R.M. Kovacevic
Department of Statistics and Operations Research (ISOR),
University of Vienna, Vienna, Austria
e-mail: raimund.kovacevic@univie.ac.at

G.Ch. Pflug (✉)
ISOR and IIASA, Laxenburg, Vienna, Austria
e-mail: georg.pflug@univie.ac.at

R.M. Kovacevic et al. (eds.), *Handbook of Risk Management in Energy Production and Trading*, International Series in Operations Research & Management Science 199,
DOI 10.1007/978-1-4614-9035-7_1, © Springer Science+Business Media New York 2013

1.1 Effects of Market Liberalization

Energy plays a central role in modern economies and the everyday life of people in industrialized countries. Consequently, it is no surprise that huge amounts of oil and coal, just as other storable and thus easily transportable commodities, have been traded around the globe in the twentieth century.

A somewhat more recent phenomenon is the emergence of markets for electricity and natural gas, which rely on a complex and expensive distribution network. Due to the huge initial investments in the infrastructure necessary for distribution and because of significant effects of scale in their production, for a long time, state-owned monopolists were the most efficient providers. In fact, electricity and natural gas were seen as typical cases of natural monopolies. This is still the state of affairs in many countries with not fully liberalized energy markets.

From the last decade of the twentieth century onwards, competitive markets for grid-based energy have been established all over the world with the hope of economic benefits from cheaper energy supply. Competition was usually established by unbundling the roles of network operation, production, and retailing. By opening access to the distribution infrastructure, market entry barriers were drastically reduced and opportunities for sourcing and trading were generated for various market participants, such as financial investors or large energy consumers.[1]

One particularity of the energy industry is the major role of uncertainty: Demand for natural gas, for example, is usually driven by several factors like temperature or even macroeconomic factors, which are hard to predict. As another example, the level of electricity supply depends not only on availability of transmission networks and power plants but also on wind, sunshine, and rainfall due to the importance of renewable energy production. To absorb variations in supply or demand it is vital for energy companies to secure access to short-term supply of energy.

As a consequence of matching short-term supply and demand, ideal[2] spot markets also create a source of flexibility[3] (i.e., short-term supply of energy) and establish the related prices. Instead of holding flexible generation assets or supply contracts like an insurance to cover the own worst-case needs, market participants are essentially enabled to use available flexibility to gain short-term profit in the spot market. Essentially, this amounts to projecting the risks of energy companies onto the spot markets, making it possible to quantify and trade them.

A side effect of the increase in market activities is an increase of exposure to (market price) risks for both suppliers and producers of energy. Therefore energy derivatives can be used to shape the risk profiles of energy portfolios.

Thus, with the emergence of liquid markets for energy, quantification and trading of risks become a crucial part of daily business for the energy industry, triggering the development of techniques tailored to energy risk management.

---

[1] For details on the rationale behind energy market liberalization we refer to [22].

[2] That is mature, liquid markets, allowing to execute trades of arbitrary size at quoted prices (cf. p. 18).

[3] cf. [8].

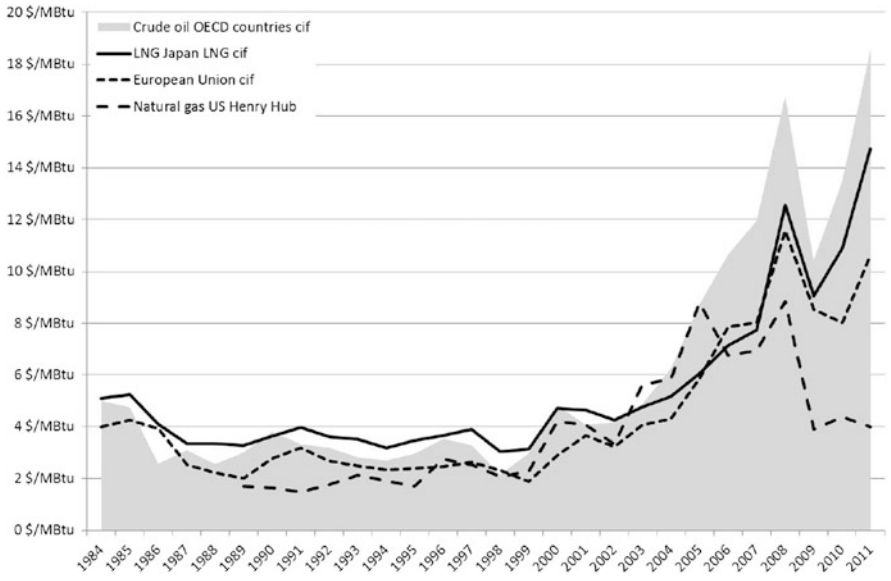

Fig. 1.1 Price levels of natural gas in different regions [source: [3], cif = cost + insurance + freight (average prices)]

1.2 The Physical Side of Markets

The main purpose of physical spot markets is the matching of supply and demand. Accordingly, markets for natural gas and electricity are principally based on physical spot products, which also serve as the main building blocks of most energy derivatives. We define spot prices as the prices for physical delivery of energy at a certain location the following day. The exact delivery period varies between products and markets. Usually electricity can be purchased for each single hour of a day or as a combination of several hours (peak/off-peak/base products) whereas natural gas spot is (mostly) delivered at a constant load during the whole following day. However, since the underlying concepts are quite similar, we will not distinguish between them.

For grid-based energies, particularly for electricity, balancing of delivery systems is crucial. Consequently, for each power market, there also exist balancing markets, where short-term generation capacities of power plants can be acquired by the system operator. We devote only a small part of our attention to balancing.

1.2.1 Natural Gas Markets

1.2.1.1 The Natural Gas Value Chain

As with oil, production of natural gas is not evenly spread over the globe. In 2011, natural gas production was dominated by the USA (20%) and the Russian Federa-

tion (18.5%), followed by Canada (4.9%), Iran (4.6%), Qatar (4.5%), and Norway (3.1%). Major exporters of natural gas are the Russian Federation, Canada, Qatar, and Norway. Similar to oil, at the moment, increasing production appears to be met with increases in confirmed reserves. According to [3] global proved gas reserves at the end of 2011 were sufficient to meet 63.6 years of production.

Before further transportation, natural gas has to be processed to remove contamination and to generate standardized products, which are tradable and fit the consumer's needs. The major share of long-range transportation is done by pipelines, which leads to a separation of the world into several markets for natural gas. Figure 1.1 illustrates the different price levels of natural gas in different regions, relying on different sources for natural gas supply.

Complementary to transport by pipeline, already in the late 1950 liquified natural gas (LNG) was used for transportation to Great Britain [8]. For this way of transport, natural gas is cooled down until it reaches a liquid state, which reduces its volume to a fraction of its original state. As a result, LNG is suitable for storage and transportation per ship. LNG plays an increasing role in connecting the different market areas for natural gas. However, the construction of the related infrastructure is expensive and time demanding and liquification as well as regasification are energy intensive. Especially due to the last point, LNG can hardly be used for arbitrage between continents, but price levels are coupled more indirectly due to the fact that LNG deliveries can be (re)directed to the market with the highest price.[4]

At a regional level, natural gas is distributed by retailers to their customers via distribution networks which are organized on a local level. Natural gas can be stored by several means, differing in size of storage and speed of input and output. In general, depleted gas fields are used for absorbing long-term fluctuations such as seasonal demand patterns, whereas salt caverns are used to flatten short-term fluctuations and to increase network stability (balancing energy).

1.2.1.2 The Development of Natural Gas Markets

In the sense of decoupling of production, transportation, and retailing, the first liberalized natural gas market appeared in the USA in the 1980s, followed by the UK in the late 1990s [11]. In pursuit of a common market for energy, several members of the European Union started implementing natural gas spot markets in the early 2000s.

Historically, natural gas markets developed very differently in various countries. Europe was reliant on the supply of a few big (stateowned) companies. This resulted in a specific European supply structure based on long-term agreements: To secure the massive investments in exploitation and transportation, producers and retailers typically negotiated long-term bilateral delivery contracts (often exceeding a time

[4] For example, the emergence of cheap USA shale gas around 2010 did not lead to significant exports of natural gas to other continents. However, market prices outside the US still dropped, to some extent caused by LNG transports stopping to deliver to the USA and serving the rest of the global market instead [1].

span of 20 years). Usually, such contracts link the price of natural gas to the price of fuels that are competitors to natural gas in the respective market region (fuel oil[5] for residential heating) or industry (oil or coal for power generation). In addition, often volume constraints are included in the contracts. These so-called take-or-pay (ToP) clauses constrain the cumulative amount to be delivered over certain time periods from below and above. This structure serves two main goals: connecting the gas price to the price of a related fuel ensures competitiveness of the price for the redistributor, while the ToP clause mitigates volume risk from the producer to the importer (see, e.g., [8, 11]). The long-term nature of the contracts amounts in a sense to vertical integration, guaranteeing a stable relation between producer and gas retailer, which simplifies long-term planning and infrastructure investments such as pipelines.

1.2.1.3 The Organization of Natural Gas Markets

As a general prerequisite, markets for trading natural gas rely on (physical) access to gas sources and potential trading partners.

The first market places for natural gas emerged at the so-called (physical) hubs, i.e., the intersections of long-distance pipelines, where gas from and to various locations could be traded between shippers. The administrator of the hub arranges the actual physical transaction and the resulting administrative actions such as transport nomination and confirmation procedures. In addition, sometimes storage (or "parking") of natural gas is offered at hubs.

This generic method of trading natural gas is based on the exact location of delivery which has one central drawback: contracts refer to more than one point of a distribution network. This may split the whole market into smaller parts with reduced liquidity. One way to bundle and increase liquidity is the introduction of virtual hubs: a virtual hub[6] is a fictive point through which *legally* all gas in a region flows. Therefore, trading for the whole region is focused at the unique virtual point. The suppression of the physical system in the market prices can lead to problems for the network operator in the case of congestion.

The actions of market participants, such as trading, delivery to customers etc. result in physical inflows and outflows at a hub. To absorb eventual imbalances in the system from imperfect matching of inflows and outflows, the transport system operator takes balancing actions by adding or extracting gas to or from the system. Usually system management relies on storage, short-term market purchase, or special flexible contracts for this task and is financially compensated by the market participants.

[5] One of the major drivers behind the large-scale usage of natural gas was to diversify energy consumption from oil in the oil crises in the 1970s, see [8].

[6] For example, NBP in Great Britain or the NCG and GP in Germany.

1.2.1.4 Stylized Facts on Natural Gas Prices

For a long time there have been strong links between natural gas prices and oil prices (cf. [19] for a model based on this). One reason is certainly the link to oil as a substitute. Another reason is the oil indexation of long-term supply contracts (discussed above). However, recently some markets experienced a decoupling of oil and gas prices (Fig. 1.1). This may be a consequence of increasing market liquidity, notably in the USA. In Europe, at the moment, it seems that the oil indexation of gas prices is losing importance along with the growing development of natural gas markets, which—provided liquidity—allows flexible sourcing at a more competitive price level but also entails more risk.

The demand for natural gas arises from electric power generation, use in energy-intensive industries (such as paper mills, cement, chemical industry), residential usage for heating and cooking and to a minor extent for transportation.

Especially large-scale usage for heating typically leads to a distinctive seasonal pattern in consumption. In addition to high price levels in winter, in some markets, high prices may occur when natural gas is used for generation at electricity peaks—for example, due to high demand in summer. This relation between natural gas demand and seasonal temperatures is reflected in the price forward curve (which is closely related to expected spot prices) (Fig. 1.2).

Many industrial users are able to switch from gas to oil. Together with the possibility of storage, this increases the elasticity of demand. As a consequence, natural gas prices in general exhibit a lower volatility and fewer spikes than electricity prices. Still, compared to stock prices, natural gas prices are quite volatile, in particular in periods of scarcity (cf. [20]). Their reliance on infrastructure is a big constraint on locational arbitrage and can result in abrupt scarcity due to physical disruptions.[7] The effects of storage are dampened by the high investment and operational cost.

In (continental) Europe, a large portion of supply still relies on long-term delivery contracts as introduced above.

1.2.1.5 Transmission Capacity Allocation in Europe

Transmission of natural gas in Europe is based on zones (largely based on member states or market areas). Transport costs between markets arise from the number of zones, which have to be crossed, and from the interzonal capacities, which have to be booked in advance. Capacities are usually classified as firm or interruptible.

In April 2013, a pan-European platform[8] was established to centralize the acquisition of capacities from transport system operators and to offer a secondary market for capacity trading between shippers. For details see http://www.prisma-capacity.eu/.

[7] For example, outages of the interconnector between GB and continental Europe or hurricanes blocking the natural gas infrastructure in the gulf of Mexico.

[8] At the moment consisting of Germany, France, the Netherlands, Belgium, Italy, Austria, and Denmark.

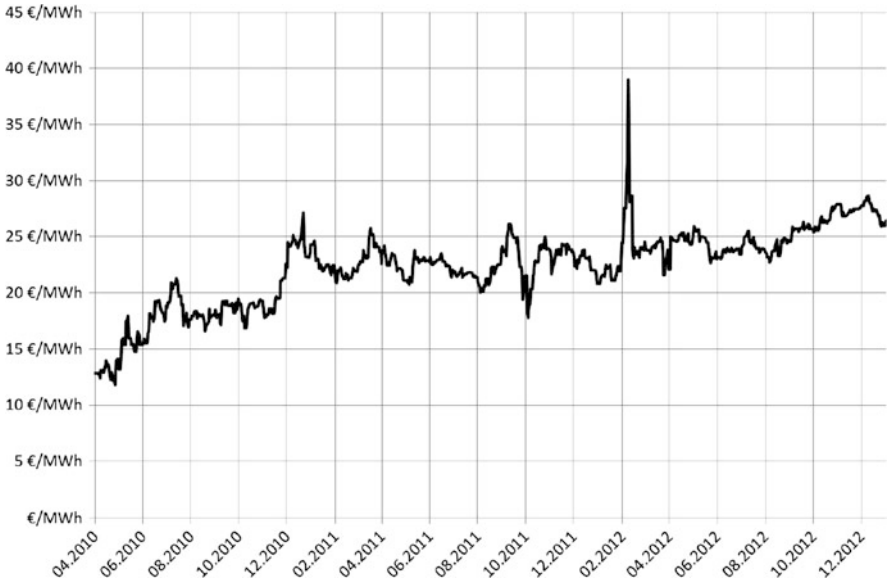

Fig. 1.2 Spot price at Zeebrugge natural gas hub (data source: http://www.net-connect-germany.de)

1.2.2 Electricity Markets

1.2.2.1 The Power Value Chain

The main market players at electricity markets are generators, distributors, and network operators. Generators convert different sources of energy into electrical energy and feed it into high-voltage networks. Distributors transform electricity from high-voltage networks into electricity with lower voltage and resell it to final consumers. Finally, network operators (independent system operators) maintain the physical network.

The main sources of power are thermal power plants (largely fossil-fueled power plants), where heat is transformed into electrical energy. The transformation can basically be achieved in two ways: The first is to generate steam which drives steam turbines. This transformation process is applicable for all fossil fuels as well as nuclear energy, waste, and some types of solar thermal plants. The second method is to directly drive turbines (basically enhanced jet engines or diesel motors) by oil or natural gas, where switching between fuels is often possible (Fig. 1.3).

The conversion factor of fuel input to electricity output is usually defined as the heat rate, which is in general a nonlinear (concave) function of the output level, i.e., electricity generation is most efficient when neither at minimal nor maximal possible output level. Often, the efficiency of thermal power plants is increased by reusing residual heat from the transformation process (cogeneration plants or combined cycle gas turbines).

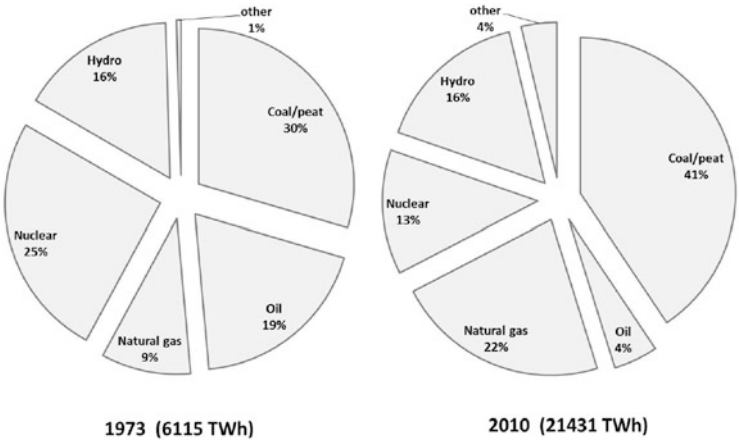

Fig. 1.3 Fuel shares in global electric power generation: 1973 (*left*) and 2010 (*right*) show an increase in relative generation by the fossil fuels natural gas and coal (data source: [15])

Depending on the layout, the electricity production of power plants is restricted by several operational characteristics (cf. [24] or [18]). The most obvious restriction on power production is the minimum and maximum production capacity of a plant, which is induced by its layout. To decrease the risk of component failures, in particular for coal and nuclear plants, it can be infeasible to turn a power plant off or on unless it has already been online or offline for a certain time period. Ramp rates specify the possible increments in power output for an online plant (due to system "inertia" or again reliability considerations). Finally, to guarantee reliability of a power plant, maintenance is necessary on a regular basis, which leads to scheduled shutdowns for all plants.

As a rule of thumb, there exists a trade-off between power production cost and system inertia, i.e., power production of plants using lignite is in general much more inflexible than of those with gas turbines. This leads to a certain degree of specialization among power plants: base load plants generate a constant power output at a low price and are sometimes turned off for maintenance only, whereas peak power plants are used to absorb high demand at high prices which compensate for their more expensive production.

Nonthermal power plants consist mainly of hydro generation, wind generation, and photovoltaic generation. In the case of hydro generation, different layouts provide very different operational characteristics. Run-of-river plants provide a constant output and are mainly used for base load generation; pondage power plants use the water from a reservoir, which provides them with a certain degree of flexibility, whereas pump storage plants can be used to store electric energy.

The special characteristic of wind and photovoltaic generation is that they are very difficult to predict due to the dependence on weather conditions (i.e., wind and sunshine) (Fig. 1.4).

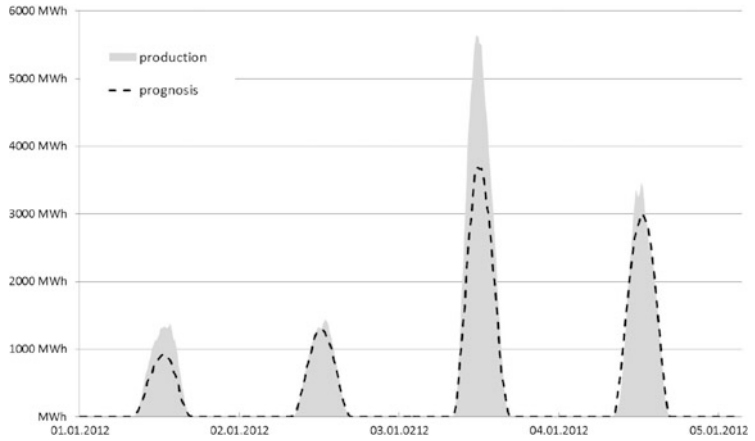

Fig. 1.4 Day-ahead forecast (*dashed*) and actual (*shaded*) production of solar power in Germany (data source: EEX)

Spot and Adjustment Markets

Short-term electricity markets typically contain a multitude of submarkets with different time horizons for delivery, reaching from day-ahead spot markets—trading power related to individual hours of the following day—to real-time markets where prices may change within minutes. In addition, power systems need reserve-capacity generators that can produce additional power within short time. Clearly, both location and time of delivery are defining elements of power products due to nonstorability and complexity of transportation.

The central goal of a spot market is the meeting of total generation and demand. Details of market design differ significantly between various market places. Following [11] (compare also [22]), we distinguish between two typical forms of power spot markets: power pools and power exchanges (bilateral markets).

- **Power pools** (single-buyer market): The distinguishing feature of this system is the central role of a system operator, who formally buys the whole generation from producers at one price.
  This can be done by collecting bids from suppliers (i.e., price for a certain generation capacity) and aggregating them from cheapest to most expensive. The result is the so-called bid stack (or merit order curve), describing the power price as a function of the demand. Intersection with the (usually inelastic) demand curve—which can be estimated or also generated by bids from buyers—gives the market price.
  In some cases,[9] the system operator collects bids from producers, containing also operational characteristics of generation units. Based on this, the system operator computes the optimal production schedule, satisfying the demand. This system

[9] Such as in the PJM market area at the east coast of the USA.

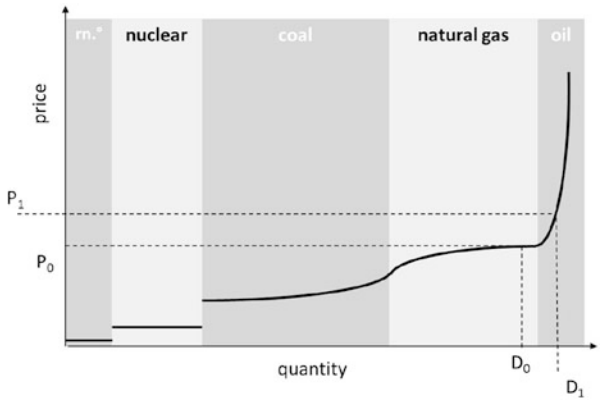

Fig. 1.5 Bid stack (schematic, rn.° = renewables). For satisfying a demand D_0 (D_1), an equilibrium price P_0 (P_1) emerges, which (simplified) corresponds to the marginal production cost of the most expensive plant producing

leads to prices which are close to the actual economic cost of production, in particular when used in connection with locational marginal pricing (as discussed below).
Figure 1.5 shows a schematic bid stack and the derivation of equilibrium prices by matching of supply and demand.

- **Power exchanges** (bilateral markets): All transactions are bilateral without connection to other trades on the markets. Since bids do not have to be publicly announced, a market clearing price (actually an exchange index) is computed based on trades (or bids/offers) and published by the exchange.

Adjustment markets work in a different way: To avoid system instability, the system operator has to maintain the target frequency in the network. Changing the real-time price of power can be used as an incentive for producers to increase or decrease their power production immediately. Reserve contingents differ with respect to the speed and likelihood of their availability: The spinning reserve encompasses units with additional capacities that are already producing, while the scheduled reserve contains units that are offline but can be brought up quickly. Generally speaking, operating reserve aims at correcting short-term disturbances and planning reserves should meet annual demand peaks. As a rule of thumb, the whole reserve capacity is held at approximately 10% of load at any time.

Besides the complex temporal distinction of different products, the location of electricity delivery has to be considered. The root of the difficulty of this issue lies at the differing cost of supply at different points in a network, which can emerge due to congestion.

To reflect the true cost of supply in market prices, the principle of locational marginal prices[10] has been introduced. For this approach, power is priced at the

[10] For details we refer to the webpage of the PJM system operator: www.pjm.com.

incremental cost of generation given the current state of the system. The benefit of being economically sensible comes at the cost of potentially unintuitive prices as well as lower liquidity due to segmentation of a market into several markets for different delivery points.

In the alternative concept of zonal prices, electricity delivery points are aggregated to zones. Transmission is only an issue between whole zones and is managed by contractual transmission rights. This approach promises higher liquidity at the cost of an imprecise transfer of actual cost (i.e., physical difficulties) of delivery.

1.2.2.2 Stylized Facts on Electricity Prices

One of the principal drivers of electricity prices are the marginal production costs, which consist largely of fuel prices and operational costs.

Some aspects of the complex interplay of electricity price, fuel prices, and demand level can be analyzed by bid stacks, as in Fig. 1.5. Since the electricity price is given by the intersection of bid stack and demand, power prices can be traced back to movements of the demand curve and movements of the bid stack.[11] If producers bid at their marginal production cost,[12] every segment of the bid stack corresponds to a power plant (with given production efficiency) and its respective production cost, which is basically driven by the fuel prices for power generation. In particular, market clearing power prices are connected to the price of the fuel which drives the generation stack at its intersection with the demand curve, the so-called marginal fuel. As a consequence, there can exist a strong dependency between electricity spot prices and the spot price of the marginal fuel. This connection can also hold for futures prices of electricity and marginal fuel.

Demand for electricity is largely inelastic due to price insensitivity of many final customers. As the demand curve can be shifted by external factors like temperature or business activity, electricity spot prices show pronounced intraday, weekly, and seasonal patterns. Note that in some markets, complex bidding behavior of producers may lead to seemingly elastic demand, for example, due to the so-called make or buy bidding: at low price levels, producers may satisfy delivery obligations by spot purchases instead of own production, leading to a price sensitive demand on the market (Fig. 1.6).

The bid stack also illustrates potential causes for price spikes and high volatility: high demand can shift the equilibrium price to regions where the offer curve is steep (due to dependency on expensive production units), leading to high, volatile prices. Finally, negative prices may be induced by a coincidence of peaking renewable production (at low price) and low demand, for example, due to holidays. This may result in a situation where base load plants face the decision of shutting down (connected to high operational costs) or paying negative prices to get rid of their production.

[11] This relation is the fundament of the so-called hybrid or structural price models which merge equilibrium and econometric models. Cf. [7] for a detailed discussion.

[12] The validity of this assumption is debatable, for example, due to strategic bidding by the producers.

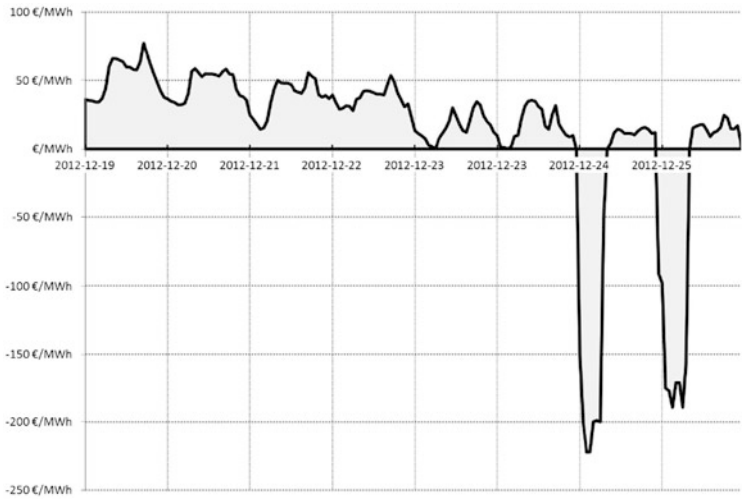

Fig. 1.6 PHELIX power spot Dec 2012 (data source: EEX)

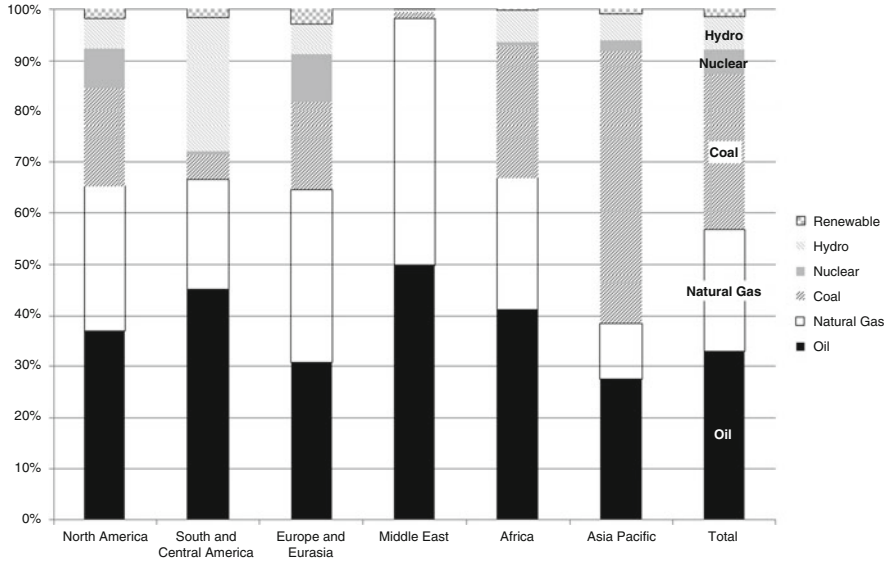

Fig. 1.7 Global consumption of primary energies (data source [3])

1.2.3 Oil and Coal Markets

Oil is a crucial commodity used for transportation, as primary energy, and in the chemical industry. As of 2011, oil had the greatest share (33.1%) in global consumption of primary energy, followed by coal with 30.3%. Whereas this marks the lowest percentage for oil in history, it is the highest share in coal since 1969 [3].

In contrast to electricity and natural gas, both oil and coal are not bound to a fixed transportation grid (although oil can be transported by pipelines as well). As a consequence there exist global markets for both commodities (Fig. 1.7).

Apparently, even the historically maximal oil consumption in recent years has been matched by updates in proved reserves. BP [3] states that world proved oil reserves at the end of 2011 were sufficient to meet 54.2 years of global production. Although the increase in proved reserves in coal does not match its production according to [3], coal has the largest reserves to production ratio of all fossil fuels.

The same ton of oil can be traded twice: once as crude oil, i.e., unrefined, and then again after refinement. The properties of crude oil vary strongly between different production locations. To achieve the grade of standardization necessary for a market, only a small number of reference qualities of oil are liquidly traded, their quotes providing a basis for trading other products. Since global transportation of oil is largely achieved by ship, delivery is usually specified at port locations. Freight rates as well as refinement capacities can have a strong influence on the prices of oil and of some refined oil products. Similar to oil, long-distance transportation of coal is usually seaborne, which makes freight rates a very important component of coal prices.[13]

In general, the oil price has the property of leading the primary energy price, i.e., the price levels of other primary energies are strongly coupled to the oil price. This is due to the frequent coupling of energy delivery contracts to oil indices as well as oil being a major competitor to other primary energies, especially for heating and power production.

In contrast to the liquid markets for some oil products, most coal trading is done via bilateral OTC contracts, due to the difficulty of standardizing the quality of coal [4].

For more detailed accounts of the oil and coal markets we refer to [4, 11].

1.3 The Financial Side of Markets: Derivatives

A derivative is a contract which can be defined as "an instrument whose price depends on, or is derived from, the price of another asset" (cf. [14]).

Although actual physical demand and physical supply meet at spot markets, huge volumes of energy are traded as derivatives on physical spot products. The main reasons for the high amount of "paper energy" in circulation are hedging and speculation.

Hedging can be defined as the execution of trades to reduce exposure to price risks. Risk reduction can require both sales and purchases of derivatives: Whereas power producers may sell parts of their production in advance, large consumers may buy parts of their demand also in advance—both to reduce their exposure to future price variability.

[13] According to [4], freight rates sometimes amount to 70% of the coal price.

Speculation is exactly the reverse of hedging: it consists of the execution of trades to deliberately gain exposure to price risks.

Without knowing the actual portfolios of market participants, it is impossible to distinguish between hedging and speculation. However, the increasing liquidity in energy markets results at least partly from increased activity of financial institutions. This development is sometimes referred to as the "financialization of commodity markets."

In the following section, we will first discuss futures contracts as the most liquid derivatives and *en passant* notice the features of exchange-traded contracts. Later we investigate the link between spot and futures prices and give a short overview of bilateral trading and nonstandard contracts.

1.3.1 Futures and Exchange Trading

Futures contracts are exchange-traded standardized contracts which specify the price of a sequence of spot deliveries, usually stretching over weeks, months, quarters, or years. For electricity there is usually a distinction between peak and off-peak (or base) futures, specifying delivery over hours of high and low demand.

The settlement of a contract specifies the mode of delivery, which can be physical (i.e., actual delivery) or financial (i.e., payment of the difference of the futures price and the spot price).

Benth et al. [2] observe that, from a financial engineering point of view, energy futures are actually swaps where a fixed payment (futures price) is swapped for a floating payment (spot price). Note that the futures price denotes the agreed delivery price per unit, but the actual volume of the contract is given by the product definition.[14]

Since futures are traded on exchanges, the exchange serves as counterparty for both buyer and seller for every trade. Since neither buyer nor seller know who is standing at the other side of the trade, exchange prices have the benefit of being nondiscriminating.

As a consequence of the intermediary position of the exchange, bankruptcy of a contract party only affects the exchange but not the participants in the trade. To avoid the risk emanating from nonpayment of a contract party (counterparty risk), the exchange imposes a margining system. This means that at every trading day, the value changes of all positions ("open interest") are financially netted among contract holders, and the contracts are replaced by current contracts with the actual market value. An initial deposit[15] from every market participant is used to cover the variation during a single day in case of insolvency of a contract holder. As a consequence of the margining systems, exchange-traded contracts are practically free of counterparty risk.

[14] Usually the contract size times length of the delivery period.

[15] Initial margin.

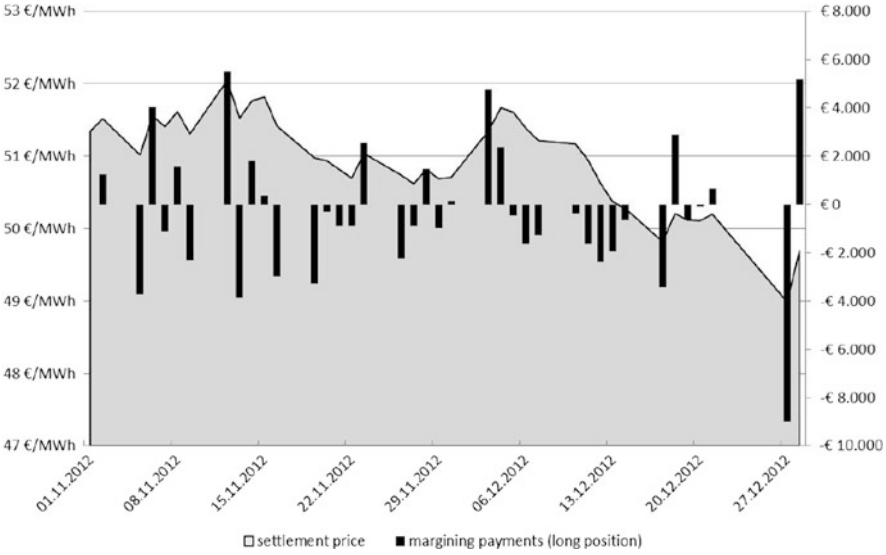

Fig. 1.8 Margining Phelix base futures Feb 13 contract (source: EEX)

The following example illustrates the procedure:

Example 1. [16] On November 1, 2012, a market participant decides to buy a constant load of 10 MW of power, deliverable in the month of January 2013. The corresponding futures contract $F(1,T)$ is bought at the exchange for the current price of 50.85 EUR/MWh. On November 2, the market price of power futures contracts with delivery period January 2013 rises, $F(2,T)$ being now 51.02 EUR/MWh. At the end of the day, the contract $F(1,T)$ held by the market participant is replaced by the new contract $F(2,T)$ (specified at a delivery price of 51.02 EUR/MWh corresponding to the current market view), and the value change is compensated by a financial payment of 1264.8 EUR,[17] which amounts to the difference in value of the two contracts. Figure 1.8 shows the whole sequence of margining payments, from November 1 until delivery at January 1.

One consequence of margining is that the values of all open positions have to be determined—even if there happened to be no active trades. Thus, a central role of an exchange is the determination of settlement prices, i.e., the prices at which open positions are valued at the end of a trading day. Weighted means of trades are often used for determining settlement prices; in the case of too few trades, the prices are to be estimated.

[16] We assume no interest rate and ignore initial margin for the sake of simplicity.

[17] $= 10\,\text{MW} \times (24 \times 31)\text{h} \times 0.17\,\text{d'/MWh}$.

1.3.1.1 Usage of Futures

Futures are to a large extent used for hedging purposes. Examples include the fixing of profit of a fossil-fueled power plant by buying fuel futures and selling power futures for the same delivery period to lock in the margin (cf. [9]). See [23] for a detailed account on hedging a natural gas portfolio with futures products. Note that even perfect hedging of risks (i.e., the exact offsetting of risky positions) can lead to losses due to margining payments.

Besides actual trading, futures provide information about the market assessment of future prices. This information can be of tremendous value and is widely used as "objective" evaluation of contracts and whole portfolios, a practice known as mark to market. Essentially this is done by valuing future obligations or deliveries by the quotes of the corresponding futures products on the market. This makes it possible to "value any contract without need to rely on the view of a trader" [11]. In practice, some effort is put into the transformation of market quotes into forward looking prices with daily or even hourly time resolution. These so-called price forward curves are often used for the valuation of daily or hourly load patterns (cf. [10, 21]).

It is especially the transparency and nondiscriminativity of futures prices—and the absence of counterparty risk premia—that recommend them as benchmarks. Note however that the concept of mark to market is theoretic to some extent: it implicitly is assumed that a whole portfolio can be sold at market prices—no matter what the size. This assumption may turn out not to be valid when a portfolio actually has to be liquidated.[18]

Several properties of a market which are prerequisites for using market prices as reference for valuation and risk management are subsummized in the term of market liquidity.

1.3.1.2 Market Liquidity

Generally spoken, market liquidity describes the possibility to trade arbitrary volumes in a market at quoted prices. Following [8], liquidity can be defined more precisely by splitting it into four distinct properties:

- *Depth*: large volumes can be bought or sold without moving the price excessively.
- *Breadth*: large number of different bids and offers are present in the market.
- *Immediacy*: the possibility to trade large volumes in a short period of time.
- *Resilience*: the ability of the market to recover towards its natural supply/demand equilibrium after having been exposed to a shock.

Note that liquidity is rather a property of single products than of markets: in many markets, (monthly) futures close to their maturity are more liquid than futures maturing far in the future.

[18] A famous example is the case of the hedge fund Amaranth Advisors L.L.C., which collapsed, unable to liquidate its huge portfolio of natural gas futures (cf. [5]).

One means by which exchanges try to establish liquidity is to establish market makers. Market makers provide liquidity by buying or respectively selling a certain quantity of energy if it is quoted outside a prespecified price range, which reduces the risk of finding no counterparty.

1.3.2 Pricing of Futures

Arbitrage-free pricing methods for futures contracts exist, both for financial and commodity markets, and are widely used. They are based on the idea of constructing a portfolio that replicates the cash flows of a derivative by borrowing money now and buying the security or commodity under consideration in order to meet the final demand ("cash and carry strategy"; cf. [14]). The proper strategy may involve rebalancing the amount of commodity held over time and is subject to interest payments, e.g., transport costs and storage costs.

Denote the forward price at delivery time T, agreed at time 0—that is today—by F_T and the current spot price with S_0. Subsuming all related costs under the terminus *cost of carry*, absence of arbitrage leads to the following relation:

$$F_T \leq S_0 + cost\ of\ carry \tag{1.1}$$

Assume inequality (1.1) is strictly satisfied. Then, in classical finance, the "reverse cash and carry" strategy could be applied by short selling spot and closing the position at time T via the futures contract. However, for commodities, this arbitrage strategy may not be viable: In the first place, short selling is not possible without having physical energy in stock. But besides this, the holder of the commodity may not even be willing to reduce the storage level for fear of not being able to satisfy the own demand. This line of thought is the idea behind the concept of convenience yield (introduced by [16, 25]). This quantity is interpreted as the financial value of having a commodity in stock and can be formally defined as the slack variable for inequality (1.1):

$$F_T = S_0 + cost\ of\ carry - convenience\ yield. \tag{1.2}$$

In situations of extreme scarcity, the convenience yield outweighs the cost of carry, thus forcing the spot above futures prices, whereas in periods of abundance cost of carry dominates the relation.

This approach suggests that the key drivers of the forward prices of storable commodities are the spot price and scarcity (given for example by inventory, demand, or production level).

Using stochastic convenience yield, Eq. (1.2) has been used as starting point for the (widely used) two-factor spot price models [13]. Empirical results in [20] also suggest connections between scarcity and volatility.

The derivation of the convenience yield relied on storage, which may be questionable for natural gas since storages have limited capacity and turnover rates.

Unfortunately, given the nonstorability of electric energy, the usual no-arbitrage arguments fully break down for power markets. A direct relation between the actual spot price and forward price is not observable for electricity. This means that forward prices contain additional information about the future and are related to expected prospective spot prices. Stoft [22] proposes to use the simple relation

$$F_T = \mathbb{E}[S_T]. \tag{1.3}$$

While admitting that this relation is not exact, Stoft consists other effects as " too subtle and too unpredictable to be of interest." Others have made the effort to extend relation (1.3), which leads to

$$F_T = \mathbb{E}[S_T] + \textit{Risk premium}(T). \tag{1.4}$$

The size (and sign) of the risk premium can be explained by the hedging pressure on producers of consumers: a positive (negative) risk premium is read as insurance premium paid by consumers (producers) to avoid price risk. Based on real data [11] shows that for some electricity markets the risk premium is positive if T is small, particularly if it corresponds to a winter or summer month, and may be negative if T is large, i.e., several years. Similar results are discussed in [12].

Equation (1.4) is the starting point for introducing risk neutral probabilities[19] Q (or equivalent martingale measures), loosely speaking by incorporating the risk premium into the probability distribution:

$$F_T = \mathbb{E}_Q[S_T]$$

Note that in the context of Eq. (1.2), choosing Q corresponds to specifying the convenience yield [2].

Other models for the spot price and/or the forward price structure—in fact the whole arsenals of econometrics and finance—have been used as well. Because of long-term equilibria of demand and supply of energy, models with mean reversion are usually preferred. See [9] for a broad overview and [2] for a rigorous modeling approach.

Summarizing, we can state that short- and long-term prices are almost entirely disconnected as soon as storability is not granted. In particular, information about future events such as planned outages of power plants or a change in the market structure affect only the prices of futures with suitable delivery period (cf. [17] for an empirical study of the effect of information on risk premia for electricity futures).

[19] Not necessarily unique; see discussion at the end of Sect. 1.3.3.

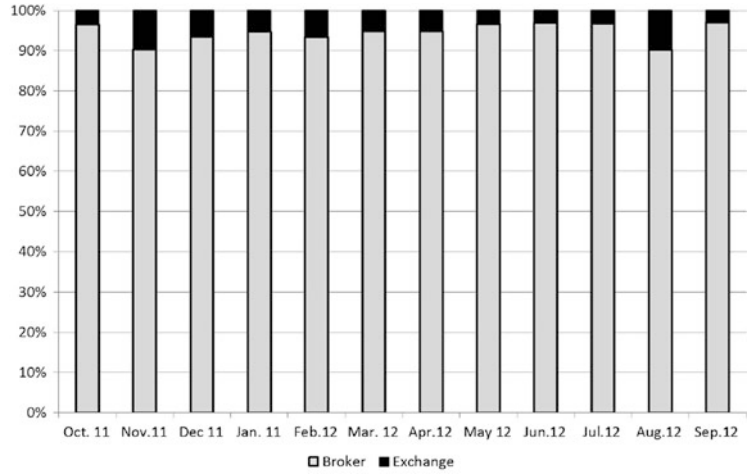

Fig. 1.9 Relative volumes traded at the German virtual natural gas hub NCG (data source: http://www.net-connect-germany.de)

1.3.3 Bilateral Trading in Energy Markets

As an alternative to trading at an exchange (mediated), there is the possibility to trade directly (bilaterally) with counterparties, the latter being called OTC (over-the-counter) trading. Compared to mediated markets bilateral markets are much more flexible with respect to the exact formulation of contracts. This is probably the reason why the major portion of energy trading still occurs bilaterally, an extreme example being the German NCG market for natural gas (see Fig. 1.9). However, writing those contracts is more involved. The main complexity arises from the tailored contractual trading agreements, which have to be negotiated. Usually bilateral trading occurs between parties who have already established contractual frameworks for trading, which are extended by standardized annexes[20] for specific trades.

In general, bilateral trades can be organized in two ways: by broker platforms (such as ICE or Spectron) that "match" the bids of participants or by direct communication between companies. As a result, counterparty risk exists but depends on the specific contractual details.

Bilateral trading occurs—at least theoretically—at individual prices, which are not published and are not necessarily connected to any other market prices. However, the data published by broker platforms show a close connection between OTC prices and exchange prices.

[20] For details see European Federation of Energy Traders, http://www.efet.org/ or International Swaps and Derivatives Association, http://www2.isda.org/.

Table 1.1 Organization forms of energy trading (see [8])

	Bilateral with broker	**Exchange**	**Bilateral without broker**
Contracts	Agreement between companies	Agreement with exchange	Agreement between companies
Trading method	Through broker	Electronic platform	Personal contact
Counterparty	Other company	Exchange	Other company
Transaction costs	Medium	High	Low
Transparency	Good, publications on end of day prices	High, information given by exchange	None
Anonymity	Identity revealed after deal	Anonymous	Identity revealed before deal
Main usage	All products	Most liquid products	Illiquid products / large volumes
Type of agreement	Framework contract	Agreement with exchange	Bilateral contract

1.3.3.1 A Survey of Bilateral Contracts

The flexibility of bilateral contracts allows potentially arbitrary conditions. However, some contract specifications proved to be especially suitable to mitigate risks in the energy industry, leading to some level of standardization even for OTC products (Table 1.1).

Probably the most common OTC products are *forward contracts* which are in fact identical to futures contracts but are traded bilaterally. The difference to futures contracts of equivalent delivery period results from the absence of margining and the presence of counterparty risk. For deterministic interest rates and absence of counterparty risk, forward and futures prices with the same maturity are equal ([11], p. 44). Consequently, futures and forwards are often used synonymously in literature. Note however the different behavior of forwards and futures in hedging.

As in financial markets, *European call* and *put options* are traded OTC, mostly on futures. European options are also traded at some exchanges, such as NYMEX, EEX, and Nordpool.

As a particularity of energy markets, *swing options* (also often called ToP contracts for natural gas) are used by market participants to hedge price risks and volume risks. Swing options have a long history as hedging instruments against those risks. At every day t of the runtime $[0,T]$ of the swing option, the holder is entitled to obtain an amount y_t of energy for a strike price K, which has been specified in advance. Usually, constraints are put on the consumption of the holder q_t, such as

$$\begin{aligned} \underline{e}_t \leq \quad y_t \quad &\leq \bar{e}_t \\ \underline{E} \leq \textstyle\sum_{t=0}^{T} y_t &\leq \bar{E}, \end{aligned}$$

bounding the single-period consumption by $\underline{e}_t$ and $\bar{e}_t$ and cumulative consumption by $\underline{E}$ and $\bar{E}$.
The strike price K can be dependent on price indices (such as oil prices for natural gas delivery contracts) but usually is independent of the current spot price of the underlying good.
Note that as a consequence of the constraints on cumulative consumption, the exercise strategy of a holder at time t depends on the whole history $y_0, \ldots, y_{t-1}$ of previous decisions since they determine the remaining quantity to be taken from the contract in the future. This makes the valuation of swing options much more complicated than the valuation of European options.

Finally, in the context of commodity markets, a *spread* refers to the difference between the prices of two products. The difference can be expressed either in terms of physical, temporal, or spatial properties. One example is the so-called *spark spread*, denoting the difference between the price of electricity and the price of the quantity of fuel required for generation. The spark spread or options on it can be used to hedge fossil-fueled power plants.

For further details on energy derivatives we refer to [4, 6, 9, 11].

References

1. Auer J (2010) Gas glut reaches Europe. Technical report, Deutsche Bank Research
2. Benth FE, Benth JS, Koekebakker S (2008) Stochastic modelling of electricity and related markets. World Scientific, London
3. BP (2012) BP statistical review of world energy June 2012. Technical report, BP. bp.com/statisticalreview
4. Burger M, Graeber B, Schindlmayr G (2007) Managing energy risk. Wiley, Chichester
5. Chincarini L (2008) A case study on risk management: lessons from the collapse of amaranth advisors llc. J Appl Financ 18:152–174
6. Clewlow L, Strickland C (2000) Energy derivatives: pricing and risk management. Lacima Publications, London
7. Coulon M (2009) Modelling price dynamics through fundamental relationships in electricity and other energy markets. PhD thesis, University of Oxford
8. Cronshaw I, Mastrand J, Pirovska M, Simmons D, Wempe J (2008) Development of competitive gas trading in continental europe. Technical report, International Energy Agency
9. Eydeland A, Wolyniec K (2003) Energy and power risk management. Wiley, New Jersey
10. Fleten S-E, Lemming J (2003) Constructing forward price curves in electricity markets. Energ Econ 25:409–424
11. Geman H (2005) Commodities and commodity derivatives, modeling and pricing for agriculturals, metals and energy. Wiley, Chichester

12. Giacometti R, Vespucci MT, Bertocchi M, Adesi GB (2010) A stochastic model for hedging electricity portfolio for an hydro-energy producer. Archivio istitutionale Università degli studi di Bergamo. http://hdl.handle.net/10446/465
13. Gibson R, Schwartz ES (1990) Stochastic convenience yield and the pricing of oil contingent claims. J Financ 45:959–976
14. Hull J (2006) Options, futures and other derivatives. Pearson Prentice Hall, New Jersey
15. IEA (2012) Key world energy statistics. http://www.iea.org
16. Kaldor N (1939) Speculation and economic stability. Rev Econ Stud 7:1–27
17. Kiesel R, Benth FE, Biegler-König R (2013) An empirical study of the information premium on electricity markets. Energ Econ 36:55–77
18. Konstantin P (2007) Praxisbuch der Energiewirtschaft. Springer, Berlin
19. Ohana S (2010) Modeling global and local dependence in a pair of commodity forward curves with an application to the us natural gas and heating oil markets. Energ Econ 32:373–388
20. Ohana S, Geman H (2009) Forward curves, scarcity and price volatility in oil and natural gas markets. Energ Econ 31:576–585
21. Ollmar F (2003) An analysis of derivative prices in the nordic power market. PhD thesis, Norwegian School of Economics and Business Administration
22. Stoft S (2002) Power System economics - designing markets for electricity. Wiley-Interscience, New Jersey
23. Treeck T (2009) The hedge effectiveness of european natural gas futures. PhD thesis, University of St. Gallen
24. Weber C (2005) Uncertainty in the electric power industry. Springer, New York
25. Working H (1949) The theory of the price of storage. Am Econ Rev 39:1254–1262

Chapter 2
Introduction to Price Models for Energy

Rita L. D'Ecclesia

Abstract The goal of this chapter is to present models which describe the dynamics of energy commodity spot prices and their forward curves. Recent developments in energy markets together with the use of new technologies caused changes in the dynamics of spot prices and there is a growing need to understand it. Given the spot price of an exchange-traded commodity we assume the forward curve, with a large set of liquid maturities is available. The forward curve provides information about the market perception of future spot prices and can be easily used to describe energy price behavior. In this chapter specific models which show to be suitable to capture the properties of energy prices are described.

2.1 Key Features for a Price Model

Quantitative methods are important requirements in energy risk management. A responsible institution will need to assess its level of various risks and monitor its level of price risk. Energy price risk management often involves a number of stages:

1. Analysis of energy price risk modelling.
2. Development of budget forecasts and potential exposures.
3. Identification of risk mitigation options. Typically this includes hedging exposures by transacting forwards, futures, or options in order to reduce budget variance and increase predictability.

In this chapter we focus primarily on the first stage: energy price risk modeling introducing most of the important statistical concepts and procedures which allow to assess and manage energy price risk. The energy marketplace is increasingly complex and dynamic. Unleaded gasoline, electricity, natural gas, heating oil,

R.L. D'Ecclesia (✉)
Sapienza University of Rome, Rome, Italy
e-mail: rita.decclesia@uniroma1.it

R.M. Kovacevic et al. (eds.), *Handbook of Risk Management in Energy Production and Trading*, International Series in Operations Research & Management Science 199,
DOI 10.1007/978-1-4614-9035-7_2,

crude oil, and emissions are exchange-traded commodities which are subject to frequent price swings on a short- and long-term basis. The complex dynamics need to be accurately modeled. Extreme levels of price volatility increase energy price risk generating, for various companies and institutions, other types of risk. For example, Californian electricity price spikes occurred in the summer of 1998 caused a dramatic increase in energy price risk. The price spike was so large that a number of energy players found themselves exposed as a result of their writing (selling) uncovered options. Therefore the possibility of a default on outstanding obligations (i.e., credit risk) increased for their counterparties. Power plant outages are another example of risk. The occurrence of an outage in a power plant increases the likelihood of outages in other power plants (thus increasing operational risk) and leads to price spikes (thus increasing energy price risk), which in turn can lead to some energy players being exposed because of their derivative liabilities (and thus credit risk is increased). Energy prices show particular features which need to be taken into account when selecting the pricing model. Natural gas exhibits seasonality and discontinuous changes; electricity prices exhibit mean reversion features and sudden, unexpected, discontinuous changes. Crude oil seems to have lost its well-known feature of mean reversion in the last decade.

2.1.1 Statistical Features of Commodity Prices

Commodities are an under appreciated and often misunderstood asset class. The drivers of commodity prices, particularly at an index level, are complex and sometimes harder to grasp than fixed income or equity assets. The means of accessing this asset class are not always intuitive either. But from an asset allocation perspective commodities have an interesting set of risk-return and correlation characteristics.

Commodities have recently become a new asset class and are more and more used by hedge funds, banks, and investors in addition to the usual producers and retailers. During the past 20 years, commodities and equities have shown different performance patterns. Commodities have outperformed global equities total returns since 1985. The reason is that the demand in energy markets has boosted commodity returns, while equity markets have gone through some major market corrections. Energy commodity prices do not generally exhibit trends over long periods; we may want to leave aside in this statement the particular situation of oil at the present time. If we look at the various energy markets as the UK natural gas, the European Exchange (EEX) electricity hourly contracts, or the Brent crude oil, different behaviors can be identified. In Fig. 2.1 the National Balancing Point (NBP) futures contract daily prices are reported, Fig. 2.2 reports the electricity base-load contract traded at EEX, and the Brent futures daily prices are described in Fig. 2.3.

Even if sharp rises are observed during short periods for specific events, such as the weather or political conditions in producing countries, energy prices tend to revert to normal levels over a long period. If demand is constant or slightly increasing over time, as it has been the case for natural gas in the European countries until

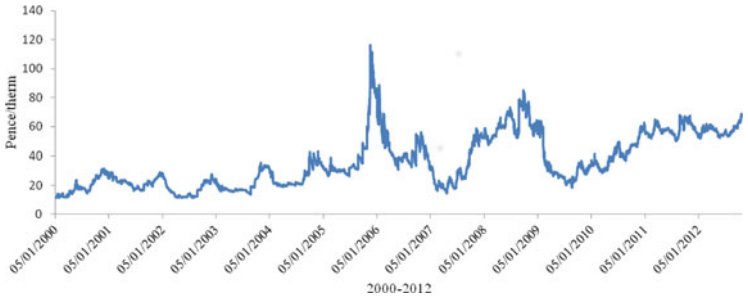

Fig. 2.1 NBP futures contract daily prices, 2000–2012

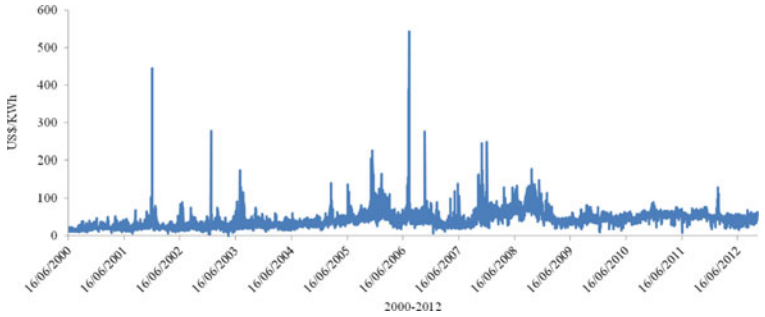

Fig. 2.2 EEX base-load contract daily prices, 2000–2012

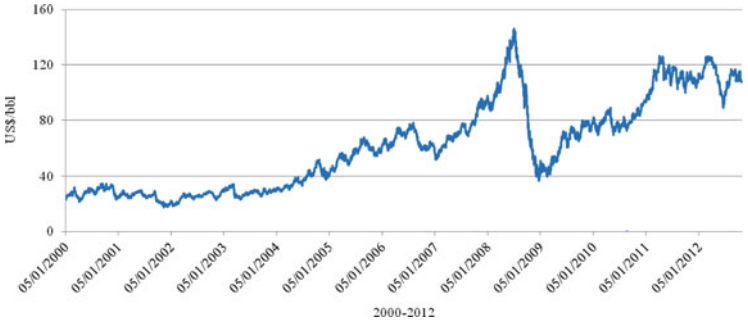

Fig. 2.3 Brent futures daily prices, 2000–2012

2005 (see Fig. 2.4) and if supply adjusts to this pattern, prices should present a stable pattern on average. This means that energy prices exhibit a mean reversion feature. In addition the occurrence of supply shocks caused by sudden unexpected change of the weather conditions or outages generates spikes. An example of this pattern is provided in Fig. 2.2.

These different features of energy prices are shown by their statistical features. Table 2.1 provides the first four moments computed on an annualized basis of daily returns for various energy commodities, where, as in finance, the return at date t is defined by:

Table 2.1 Statistical features of energy prices

Statistics	Natural gas (GBP/therm)	Electricity (USD/MWh)	Brent (USD/bbl)	Coal (USD/st)
Min	8.53	0.8	9.64	5.63
Max	116.3	543.72	146.08	19.72
Mean	32.87	46.97	48.66	8.25
Variance	385.65	761.79	1027.32	8.09
Kurtosis	−0.47	48.344	−0.35	3.06
Skewness	0.68	4.28	0.845	1.94

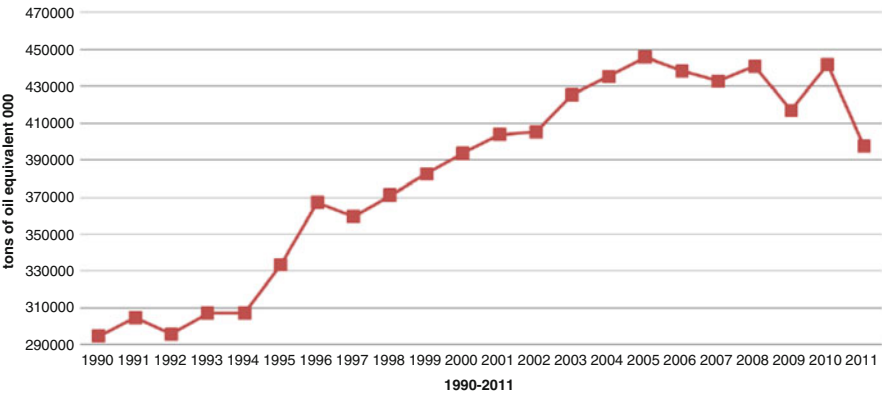

Fig. 2.4 Natural gas consumption in the EU27, 1990–2011

$$R_t = ln(\frac{S_t}{S_{t-1}}) \sim \frac{S_t - S_{t-1}}{S_t} \tag{2.1}$$

and S_t denotes either the spot price of the commodity at date t or the price of the first nearby futures contract. Given a time series $\{R_1, R_2, \ldots, R_T\}$ of length T, the mean $E(R)$ is computed as

$$E(R) = \frac{1}{T}\sum_{i=1}^{T} R_i \tag{2.2}$$

and the variance as

$$\sigma_T^2 = \frac{1}{T-1}\sum_{i=1}^{T} (R_i - E(R))^2. \tag{2.3}$$

The square root of σ_T^2 is the volatility of the rate of return. The length T of the time interval is computed as a fraction of the year. So the volatility σ_T must be converted into its annual equivalent through the formula

$$\sigma = \sigma_T \sqrt{\frac{250}{T}} \tag{2.4}$$

if the number of trading days is 250. In the case of electricity, however, we need to consider 365 days given that it is traded every day of the year.

Commodity prices show different dynamics compared to the stock prices. In general the volatility of commodities daily returns is not comparable to the equity or interest rate markets, for the latter annual volatility ranges between 10% and 12%, while for commodities like natural gas annual volatility is about of 68% and in the electricity market a it may reach 500%. In the case of the EEX electricity contract over the period 2000–2012 an annual volatility equal to 592% is reported. In Fig. 2.5 the dynamics of the daily rate of returns of the NBP natural gas contract is reported.

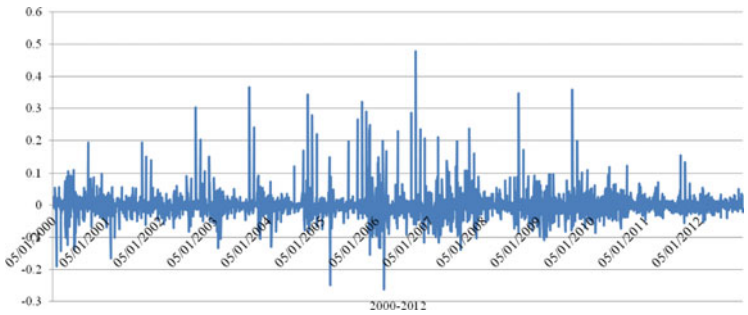

Fig. 2.5 NBP daily returns, 2000–2012

2.2 The Basic Stochastic Process

We consider the case of a single commodity spot price which is represented by a stochastic process S_t. If not otherwise specified, the current date is denoted as 0. Our concern is twofold:

1. To find the most appropriate mathematical structure for S_t, i.e., the type of process: Geometric Brownian Motion (GBM) versus jump diffusion or any other structure, using the properties observed in the historical database of spot prices with possible adjustments for economic growth or technological change. The choice of this stochastic process S_t should lead to a probability distribution for the random variable S that agrees in particular with the empirical moments and other known features of this distribution. The statistical features have to be consistent with the observed dynamics, i.e., the properties involved in the change of S between two dates.
2. Once we have chosen the "backbone" of S_t there will be parameters attached to this stochastic process. These parameters will be estimated from market data. Hence, we need liquid markets and "clean" data to make the right decision.

The empirical characteristic of sudden jumps t_i and t_{i+1} in many energy products has been invoked to explain the discrepancies observed between actual pricing of energy exotic options and theoretical predictions. More generally, ignoring heavy

tails results in extreme price changes being underestimated and could lead to ineffective hedging strategies and/or to mispricing assets. A sensible model for energy products should capture the consequence of temporary supply shocks which result in price jumps. Merton [13] introduced a model which captures this idea. It describes the evolution of the asset price by a continuous diffusion part and a discontinuous jump part.

2.2.1 Arithmetic Brownian Motion

A process X is called a Brownian motion with drift (or an Arithmetic Brownian Motion or the Bachelier model [1]) if it satisfies the stochastic differential equation:

$$dX_t = \mu dt + \sigma dW_t \tag{2.5}$$

where:

- μ and σ are real numbers, σ being strictly positive.
- dX_t represents the change in X over an infinitesimal time interval dt (dt=1 day) or
$$dX_t = X(t+dt) - X(t). \tag{2.6}$$
- dW_t represents the differential of Brownian motion W_t. For each $t > 0$ the random variable $W_t = W_t - W_0$ is the increment in $[0;t]$ it is normally distributed with zero mean, $E(W_t) = 0$, standard deviation $\sqrt{E(W_t^2)} = \sqrt{t}$, and density
$$f(t,x) = \frac{1}{\sqrt{2\pi t}} e^{\frac{-x^2}{2t}}. \tag{2.7}$$
For $p \in [0,1]$ the p-th percentile of W_t is $\sqrt{t}N^{-1}(p)$, where N^{-1} is the inverse function of the standard normal distribution function
$$N(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-u^2}{2}} du. \tag{2.8}$$

An example of the density function of a Wiener process is described by Fig. 2.6. Equation (2.6) implies that

- The expectation of the increment is $E(dX_t) = \mu dt$. So $\mu = \frac{E(dX_t)}{dt}$ represents the expected change in X per unit of time and is called the drift of the arithmetic Brownian motion.
- The variance of the increment is
$$Var(dX_t) = \sigma^2 Var(dW_t) = \sigma^2 dt. \tag{2.9}$$
Hence, dispersion of the change in X around its mean, μ, increases with σ, the fundamental volatility parameter. The price changes increase with the length of the time interval dt.

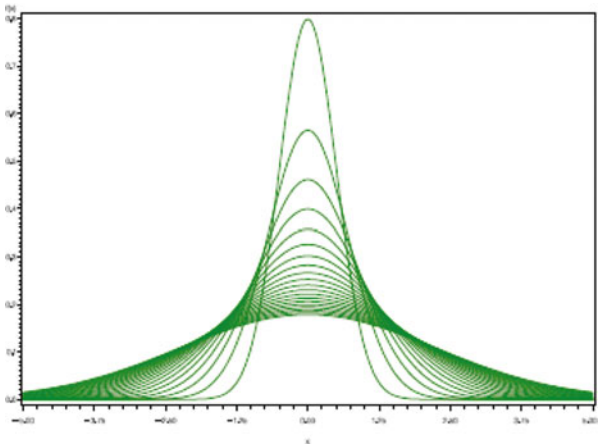

Fig. 2.6 Wiener process, density function for $t = 0.25, 0.5, \ldots$

- The price changes dX_t are
  - Independent, i.e., the change between Monday and Tuesday has no impact on the change from Thursday to Friday.
  - Stationary (the change in X over one day has the same distribution over time).

  The distribution of the price change dX_t depends on the length of the interval $(t + dt)$ but not on its origin X_t.
- If the drift is zero, i.e., $\mu = 0$, then

$$E(dX_t) = \sigma E(dW_t) = 0. \tag{2.10}$$

  The process (X_t) can be seen as a continuous version of a symmetric random walk. Its expectation equals the starting point X_0 for all t. This model is appropriate to represent the spread between two commodities whose price differences remain constant on average over time.

Energy commodities as natural gas and electricity exhibit seasonality features which contradict this property; however, the introduction of a seasonal deterministic component an solve the problem. Note that (2.6) may also be written as

$$X(t + dt) = X(t) + \mu dt + \sigma dW_t. \tag{2.11}$$

According to (2.11) the price at time $t + dt$ depends only on $X(t)$ and it does not depend on any preceding value of X. This means that assuming (2.6) is a valid model for the process $X(t)$, the distribution of X at a future date $(t + dt)$ only depends on the current value $X(t)$. So $X(t)$ satisfies the *Markov property*, i.e.,

$$P\left[X(t_{k+1}) \in B \middle| X(t_k), \ldots, X(t_0)\right] = P\left[X(t_{k+1}) \in B \middle| X(t_k)\right] \tag{2.12}$$

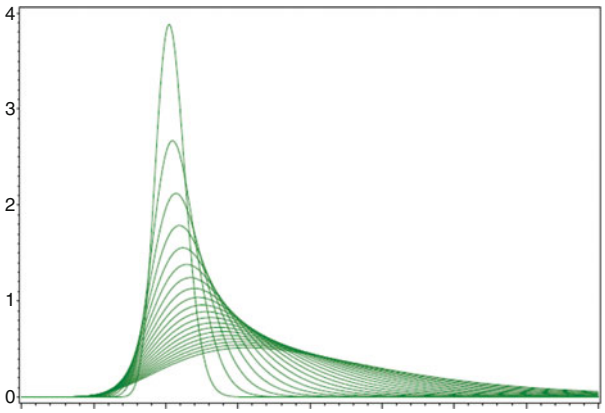

Fig. 2.7 GBM for $\alpha = 0.14$, $\sigma = 0.20$, $S_0 = 1$. The (lognormal) density functions for $t = 0.25, 0.5, \ldots, 1$

and this will play a key role in the founding models of option pricing. This feature is not present in the practice of technical analysis where the historical prices are considered to provide buying or selling signals. Usually the current price is compared to the average of the previous t_{k_i} days. Trading strategies involving *head and shoulders* or the comparison of short-term and long-term moving averages are very popular among commodity trading advisors (CTAs).

Equation (2.11) is an affine function of dW_t and so also $X(t+dt)$ follows a normal distribution and this implies that the process $X(t)$ may assume negative values. These possible negative values for the stock price, which contradict the principle of limited liability attached to a stock (whose market price cannot go below 0), represent the main limitation of this model. The arithmetic Brownian motion is a very popular way of representing a quantity which may be positive as well as negative, and so it can easily represent the behavior of spread contracts between two different commodities or two types of crude oil, when one price has no reason to be consistently higher than the other. So in the case of the valuation of an option on a commodity spread, it will be shown that the computations of expectations involved in this valuation are particularly simple in the case of the ABM.

2.2.2 Geometric Brownian Motion

In 1965, an alternative to Bachelier's model was introduced by Paul Samuelson [14] where the rates of returns, not the stock prices, follow an ABM (Fig. 2.7):

$$\frac{dS_t}{S_t} = \alpha dt + \sigma dW_t \tag{2.13}$$

Basically $\frac{dS_t}{S_t}$ is the return obtained by investing in the stock for the period dt in the case of no dividend payment, the return is solely generated by the price change. The quantity $\alpha dt + \sigma dW_t$ follows a normal distribution.

A straightforward application of Ito's lemma to $ln(S)$ yields the solution

$$S_t = e^{ln(S_0)+\hat{\alpha}t+\sigma W_t} = S_0 e^{\hat{\alpha}t+\sigma W_t}, \tag{2.14}$$

where $\hat{\alpha} = \alpha - \frac{1}{2}\sigma^2$. Therefore S_t is lognormally distributed with

- $E(S_t) = S_0 e^{\hat{\alpha}t}$
- $Var(S_t) = S_0^2 e^{2\hat{\alpha}t}(e^{\sigma^2 t} - 1)$
- Density function

$$f(S,t) = \frac{1}{\sigma S\sqrt{2\pi t}} e^{-(lnS - lnS_0 - \hat{\alpha}t)^2 \frac{1}{2\sigma^2 t}} \tag{2.15}$$

For $p \in [0,1]$ the *p-th percentile* of S_t is $S_0 e^{\hat{\alpha}tN^{-1}(p)\sigma\sqrt{t}}$. Since S_t is the exponential of an ABM, it is called a GBM. If we compute the expectations of both sides of (2.13) we obtain

$$\alpha = \frac{1}{dt} E\left(\frac{dS_t}{S_t}\right), \tag{2.16}$$

which is the *drift* of the process and measures the expected return per unit of time. Estimating the variance of the two sides of (2.13) we obtain

$$\frac{1}{dt} Var\left(\frac{dS_t}{S_t}\right) = \sigma^2. \tag{2.17}$$

It measures the variance of the return per unit of time. Equation 2.13 implies that the expression of S at $t + dt$ is given by

$$\frac{S_{t+dt} - S_t}{S_t} = \alpha dt + \sigma dW_t \tag{2.18}$$

or

$$S_{t+dt} = S_t[1 + \alpha dt + \sigma dW_t], \tag{2.19}$$

which shows that S_{t+dt} only involves S_t and no prior values of S. Like the ABM, the GBM is also a Markov process. Finally, the assumption of constant volatility is obviously not consistent with the behavior of commodity prices. This volatility may be changed to a deterministic function of time at a low mathematical cost. It may be made stochastic either through the introduction of stochastic volatility or by adding a jump component to the model. Figures 2.8–2.10 show that natural gas volatility exhibits seasonality and spikes, crude oil volatility presents spikes, and that there is no particular pattern in electricity volatility.

We have to emphasize that the sign of α is typically perceived as positive since no investor would buy a stock offering a negative expected return. The capital asset pricing model (CAPM) states that, under some equilibrium conditions, the expected return on a risky stock is equal to the risk-free rate, r, plus a risk premium (involving the beta of the stock and the expected excess of performance of the market portfolio over the risk-free rate). In general when we have a risky security whose

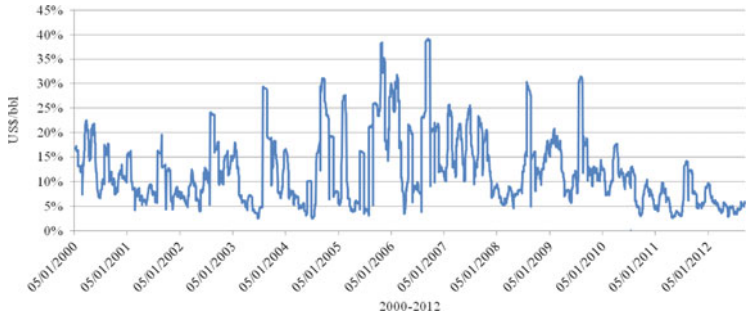

Fig. 2.8 Volatility of NBP returns

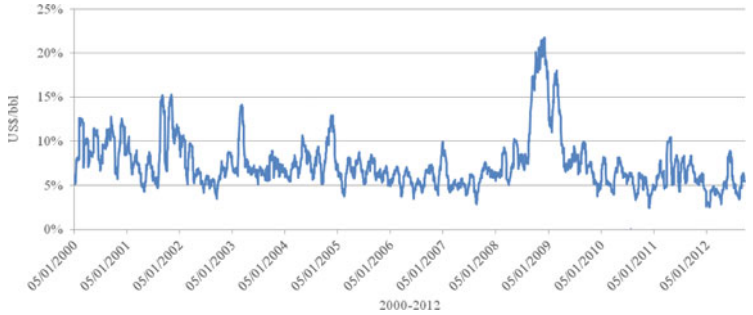

Fig. 2.9 Volatility of EEX returns

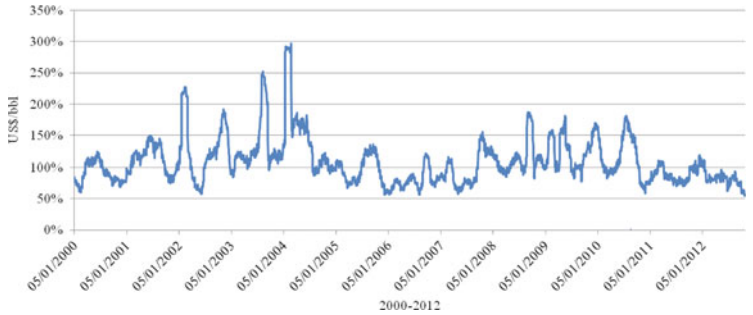

Fig. 2.10 Volatility of Brent returns

dynamic is described by (2.13) the drift is usually higher than the risk-free rate, and positive. This has a fundamental consequence that needs to be kept in mind: besides normality of returns the GBM's key mathematical assumption implies that the stock price grows on average over time. This assumption is the building block to derive the Black–Scholes formula. If we do not believe this property is true for energy commodity prices, the Black–Scholes formula in its original form should not be used to price options written on the spot prices of commodities.

2.2.3 Modeling Seasonality

Natural gas and electricity prices exhibit seasonality in prices, due to changing consumption as a result of weather patterns, as it is shown in Figs. 2.1 and 2.2. To take into account for the seasonal feature in price dynamics a simple and popular representation for commodity spot prices under the real probability measure P may be:

$$ln(S_t) = f_t + X_t, \tag{2.20}$$

where f_t is a deterministic component accounting for the seasonality of prices. and X_t represents the random component which may be described by the most adequate process for the specific commodity price.

The use of the log price guarantees positive values for the prices. The deterministic function f_t can be expressed as a *sin* or *cos* function with annual or semi-annual periodicity, as well as the parameters derived from a database of spot prices. An alternative representation of the seasonal component is the following:

$$f_t = \mu + \sum_{j=1}^{12} \gamma_j D_{jt}, \tag{2.21}$$

where D_{jt} is a dummy variable that takes value 1 if the price we are dealing with refers to a specific month of the year and 0 otherwise. In the case of electricity we may also include the weekly seasonal component:

$$f_t = \mu + \beta_1 D_{1t} + \beta_2 D_{2t} + \sum_{j=1}^{12} \gamma_j D_{jt}, \tag{2.22}$$

where $D_{1t} = 1$ when the price refers to a Saturday and 0 otherwise and $D_{2t} = 1$ when the price refers to a Sunday and 0 otherwise.

2.2.4 The Mean Reversion Feature

An empirical property of commodity prices is the tendency toward lower levels (higher levels), when they are too high (low). This property is called mean reversion and can be modeled using a so-called mean-reverting (MR) process. In the case of a GBM the variance of the distribution of the price return grows linearly with time. In other words, the further out in time we move, the greater is our uncertainty about the value of the factor. However, for commodities, such as natural gas coal, and electricity, we would expect supply shocks to be reasonably short lived with prices in general fluctuating around values determined by the cost of production and changes in demand. So we might therefore reasonably expect energy prices to revert back to their long-term mean.

A Gaussian MR process is described by the following stochastic differential equation:

$$dX_t = \alpha(\theta - X_t)dt + \sigma dW_t \tag{2.23}$$

with $\alpha > 0$.

This model has been introduced in finance in 1977 by Vasicek [16] to represent the random evolution of interest rates; according to (2.23) the term structure moves through changes in the spot rate r_t. The process described by (2.23) prevents the large windings of the ABM described in (2.6) (which may go to infinity positively and negatively and is inappropriate for interest rates that move in a narrow band of the type $[0, 0.2]$; and to avoid the growth over time contained in (2.11) unsuited for interest rates. Equation (2.23) is also known as the Ornstein–Uhlenbeck process. A key feature of this process is obtained by looking, from date t, at the expected change

$$E_t[dX_t] = \alpha(\theta - X_t)dt \tag{2.24}$$

so that, assuming $\alpha > 0$, $E_t[dX_t] > 0$ when $X_t < \theta$, i.e., we expect an increase (decrease) in the stochastic variable level when we are below (above) the level θ. The higher the value of α, the faster the return toward the level θ; α is called speed of reversion, while θ determines the long-run mean level. The distribution of X at any future time is Gaussian, so it allows for negative values. The solution of (2.23) is

$$X_t = e^{-\alpha t}X_0 + \theta(1 - e^{-\alpha t}) + \sigma \int_0^t e^{-\alpha(t-s)}dW_s. \tag{2.25}$$

In addition, we also have that

$$E[X_t | X_0] = e^{-\alpha t}X_0 + \theta(1 - e^{-\alpha t}) \tag{2.26}$$

$$Var[X_t | X_0] = \sigma^2 \int_0^t e^{-2\alpha(t-s)}sd(s) = \frac{\sigma^2}{2\alpha}(1 - e^{-2\alpha t}). \tag{2.27}$$

An extension that guarantees positive values for X has been proposed by Cox, Ingersoll, and Ross (CIR model) [2]:

$$dX_t = \alpha(\mu - X_t)dt + \sigma\sqrt{X_t}dW_t. \tag{2.28}$$

This model is used to describe the dynamics of the instantaneous interest rate. The peculiar form of the diffusion coefficient has been chosen to ensure that the process does not achieve negative values. This model shares with the Vasicek model the form of the drift term, so that it allows for mean reversion and the values of interest rates cannot explode. The level of absolute variance increases with increasing interest rates. Unfortunately, this SDE does not admit an explicit solution, such as for the Vasicek model. This model is less tractable than the Vasicek one: the stationary distribution of the X is related to the noncentral chi-square distribution. We can obtain few properties of the solution such as expected value, variance, and distribution. In particular, the stationary expectation of X_t is the same as in the Vasicek model.

Example: $ln(S_t)$ represents the spread between the UK NBP price and the price of natural gas at the Dutch Title Transfer Facility (TTF Hub) Netherlands. If the price at NBP is greater than the price at TTF plus transportation costs between the two locations then the spread $ln(S_t)$ will be above the long-term level S. In this situation market players will buy at TTF and sell at NBP. This will reduce the demand at NBP and simultaneously drive up the demand at TTF, thereby placing a downward pressure on the price at NBP it and therefore and the spread will move back toward the long-term level. This reaction is represented by the drift term $\alpha(\theta - X_t)dt$, which will be negative. Similarly, if the spread is below the long-term level it will gradually move back toward drift term will be positive. The rate of adjustment back toward the long-term level is determined by the mean-reversion parameter θ. The larger the θ the quicker the price moves toward the mean. The mean-reversion model has experienced considerable empirical success particularly in the natural gas market.

2.2.5 Stochastic Processes with Jumps

The Brownian motion is a process which is continuous in time and space so it cannot capture extreme movements. The Brownian motion is Gaussian and consequently has symmetric distribution with zero excess kurtosis. A way to consider extreme movements, i.e., skewness and excess kurtosis, is to introduce some discontinuity in space, i.e., jumps. Jump diffusion processes may allow to take into account these features, for instance, the Merton Jump Diffusion (MJD) [12] or the Kou Jump Diffusion (KJD) [11] as well as the time-changed Brownian motion, like the Variance Gamma (VG) process.

The Merton jump model takes the form

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t + J_t\, dq_t \tag{2.29}$$

where

- (dq_t) is a Poisson process with intensity λ.
- The process q counts the number of jumps that have occurred and J_t represents the magnitude of the jumps. The jumps are assumed to be independently identically log-normally distributed with parameters (α, β^2). It is also assumed that the jump heights are independent form the jump times.
- The jump process q_t is a discrete event process in the sense that only a finite number of jumps occurs w.pr.1 in finite time intervals.

In (2.29), the increments of the return process consists of three components: a linear drift μdt, a Brownian motion σdW_t representing *normal* price fluctuations that are due to normal clearing imbalances in demand and supply, and a compound Poisson process dq_t that accounts for jumps in prices due to the arrival of news. Thus the continuous part of the stochastic differential equation $\mu dt + \sigma dW_t$ accounts for

the usual fluctuations in S_t, and the jump part dq, driven by a Poisson process, accounts for the extreme events.

For every small time interval, say Δ_t the probability that a single jump occurs is $\lambda\delta_t$, while the overall probability that exactly k jumps occur is

$$(\lambda t)^k \frac{e^{-\lambda t}}{k!} \tag{2.30}$$

and the average time between two jumps is equal to $1/\lambda$. When $dq = 0$ the process behaves as a GBM.

By introducing jumps, we can better capture the abrupt market price changes driven by unpredictable events such as abnormal weather or forced capacity outages. More formally it can be shown that the k period returns under GBM have both skewness and relative kurtosis of zero. However, under the MJD the skewness is given by

$$\frac{\lambda\alpha^3}{(\sigma^2+\lambda(\beta^2+\alpha^2))^{1.5}\sqrt{k}} \tag{2.31}$$

and the relative kurtosis

$$\frac{\lambda(\alpha^4+3\beta^4)}{(\sigma^2+\lambda(\beta^2+\alpha^2))^2 k}. \tag{2.32}$$

The MJD [12] model has gained considerable popularity in the finance community because it can lead to analytical solutions for call and put options and interest rate derivatives such as caps, floors, and swaptions. However, it does not have very desirable properties in terms of analytical tractability in the case of path-dependent (exotic) options or econometric estimation. In addition since there is only one jump component *good news* and *bad news* are not distinguished by their intensity or distributional characteristics. In the case of energy price returns there appears to be an asymmetric effect in terms of the response to good and bad news. Bad news, resulting in a negative return, increases volatility by more than good news, resulting in a positive return, of the same magnitude. One way to incorporate this would be to have J_t drawn from a normal distribution, but with different jump volatilities for good and bad news. Such an extension, although relatively easy to implement, leads to a loss of analytical tractability of MJD.

An alternative model which is becoming quite popular is the double exponential Jump diffusion model of Kou [11].

The Kou Jump Diffusion model (KJD) assumes that the jump sizes follow a double exponential distribution with parameters (p, η_1, η_2). It assumes that the sequence of nonnegative random variables, J_t^i are i.i.d. (indipendently, identically distributed) such that the natural logarithm given by $ln(J_t^i)$ has an asymmetric double exponential distribution:

$$J_t^i = \begin{cases} J_t^{i+} & w.pr.\ p, \\ J_t^{i-} & w.pr.\ 1-p, \end{cases} \tag{2.33}$$

where $J_t^{i+} \sim$ Exponential(η_1) and $J_t^{i-} \sim$ Exponential (η_2).

In this case ξ^u and ξ^d are exponential random variables with means $1/\eta_1$ and $1/\eta_2$, respectively. This model allows the price return distributions to be asymmetric and fat tailed; in addition it leads to nearly analytical solutions to many option pricing problems.

2.2.6 Mean-Reverting with Jumps

To combine the mean-reversion feature with the jump component, a typical feature of energy commodity prices, a natural extension of the mean-reversion model is proposed. A jump component added to the standard mean-reverting process in order to reproduce the spiky behavior. In its simplest formulation, the jump component is a standard compound Poisson process similar to that proposed by Merton. The associated stochastic differential equation can be written as follows:

$$dS_t = -\lambda\,(ln(S_t) - l(t))S_t dt + \sigma_t S_t dW_t + d\Pi_t, \tag{2.34}$$

where $d\Pi_t = J_t\, dq_t$ is a compound Poisson process.

This class of processes can be used in the electricity trading sector since it merges together the two main features of electricity spot price dynamics: mean reversion and jumps. Unfortunately, the ways in which this class of models merge mean reversion and jumps cannot be considered very particularly slow. realistic, given that the reversion intensity is constant for both normal and spike regimes, while empirically we observe that the spike-reversion intensity is much more significant than the standard mean reversion [14]. This implies that when a positive jump occurs the reversion to the normal regime is slower. Moreover, the simple compound Poisson process Π_t is characterized by a constant jump frequency while we know also that the probability of a spike occurrence is not constant over time, but is often cyclical, since it depends on some price determinants which are themselves periodical.

Parameter estimation also requires some particular techniques, for instance, filtering methods have to be applied in order to separate, in the original empirical time series, jump components from diffusive components, and this operation may have large influence on the estimate of the whole set of parameters.

2.2.7 Stochastic Volatility

The assumption of a constant volatility parameter is another unrealistic assumption that has to be released in most of the energy pricing models. Many different theories have been suggested to deal with the idea of nonconstant volatility, the application of GARCH-type models is becoming increasingly popular as well as the use of mixed distributions. GARCH models have been the most frequently applied class of time-varying volatility models in empirical research. This is mainly due to the problems

which arise as a consequence of the intractability of the likelihood function of the stochastic volatility model which prohibits its direct evaluation. We suggest here to directly specify a stochastic differential equation with a time-varying volatility parameter. In this approach the asset price S_t satisfies the following stochastic differential equation:

$$\frac{dS_t}{S_t} = \alpha dt + \sigma_t dW_t, \tag{2.35}$$

where $\sigma_t = f(Y_t)$ and $dY_t = (a + bY_t)dt + cY_t dV_t$ and $dV_t = \rho dW_t + \sqrt{1-\rho^2}dZ_t$, with W_t and Z_t independent Brownian motions.

In this setup the model constants are given by the parameters α, a, b, and c. The parameter ρ is the correlation between Z and V. It is assumed that random shocks affecting the variance are correlated to the random shocks of the asset price. In this model there are two sources of risk, namely the future path of the asset price and the future path of volatility.

2.2.7.1 Models for Stochastic Volatility

One of the most popular model for stochastic volatility was proposed by Hull and White [10] and assumes $\sigma_t = \sqrt{Y_t}$.

In the Hull and White model high variance of the volatility parameter drives the fat tails of the price distribution. Thus extreme returns of positive and negative sign are more likely than in the GBM where the asset price follows a lognormal distribution. If $\rho = 0$ so that shocks to returns and shocks to volatility are uncorrelated the price distribution is symmetric and leptokurtic. In fact the sign of the correlation determines the symmetry of the distribution; negative correlation results in the left tail of the price distribution containing more probability mass than the right tail.

The model proposed by Scott [13] in 1987 is furnished by a pair of Ito stochastic differential equations:

$$\begin{aligned} \frac{dS_t}{S_t} &= \mu dt + m e^{Y_t} dW_t^{(1)} \\ dY_t &= -\alpha Y_t dt + k dW_t^{(2)} \end{aligned} \tag{2.36}$$

where S_t is the asset price. The parameters α, m, and k are positive and nonrandom quantities and are uncorrelated Wiener processes in the original formulation. Some extension assume $dW_t^{(i)} = \eta_t^{(i)} dt, i = 1,2$ are correlated Wiener processes, i.e., $\eta_t^{(i)}$ are zero-mean Gaussian white noise processes with cross correlations given by

$$\left\{ \eta_{t1}^{(i)} \cdot \eta_{t2}^{(j)} \right\} = \rho_{ij} \delta(t_1 - t_2)$$

where $\rho_{ii} = 1, \rho_{ij} = \rho (i \neq j, -1 \leq \rho \leq 1)$.

From Eq. (2.36) we see that

$$Y_t = Y_0 e^{-\alpha(t-t_0)} + k\int_{t_0}^{t} e^{-\alpha(t-s)} dW_s^{(2)}$$

where we assume that the volatility process Y_t starts at certain initial time t_0 (which can be set equal to 0) with a known value $Y_{t0} = Y_0$. The process Y_t is Gaussian with conditional first moment and variance given by

$$\begin{aligned} E[Y_t|Y_0] &= Y_0 e^{-\alpha(t-t_0)} \\ var[Y_t|Y_0] &= \frac{k^2}{2\alpha}\left(1 - e^{-2\alpha(t-t_0)}\right). \end{aligned}$$

2.3 State Variable Models for Energy Prices

Besides improvements to the price process dynamics brought by mean reversion or the jumps discussed in previous sections, another way to increase the quality of modeling is not to be limited to a single-state variable (namely, the spot price), but to enrich the representation by the introduction of one extra or several state variables. For energy commodities these potential state variables may be the convenience yield, the long-term value of mean reversion, and the stochastic volatility.

2.3.1 A Stochastic Convenience Yield Model

Gibson and Schwartz [8] note that the convenience yield for crude oil has shown to be, in theoretical and empirical research, a key factor driving the relationship between spot and futures prices. Gabillon [6] also comments that the shapes of backwardation (future price below expected spot price) and contango (future price above expected spot price) successively displayed over time by oil forward curves are inconsistent with a constant convenience yield. Hence, the following two-state variable model for oil-contingent claim pricing was proposed:

$$\begin{cases} \frac{dS_t}{S_t} = \mu dt + \sigma_1 dW_t^1, \\ dy_t = k(\alpha - y_t)\,dt + \sigma_2 dW_t^2, \\ dW_t^1 \cdot dW_t^2 = \rho dt. \end{cases} \tag{2.37}$$

The first equation describes a classical GBM for the oil spot price. The second equation defines the Ornstein–Uhlenbeck process driving the convenience yield and leading to positive and negative values: For different period and commodity, the convenience yield may have a different sign.

2.3.2 A Stochastic Volatility Model

Eydeland and Geman [4] propose an extension of the Heston [9] stochastic volatility model to describe natural gas or electricity price dynamics. They introduce mean reversion in the spot price and propose the following two-state variable model:

$$\begin{cases} dS_t = k(a - ln(S_t))S_t dt + \sigma_t S_t dW_t^1, \\ d\Sigma_t = b\,(c - \Sigma_t) + e\sqrt{\Sigma_t} dW_t^2, \\ dW_t^1 \cdot dW_t^2 = \rho dt. \end{cases} \tag{2.38}$$

Here $\Sigma_t = \sigma_t^2$. The parameters k, a, b, c, and are all positive. The correlation coefficient ρ is in general negative since, in contrast to stock prices, the volatility of commodity prices tends to increase with prices, this is known as the inverse leverage effect which leads in option prices to a volatility smile skewed to the right [5]. The parameters in (2.38) need to be estimated under the real probability measure P by methods, such as maximum likelihood, from a database of commodity spot prices. We can observe that, the dynamic of the second state variable, Σ_t, in contrast to the second equation in (2.37) where the convenience yield may be negative, thanks to the presence of $\sqrt{\Sigma_t}$ in the coefficient of dW_t^2 where $\sigma_t^2 = \Sigma_t$, remains positive. The parameter a represents the equilibrium/cost of production value of the commodity under analysis. The process $(\Sigma_t)_{t\geq 0}$ allows to introduce positivity in the MR Vasicek model. Mathematically, it is a much more complex process, since the quantity Σ_t is no longer normally distributed.

The model may be enriched adding a jump component to the first equation to take into account for random changes in the total volatility of S_t due to arrival of jumps and to random moves of σ_t in the diffusion component.

2.3.3 A Three-State Variable Model for Oil Prices

If we consider any commodity at the beginning of a bullish cycle, the mean-reversion feature is no longer so clear, at least toward a fixed level as described in the above subsection.

A valid model [7], if we do not wish to introduce jumps in trajectories, may be the following:

$$\begin{cases} dS_t = k(L_t - ln(S_t))S_t dt + \sigma_t S_t dW_t^1, \\ \frac{dL_t}{L_t} = \mu dt + \sigma_2 dW_t^2, \\ d\Sigma_t = b\,(c - \Sigma_t)\,dt + e\sqrt{\Sigma_t} dW_t^3, \\ dW_t^1 \cdot dW_t^3 = \rho dt. \end{cases} \tag{2.39}$$

where

- $\Sigma_t = [\sigma_t]^2$

- The correlation between W^1 and W^2 resp. W^2 and W^3 may be different from zero; however, in some models independence is assumed

A positive drift in the second equation generates a rise, on average, of the value L_t toward which the commodity spot price S_t tends to revert, while this spot price itself may fluctuate significantly around L_t depending on the arrival of positive or negative news about the situation of world- and company-specific reserves. We should note in particular that the stochastic volatility (σ_t) may take large values on some days and be the mathematical explanation of major movements in spot price S_t.

2.4 Regime Switching Models

In the previous sections we have mentioned the fact that electricity spot price dynamics is characterized by normal and spiky behaviors or regimes. Hence, a natural way of mathematically representing this feature is through the class of multiple regime processes. According to this modeling approach, the electricity spot price is assumed to follow two different and independent regimes. The first one, the "mean-reverting regime," is intended to describe the non-spiky behavior of the dynamics, while the second is intended to replicate the "spiky regime" of the process. The spikes in the second regime are modeled with a simple lognormal behavior whose mean and standard deviation are much higher than those of the mean-reverting regime process.

Formally, we have the following specification:

$$\begin{cases} dln(S_t) = -\lambda(ln(S_t) - \mu_t)dt + \sigma_t dW_t, \\ dln(S_t) = \mu_t dt + \sigma_t dW_t. \end{cases} \tag{2.40}$$

The first equation describes the dynamics in the stable regime, while the second equation describes the spiky ones. The switching from one regime to the other one is governed by a two-state Markov process M_t for each t and $T > 0$ with $T > t$, characterized by the following transition matrix $P(t,T)$:

$$P(t,T) = \begin{pmatrix} Pss(t,T) & Psr(t,T) \\ Prs(t,T) & Prr(t,T) \end{pmatrix}. \tag{2.41}$$

The role of the transition probability matrix is very important since it determines the likelihood to jump from one regime to the other at every time. Conditional on the regime state, the parameters of the two processes can be easily estimated by means of ML estimators based on the normal distribution of the stochastic terms of the two processes. However, in practice, parameter estimation is not that easy since we do not know the underlying regime at each single time (the regime is a latent variable). The Kalman filtering methodology helps us to solve estimation problems, but again we stress the fact that filtering procedures usually have a big impact on estimation results. Another source of complexity is represented by the fact that the regime is not the only latent variable, prices are as well. Prices in the mean-reverting regime

continue as a latent process during a spike regime and thus are unobservable there. From a mathematical point of view, we have a bi-dimensional Markov process, and this creates serious problems, when estimating the parameters. Deng (1999) proposed a methodology to circumvent estimation problems for Markov switching processes. Excluding estimation problems, however, this class of processes has the important advantage of producing closed form solution both for forward and plain-vanilla options prices, as linear combinations of prices under the two regimes and regime probabilities. Finally, though, for this class of models, the problem of market incompleteness arises since the source of uncertainty is multiple and not completely avoidable.

Acknowledgement. The author thanks B. Analui, R. Kovacevic, and G. Pflug for carefully reading and revising the manuscript.

References

1. Bachelier L (1900) Théorie de la speculation. Annales Scientifiques de École Normale Superieure 17:21–86
2. Cox JC, Ingersoll JE, Ross SA (1985) A theory of the term structure of interest rates. Econometrica 53:385–408
3. Deng S (2000) Stochastic models of energy commodity prices and their applications: Mean-reversion with jumps and spikes. University of California Energy Institute Working Paper No. 73
4. Eydeland A, Geman H (1998) Pricing power derivatives. RISK, September 1998
5. Eydeland A, Wolyniec K (2003) Energy and power risk management: new developments in modeling, pricing and hedging (Wiley Finance Series). Wiley, Chichester
6. Gabillon J (1991) The term structure of oil futures prices. Working Paper 17, Oxford Institute for Energy Studies
7. Geman H (2000) Scarcity and price volatility in oil markets. EDF Trading Technical Report
8. Gibson R, Schwartz ES (1990) Stochastic convenience yield and the pricing of oil contingent claims. J Financ 45:959–976
9. Heston SL (1993) A closed-form solution for options with stochastic volatility with applications to bond and currency options. Rev Financ Stud 6(2):327–343
10. Hull J, White A (1987) The pricing of options on assets with stochastic volatility. J Financ 42:281–300
11. Kou S (2002) A jump diffusion model for option pricing. Manage Sci 48(8):1086–1111
12. Merton R (1976) Option pricing when underlying stock returns are discontinuous. J Financ Econ 3:125–144
13. Samuelson P (1965) Rational theory of warrant pricing. Ind Manage Rev 6(Spring):13–31

14. Schwartz ES (1997) The stochastic behavior of commodity prices: implications for valuation and hedging. J Financ 52(3):923–973
15. Scott IO (1987) Option pricing when the variance changes randomly. Theory, estimation and an application. J. of Financial and quantitative analysis 22: 419–438
16. Vasicek O (1977) An equilibrium characterization of the term structure. J Financ Econ 5(3):177–188

Chapter 3
Price Dynamics in Electricity Markets

Florentina Paraschiv

Abstract With the liberalization of global power markets, modeling of exchange-traded electricity contracts has attracted significantly the attention of both academic and industry. In this paper we offer an overview of the most common deseasonalization techniques and modeling approaches in the literature. We extract the deterministic component of EEX Phelix hourly electricity prices and we discuss different financial and time-series models for their stochastic component. Additionally we apply extreme value theory (EVT) to investigate the tails of the price changes distribution. Generally our results suggest EVT to be of interest to both risk managers and portfolio managers in the highly volatile electricity markets.

3.1 Introduction

Finding realistic models to describe electricity prices is essential for the valuation of power grids, for the risk managers in the estimation of risk measures as well as for portfolio managers to determine worst-case scenarios in very turbulent markets.

Electricity prices pose a particular challenge for researchers, given their main characteristics: seasonalities, mean reversion, extremely large price movements as well as negative prices. Seasonality represents the deterministic component of the prices. A successful modeling approach is based on a rigorous deseasonalization technique. Therefore the seasonal components of electricity prices are discussed here in detail and we further offer a review of the main procedures used in the literature to deseasonalize them.

F. Paraschiv (✉)
Institute for Operations Research and Computational Finance, University of St. Gallen, Bodanstrasse 6, CH-9000, St. Gallen, Switzerland
e-mail: florentina.paraschiv@unisg.ch

R.M. Kovacevic et al. (eds.), *Handbook of Risk Management in Energy Production and Trading*, International Series in Operations Research & Management Science 199,
DOI 10.1007/978-1-4614-9035-7_3,

Beside the deterministic component, electricity prices have also a stochastic component, given by the inefficient storing capacities for electricity or by the intense use of renewable energy over the last years for power generation. Thus, the production has to follow the more or less inelastic demand and traders of electricity with physical delivery are forced to balance their accounts in every single hour, independently of the actual offers. This leads to extreme fluctuations in electricity prices, which make them difficult to forecast. The most frequently applied models for the stochastic component of electricity prices are financial and time-series models. In this paper we offer a comparative view of three popular financial models: Brownian motion, Ornstein–Uhlenbeck process, and the well-known Pilipovic model. From the time-series models we selected for discussion the most popular ones: ARMA (autoregressive moving average) models as well as GARCH (generalized autoregressive conditional heteroscedasticity) models. Given the extremely large price movements in electricity prices, we model extreme tail quantiles with extreme value theory (EVT) applied to EEX Phelix hourly electricity prices. We show that this procedure describes more realistically extreme tail quantiles than the classical time-series models.

In Sects. 3.2 and 3.3 we offer a description of the main characteristics of electricity prices and we discuss the main deseasonalization techniques. Section 3.4 classifies and discusses comparatively different modeling approaches for electricity prices. Section 3.5 shows an application of EVT for modeling extremely large electricity price changes. Section 3.6 offers a summary of the paper.

3.2 Characteristics of Electricity Prices

Electricity markets have particularities which clearly distinguish them from other commodity markets. Given the lack of efficient storing opportunities for electricity, which prevents intertemporal smoothing of the demand by holding storages, extremely large price movements (spikes) as well as various cyclical patterns of behavior occur. Supply and demand determine market prices which have to correspond exactly at any location and at any time. Because of limited efficient storing capacities, the grid operators have difficulties to balance out hard-to-predict variations in power production and consumption in order to cover peak loads. The main characteristics of electricity prices are seasonalities, mean reversion, and negative prices.

3.2.1 Seasonalities

The seasonal behavior of electricity prices is one of the most complicated ones among all commodities. It is predominantly caused by the almost inelastic, at least in the short term, demand for electricity, which by itself shows pronounced patterns caused by economic and business activities.

Electricity prices reveal three types of cyclical patterns: daily, weekly, and yearly seasonality (see [2]). As emphasized in [24], the amount of electricity demanded depends mostly on the level of human activity, as well as on the weather and climate conditions. Essentially, two Off-Peaks can be observed: Off-Peak 1 and Off-Peak 2. Off-Peak 1 is represented by the first 7h of the day, when most people sleep and fewer businesses are operating. Off-Peak 2 are the last four hours of the day, when most human and business activities have ceased. In between, during the Peak hours (between 8 a.m. and 8 p.m.), electricity demand increases drastically, an intensification that is linked to people getting ready for the day ahead and to the start of the business activities (see [4]). Figure 3.5 summarizes the autocorrelation function for the baseload hours (prices for the 24h of the day) versus the Off-Peak and Peak hours. On average, electricity prices are relatively constant during the working week (holidays have to considered separately), whereas at week ends and during holidays electricity prices fall. In addition to this, the hourly pattern differs depending on the season. While in winter we can observe two peaks (at noon and one evening peak, at 7 p.m.), in summer we typically observe only one peak at noon (see Fig. 3.1).

3.2.2 Mean Reversion

In the short term electricity prices are characterized by jumps or spikes. However, in the long run they revert to the mean-reverting level (MRL) (see [5]). The long-term price level is characterized by the marginal costs of production. These can be constant, periodic, or periodic with trend. This argument refers to the theory of perfect competition, i.e., if demand for electricity is high, production capacities with high marginal costs are implemented, whereas if demand is low, production capacities with low marginal costs are used and consequently prices fall. The alignment from capacities with lowest to ones with highest marginal costs is depicted by the so-called merit order. This notion implies that there is more than one constant mean-reversion level, depending on the time of the day, of the week, and of the year (see previous section). Hence, the concept of mean reversion implies that electricity prices return to their respective usual level (see [10]).

3.2.3 Negative Prices

The limited storing capacities and the limited load change flexibility caused the negative electricity prices at EEX. From an economic perspective, negative prices can be rational, e.g., if the costs to shut down and ramp up a power plant unit exceed the loss for accepting negative prices (see [14]). Since September 1st 2008, negative price bids have been allowed at the German power exchange EEX as the first energy exchange in Europe. In our analysis we will refer to the German power exchange EEX Phelix hourly electricity prices between September 2008 and December 2011.

Historical spot market data over the investigated period shows a total amount of about 100h with negative prices. Mostly, they occur in the night and morning hours (23:00 to 08:00) as displayed in Fig. 3.6. Furthermore, they occur with higher frequency on Mondays and Sundays (Fig. 3.7). As shown in the histogram in Fig. 3.8, the absolute frequency of the prices presents clusters of 2EUR/MWh.

3.3 Deseasonalization Techniques

The electricity prices are explained by two fundamental components: a deterministic component represented by the typical seasonality pattern and the price uncertainty as stochastic component of the prices. The load, as one main driver of electricity prices, shows some noticeable patterns, such as the peak at midday in summer days. The electricity prices follow more or less typical seasonality patterns, which are described in the literature by deterministic functions. However, as discussed in [14], beside the deterministic impact factors, electricity spot prices are also influenced by uncertain parameters like power plant outages and fluctuant renewable electricity generation. These uncertainties are drivers of the stochastic component of the spot prices. The current section offers an overview of the deseasonalization techniques applied in the literature for electricity prices, while the next section will classify modeling approaches for the stochastic component.

3.3.1 Preliminaries

In order to incorporate the seasonal feature while taking the property of mean reversion into account, the spot price can be expressed as a combination of these two components:

$$P_t = f_t + \sum_{i=1}^{n} X_{i,t}, \tag{3.1}$$

where f_t is the deterministic component seasonality and X_t is the stochastic part. There is a discussion in the literature whether level or logarithmic prices should be modeled. Accordingly to [27] or [13] the drawback of deseasonalizing day-ahead spot prices is that the residuals of $P_t - f_t$ can become negative, which prevents the use of logarithms on the deseasonalized spot prices. In addition, the authors mention as well that the advantage of using the spot prices instead of their logarithms is that on average they yield better parameter estimates. In the current paper we will look at the level of the prices.

Demand and supply of electricity show seasonal fluctuations, which traduce into the seasonal behavior of spot electricity prices. Figure 3.1 depicts the average price per month (January and July) and type of day (week day and weekend day/holiday, respectively). This chart confirms the seasonality pattern of the electricity prices.

Over the 24h time period, prices move in a distinct hourly pattern, which follows the demand for electricity. As we have already mentioned, prices start increasing when people get ready for the day (around 6–7 a.m.) and decrease after 8 p.m., when business activities are over. We observe differences in the prices between winter and summer time as well as between workdays and weekend days (see [10]). Often researchers look distinctively at different hours within a day, given their distinctive patterns: they distinguish between baseload and Peak hours and even more between Off-Peak I and Off-Peak II hours.

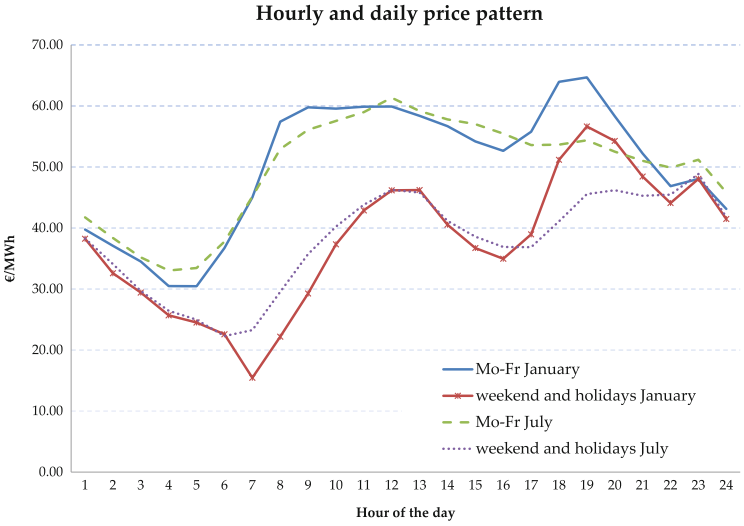

Fig. 3.1 Hourly and daily day-ahead price patterns for EEX Phelix

3.3.2 Overview of Deseasonalization Techniques

The aim of deseasonalization techniques is to reduce the predictable pattern of electricity prices, in order to delimitate the stochastic component of the prices. We firstly remove the long-term trend from the hourly electricity prices:

$$f_t^{trend} = a^{trend} + b^{trend} t. \tag{3.2}$$

The constant term a of the equation may be interpreted as the fix costs of power production. The term t represents the "long-run linear trend in the total production cost," which is related to macroeconomic variables like inflation and, hence, depicts a positive trend.

Secondly, the seasonality of de-trended prices should be removed. Based on the discussion in [10], we give an overview of the several suggestions in the literature to describe the daily, weekly, and annual cycles of electricity prices. There are many deseasonalizing techniques in the literature. Thus, Knittel and Roberts [15] imple-

ment piecewise constant functions; [7, 22, 28] adopt sinusoidal functions; [17] use a combination of both.

3.3.2.1 Yearly Seasonality

The yearly seasonality can be modeled with the classical trigonometric functions:

$$f_t^{season} = a + b_{1,t} \cos\left(\frac{2\pi}{8,760} \cdot (t-\tau)\right) + b_{2,t} \sin\left(\frac{2\pi}{8,760} \cdot (t-\tau)\right). \tag{3.3}$$

The parameter τ defines the phase shift, i.e., the starting point of a seasonal oscillation, and 8,760 is the number of hours in one year. The use of trigonometric functions to define the yearly season is a common method in the literature. However, this method alone does not deliver satisfactory results (see [14]). The use of only trigonometric functions for the EEX prices is indeed not suitable, as they do not show a strong seasonality over the year—after all, some winters have almost spring-like temperatures, and vice versa. In order to make the explanatory power of the trigonometric functions stronger, other variables need to be added, e.g., time (see [27]).

Another method suggested by [14] is the use of the monthly average prices $\bar{P}_{m'}$ as a seasonality factor over the year. The seasonality over the year in this case defines the mean values of the hours of the respective months as a seasonality component:

$$f_t^{season} = \sum_{m'=1}^{12} \bar{P}_{m'} * 1(m' \mid m' = m(t)). \tag{3.4}$$

3.3.2.2 Weekly Seasonality

The weekly cycle can be established using several methods. The first one is called adjusted absolute sinus-function (aasf) (see [14]):

$$f_t^{weekly} = \alpha + d \left| \sin\left(\frac{\pi * t}{168} - \varphi\right) \right|. \tag{3.5}$$

The phase-shift parameter φ is determined as the deviation from the point in which the weekly cycle reaches its minimum in the observation set. For this purpose, we calculate the mean values of the electricity prices for each day of the week. This calculation delivers the seventh hour of Sunday as the minimum price. Then we take the distance of all hours of the week from the respective Sunday's seventh hour (see [14]). An alternative to verify to what extent a (daily) seasonality exists within the week is to include dummy variables. For more information about this method, see [2] or [10].

3.3.2.3 Daily Cycle

The daily cycle is defined as the mean over the 24h of the day and is then removed. Different daily cycles are determined for each season: winter, spring, summer, and autumn. Thus, we take daily average prices for each weekday d dependent on the season:

$$f^{daily}_{i,season} = \frac{24}{T_{season}} \sum_{d=0}^{\left(\frac{T_{season}}{24}\right)-1} P_{i+24d,season}, \tag{3.6}$$

where d is one of the weekdays, i is the hour of the day, and T is the maximum number of hours overall d weekdays in one season (winter, spring, summer, or autumn). There are two ways used in the literature to deseasonalize: to split the seasonal decomposition into daily, weekly, and yearly seasonality and to estimate them separately or to consider them simultaneously (see [14]):

$$f_t = a^{trend} + b^{trend}t + c\frac{24}{T_{season}} \sum_{d=0}^{\left(\frac{T_{season}}{24}\right)-1} P_{i+24d,season} + d\left|\sin\left(\frac{\pi * t}{168} - \varphi\right)\right| + \\ + e\sum_{m'=1}^{12} \bar{P}_{m'} * 1(m'|m' = m(t)). \tag{3.7}$$

3.3.3 Application

In this section we derive the seasonality shape for EEX Phelix hourly electricity prices quoted at the European energy exchange (EEX), between September 2008 and December 2011. The derivation of the deseasonalization shape follows the procedure described in [2]. In a first step, we identify the seasonal structure during a year with daily prices. In the second step, the patterns during a day are analyzed using hourly prices. Let us define two factors, the factor-to-year ($f2y$) and the factor-to-day ($f2d$) (following the usual notation in [2]). By $f2y$ we denote the relative weight of an average daily price compared to the annual base of the corresponding year:

$$f2y_d = \frac{S^{day}(d)}{\sum_{k\varepsilon \text{year}(d)} S^{day}(k)\frac{1}{K(d)}}. \tag{3.8}$$

$S^{day}(d)$ is the daily spot price in the day d, which is the average price of the hourly electricity prices in that day. $K(d)$ denotes the number of days in the year when $S^{day}(d)$ is observed. The denominator is thus the annual base of the year in which $S(d)$ is observed.

To explain the $f2y$, we use a multiple regression model (similar to [2]):

$$f2y_d = \alpha_0 + \sum_{i=1}^{6} b_i D_{di} + \sum_{i=1}^{12} c_i M_{di} + \sum_{i=1}^{3} d_i CDD_{di} + \sum_{i=1}^{3} e_i HDD_{di} + \varepsilon \tag{3.9}$$

- $f2y_d$: factor-to-year, daily-base-price/yearly-base-price
- D_{di}: six daily dummy variables (for Mo–Sat)
- M_{di}: twelve monthly dummy variables (for Feb–Dec); August will be subdivided in two parts, due to summer vacation
- CDD_{di}: cooling degree days for three different German cities
- HDD_{di}: heating degree days for three different German cities

where CDD_i/HDD_i are estimated based on the temperature in Berlin, Hannover, and Munich:

- Cooling degree days (CDD) = $\max(T - 18.3°\text{C}, 0)$
- Heating degree days (HDD) = $\max(18.3°\text{C} - T, 0)$

We transform the series $f2y_d$ from daily to hourly, by considering the same factor-to-year $f2y_d$ for each hour t observed in the day d. In this way we construct hourly $f2y_t$ series, which later enter the shape s_t. The $f2d$, in contrast, indicates the weight of the price of a particular hour compared to the daily base price:

$$f2d_t = \frac{S^{hour}(t)}{\sum_{k\varepsilon \text{day}(t)} S^{hour}(k)\frac{1}{24}} \tag{3.10}$$

with $S^{hour}(t)$ being the hourly spot price at the hour t. We know that there are considerable differences both in the daily profiles of workdays, Saturdays, and Sundays, but also between daily profiles during winter and summer season. Thus, following [2] we suggest to classify the days by weekdays and seasons and to choose the classification scheme presented in Table 3.1. The workdays of each month are collected in one class. Saturdays and Sundays are treated separately. In order to obtain still enough observations per class, the profiles for Saturday and Sunday are held constant during three months.

Table 3.1 Assignment of each day to 1 out of the 20 profile classes

	J	F	M	A	M	J	J	A	S	O	N	D
Week day	1	2	3	4	5	6	7	8	9	10	11	12
Sat	13	13	14	14	14	15	15	15	16	16	16	13
Sun	17	17	18	18	18	19	19	19	20	20	20	17

The regression model for each class is built quite similarly to the one for the yearly seasonality. For each profile class $c = \{1, \ldots, 20\}$ defined in Table 3.1, a model of the following type is formulated:

$$f2d_t = a_o^c + \sum_{i=1}^{23} b_i^c H_{t,i} + \varepsilon_t \quad \text{for all } t\varepsilon c, \tag{3.11}$$

where $H_i = \{0, \ldots, 23\}$ represents dummy variables for the hours of one day.

The seasonality shape sw_t can be calculated by $sw_t = f2y_t \cdot f2d_t$. sw_t is the forecast of the relative hourly weights and it is additionally multiplied by the yearly average prices, in order to align the shape at the prices level. This yields the seasonality shape s_t which is finally used to deseasonalize the electricity prices. Figure 3.2 shows the autocorrelation function of the hourly prices before and after deseasonalizing.

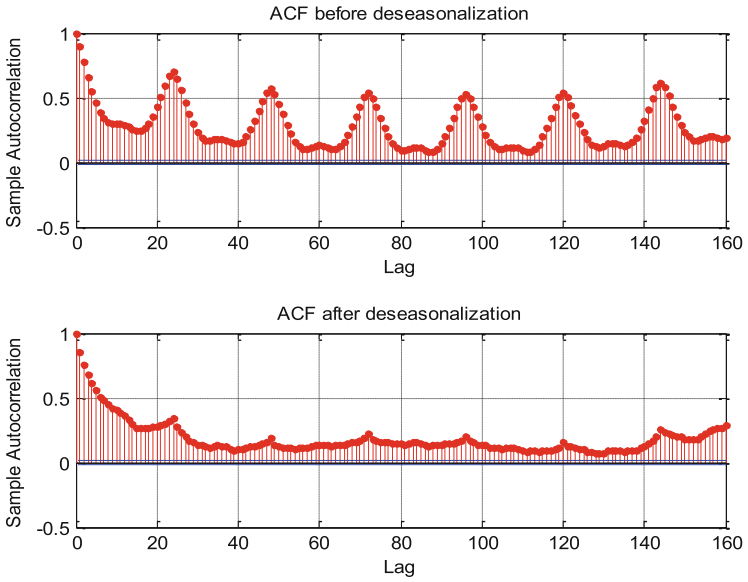

Fig. 3.2 Autocorrelation function before and after deseasonalization

The deseasonalized series is assumed to contain only the stochastic component of electricity prices, such as the volatility and randomly occurring jumps and peaks, which can be simulated via different stochastic processes, as described in the following section.

3.4 Modeling Approaches for Electricity Prices

There are many different theoretical methods that can be applied for electricity price simulations depending on the research question or planning tasks. Thus, the different methods cannot be directly compared with each other as each method has its strengths and its weaknesses. Accordingly to [26], these methods can be classified as:

- Fundamental models
- Game theoretic models

- Financial mathematics models
- Statistical and econometric time-series models
- Technical analysis or expert system

3.4.1 Fundamental Models and Game Theoretic Approaches

Fundamental models consider modeling of the whole electricity system with all suppliers, whereas each single power plant or technology classes are described separately in the modeling approach. This type of models include a detailed analysis about the electricity demand as well as capacity use and maintenance hours of power plants. They are used to produce scenarios for electricity prices, which are further integrated for middle- to long-term planning tasks and price forecasts (see [14, 19, 25]).

Game theoretic approaches consider the strategic behavior of different market stakeholders [14]. These models simulate competitive electricity markets and analyze long-term equilibriums on the whole-sale market in general based on a Cournot–Nash framework (see [12, 16]). This type of models is used to test different market design options and to analyze the behavior of market participants (see discussion in [14]).

3.4.2 Financial Models

Financial and time-series models are calibrated on historical hourly or daily prices and used for short-term price forecast in risk management. Financial mathematical models such as geometric Brownian motions or mean-reversion processes are one of the most applied stochastic processes for electricity prices. They deal with the volatility of electricity prices and can be used for derivative pricing or real options in energy markets [11]. As an example for mean-reversion processes is the Ornstein–Uhlenbeck process which, formulated for electricity price changes, reads

$$dX(t) = k_1(\mu_1 - X(t))dt + \sigma_1 dW_1(t). \tag{3.12}$$

The first term of the mean-reversion process is the drift component. The parameter k_1 describes the speed of the reversion of the stochastic component of the electricity prices to their long-term mean μ_1. The economic interpretation of this mean-reversion component is that stochastic price fluctuations around the mean and price peaks are only temporary, caused by, e.g., power plant outages or capacity storages [14]. The second term, the stochastic component $dW_1(t)$, corresponds to the standard Brownian motion.

Another class of financial models is represented by two-factor models, which distinguish between the short- and long-term dynamics of the prices. Examples in

this sense are the Pilipovic model or the model proposed by [23] who applied it to oil markets and expanded by [17, 18]. The Pilipovic model is well established in commodity markets and known as long-term/short-term model. The short-term deviations are explicitly modeled as the deviations from the long-term mean.

The two-factor [21] model under $\mathscr{P}$ is

$$\begin{aligned} dX_t^{Und} &= k_2(L_t - X_t^{Und})dt + X_t^{Und}\sigma_2 dW_{2t}, \\ dL_t &= \mu_2 L_t dt + L_t \sigma_3 dW_{3t}, \\ dW_{2t}dW_{3t} &= 0, \end{aligned} \tag{3.13}$$

where

X	Spot price
X^{Und}	Underlying spot price value
L_t	Equilibrium price
t	Time of observation
k_2	Rate of price mean reversion
σ_2	Volatility
μ_2	Drift of the long-term equilibrium price
σ_3	Volatility of the long-term equilibrium price
dW_2	Random stochastic variable defining the randomness in the spot prices
dW_3	Random stochastic variable defining the randomness in equilibrium prices

The drift of the second factor reflects the expectations about available production capacities in the future, trend of demand, or political or regulatory effects. The first factor, in contrast, models the differences between current value and a stochastic equilibrium level. This means that the level of mean reversion is not constant, but it depends on the time of the day, of the week, and of the year. These deviations reflect short-term effects that result, for example, from variations in weather conditions or intermittent supply shortages (see [2]).

To see the differences among the three popular financial models, we simulate electricity spot prices employing a Brownian motion, an Ornstein–Uhlenbeck process, and the Pilipovic model. Firstly, we simulate the stochastic residuals using Eqs. (3.12) and (3.13), and in a second step, to obtain spot prices, we add the seasonality shape, as derived in Sect. 3.3.3. We simulate 1,000 scenarios and we look comparatively at the probability distribution function of the simulated prices after 1-, 6- and 12-month horizon. The parameters used for our simulation are summarized in Table 3.2. Results are displayed in Fig. 3.3.

We observe that the distribution of spot electricity prices simulated by the Pilipovic model is more skewed to the right than in the case of the prices obtained with the other two model versions. This is actually the more realistic distribution, given the extremely large price movements observed in electricity prices. The difference between the Pilipovic model and the other two (Brownian motion

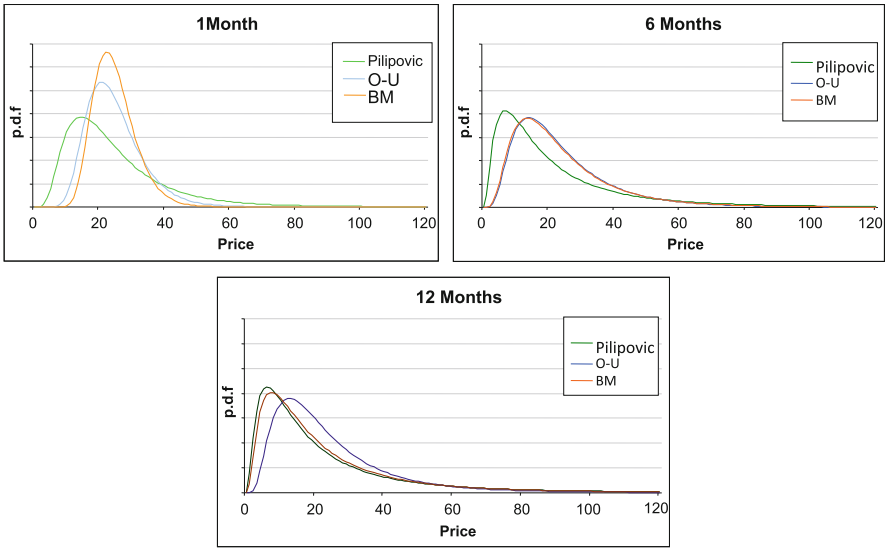

Fig. 3.3 Distributional information over scenarios generated with different financial models

and Ornstein–Uhlenbeck models) is given by the fact that the former distinguishes between the short-term and the long-term dynamics of electricity prices. Furthermore, the log spot price is assumed to mean revert towards an equilibrium level, which itself is stochastic, while in the Ornstein–Uhlenbeck process the equilibrium level is assumed constant. All three models simulate the skewed shape of the distribution of electricity prices, with an increase in the planning horizon. However, the Pilipovic model offers more realistic short-term forecasts than the other two models.

Table 3.2 Simulation parameters per annum

Geometrical Brownian motion	
σ_0	$0.9^{-1/2}$
Ornstein–Uhlenbeck process	
k_1	1.7^{-1}
μ_1	0
σ_1	$0.74^{-1/2}$
Pilipovic model	
k_2	3^{-1}
σ_2	$0.74^{-1/2}$
σ_3	$0.25^{-1/2}$
μ_2	0

3.4.3 Time Series Models

Due to their widespread use and their comprehensibility we discuss now an important class of time-series models—the family of ARMA models. This forecasting method is based on the assumption that data have an internal structure such as autocorrelation. ARMA processes enable the simulation of time dependencies within a time series and consist of two parts, the autoregressive and the moving average part. The autoregressive component considers the lagged p-price values for computing the stochastic component of the electricity price X_t in t. The moving average component takes the weighted mean of the last q error terms into account. The calculation of the price in t depends at least on a new error term ε_t, which can be, e.g., normally or Laplace distributed:

$$X_t = \sum_{i=1}^{p} \alpha_i X_{t-i} + \sum_{j=1}^{q} \beta_j \varepsilon_{t-j} + \varepsilon_t. \tag{3.14}$$

The parameters α_i describe the impact of the values X_{t-i} at the actual value X_t for all $i = 1, \ldots, p$. The parameters β_j define the weights of the error terms (innovations) ε_j within the moving average component. For an extensive discussion and applications concerning ARMA models for electricity prices see [14].

Typically ARMA models are used in time-series analysis to account for linear serial dependence. They provide the possibility to condition the mean of the process on past realizations which has often produced acceptably accurate predictions of time series in the short term. However, the assumption of autoregressive model of conditional homoscedasticity is too constricting, as electricity prices usually display volatility clusters or spikes. That is, the variance is not constant within all parts of the time series, but there are rather phases of higher and lower volatility. During the phases of high volatility, markets are often nervous and electricity prices jump and remain longer in the jump regime, implying a higher conditional probability for high price changes, when such price movements have already occurred in the recent past (see [14]).

Within the GARCH approach the assumption of homoscedasticity dropped for a heteroscedastic variance, meaning that the variance is not constant within all parts of the time series. The GARCH(p,q) process accordingly to [3, 9] reads:

$$\sigma_t^2 = \phi_0 + \sum_{z=1}^{m} \phi_{1z} \sigma_{t-z}^2 + \sum_{y=1}^{n} \phi_{2y} \varepsilon_{t-y}^2. \tag{3.15}$$

The time-variant variance σ_t^2 is driven by a constant component ϕ_0, the autoregressive part of order m and a moving average part of order n. The variance at any time t must be positive and in consequence the parameters ϕ_0, ϕ_{1z} and ϕ_{2y} can take only positive values, or equal zero at any time.

Researchers applying GARCH processes to model electricity prices assume that these can handle the heteroscedasticity caused by jumps in the case of electricity prices. However, besides seasonality and volatility clustering, an

important characteristic to be considered is the large number of extreme price changes. The spiking behavior is often described in the literature by regime-switching models [2, 14] which allow electricity prices to switch between the normal or "base" regime and the "jump regime." Another modeling approach applies EVT to model the extreme tails of the electricity prices [6]. We offer such an example in the next section.

3.5 Extreme Value Theory for Tail-Quantile Estimates

One method to deal with extremely large price movements is to delimitate the extreme tail of the distribution and to model it independently with the POT (peaks over the threshold) method, applying EVT and GPD (generalized Pareto distribution). We show the comparison of the modeling performance of an AR-GARCH model with normal or t-distributed innovations against the conditional GDP and POT method applied to tail quantiles for EEX Phelix hourly electricity prices, for the sample period August 2008–2011. Compared to typical financial assets like stocks and bonds, the magnitude of the price changes in the EEX Phelix is extreme. As seen in Fig. 3.9, the exchange electricity prices have some hours go as high as 496.26EUR/MWh and some hours go as low as -500EUR/MWh.

Since the price changes are so extreme in some hours and we focus on the extreme quantiles of the distribution, we have chosen to use simple net returns $r_t = (P_t - P_{t-1})/P_{t-1}$ instead of logarithmic returns. A similar methodology was suggested by [6]. The drawback of working with simple net returns if one is interested in large price drops is that there is a lower bound of -100%. We therefore focus in this study on the positive tail.

We model the seasonality of electricity prices in the mean equation of the AR-GARCH specification. We therefore include AR(1) and AR(24) terms in the model to account for the daily electricity prices seasonality. We end up with the following specification:

$$r_t = a_0 + a_1 r_{t-1} + a_2 r_{t-24} + \varepsilon_t,$$
$$\sigma_t^2 = \phi_0 + \phi_1 \varepsilon_{t-1}^2 + \phi_2 \sigma_{t-1}^2,$$

where σ_t^2 is the conditional variance of ε_t. ε_t is equal to $\sigma_t \eta_t$ with η_t $N(0,1)$ or Student's t-distributed i.i.d. innovations (scaled to have variance one) with mean 0, variance 1, and degree of freedom ν. The reason for including the t-distribution is that empirical evidence strongly rejects the idea that electricity price changes are normally distributed (see [6]).

We fit both versions of the AR-GARCH model to data by maximizing the likelihood function. Results are available in Table 3.3. Likelihood ratio test results show a better performance of the AR-GARCH model with t-innovations against the version with normal innovations. For both AR-GARCH models (with Gaussian and t-distributed errors) we get significant parameter estimates and "a"-parameters that are positive; a fairly high R^2 shows that the model explains successfully the data.

The sum of the "a"-parameters is not significantly lower than one; however, an infinite unconditional variance cannot be rejected for any of the two models. This is not surprising considering the extremely fat-tailed data in our study and it is further supported by the large positive tail-index estimate for the residual series and, particularly, for the original return series.

To check how much of the autocorrelation has been removed by our AR-GARCH volatility model, we look at the autocorrelation function of the standardized residuals z_t. To see how much of the heteroscedasticity has been removed by the GARCH model, we will analyze the filtered residuals ε_t. Figure 3.10 shows that an independent and identical distribution (i.i.d.) series is now approximately given for the standardized residuals of EEX Phelix price returns. The standardized residuals are meant to constitute the i.i.d. series with zero mean and unit variance that is used to estimate the tails of the sample cumulative distribution function with EVT. In Fig. 3.11 one can clearly see the heteroscedasticity in the filtered residuals. Most of the heteroscedasticity of the original data is reflected by our GARCH variance model.

We further investigate the tails of the price change distribution and estimate tail quantiles $\alpha_{t,p}$. We take standard quantiles of the normal distribution or of the t-distribution, multiply them with estimates of σ_t, and derive conditional tail quantiles $\alpha_{t,p}$ calling the mean equation of our AR-GARCH model:

$$\alpha_{t,p} = a_0 + a_1 r_{t-1} + a_2 r_{t-24} + \sigma_t \alpha_p. \tag{3.16}$$

Table 3.3 AR-GARCH parameters

AR-GARCH parameters	Normal		Student's t	
	Coeff.	Std. errors	Coeff.	Std. errors
$a_0 * 10^2$	1.5	(0.00015)	0.037	(0.00008)
a_1	0.0110	(0.0023)	−0.0012	(0.0004)
a_2	0.8640	(0.0016)	0.9817	(0.0004)
$\phi_0 * 10^4$	5.1602	(0.000005)	0.0416	(0.000007)
ϕ_1	0.5788	(0.0048)	0.4930	(0.0036)
ϕ_2	0.4212	(0.0022)	0.5068	(0.0036)
ν				2.5407
Likelihood	35'100		54'670	
R^2	0.853			
DW stat.	1.856			

Own calculations

We further model extreme tail quantiles with EVT. That is, we focus on the observations in the residuals η_t from the AR-GARCH model with normal innovations applying the POT method, following the procedure in [6]. We thus collect observations in the residual series that exceed a certain high threshold u (see [8], as cited by [6]). The excess distribution $F_u(y)$ is given by:

$$F_u(y) = P(R - u \leq y | R > u) = \frac{(F_R(u+y) - F_R(u))}{1 - F_R(u)}, \qquad 0 \leq y \leq, R_F - u \tag{3.17}$$

where y is the excesses over u and R_F is the right endpoint of F_R, which is the assumed distribution of the residuals η_t. For a high enough threshold u, [1, 20] show that for a large class of distributions F_R the excess distribution $F_u(y)$ can be approximated by the so-called GPD:

$$G_{\xi,\alpha}(y) = \left[1 - \left(1 + \frac{\xi}{\alpha} y\right)\right]^{-1/\xi}, \qquad \text{if} \quad \xi \neq 0, \tag{3.18}$$

$$G_{\xi,\alpha}(y) = 1 - e^{-y/\alpha}, \qquad \text{if} \quad \xi = 0, \tag{3.19}$$

for $0 \leq y \leq R_F - u$. ξ is the tail index and $\alpha > 0$ is just a scaling parameter. The parameters are determined by fitting the GDP to the actual data and by estimating the parameters with the maximum likelihood method. In general, the threshold u is chosen within reasonable limits of 5–13% of the data. We look at the most extreme 10% upper tail of the standardized residuals and fit the GPD to the upper tail excesses over the threshold. Thus, our upper tail "starts" at 0.9339. The maximum likelihood estimators for the GPD are given in Table 3.4. The model fit for the upper tail of residuals is shown in Fig. 3.4.

Table 3.4 Maximum likelihood estimators for the GPD parameters

Lower tails		Upper tails	
ξ	α	ξ	α
0.5913	1.7230	0.6909	1.8377

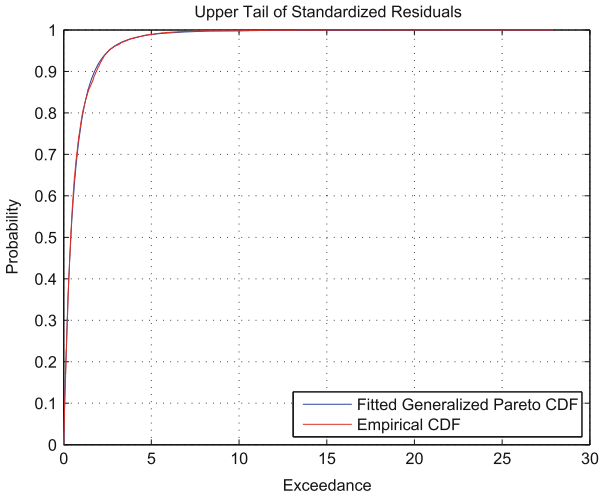

Fig. 3.4 Generalized Pareto upper tail of residuals

Accordingly to [6], the unconditional EVT tail quantiles α_p of the residual distribution with certain probabilities p is given by

$$\alpha_p = u + \frac{\alpha}{\xi}\left(\left(\frac{n}{N_u}p\right)^{-\xi} - 1\right), \tag{3.20}$$

where n is the total number of observations and N_u is the number of observations above the threshold u. Conditional GPD tail quantiles are further obtained by calculating Eq. (3.16) for α_p.

By counting the number of returns that are larger than the estimated tail quantile for each model version against the theoretical number of exceedences, we get a number that represents the accuracy of these estimates. Our estimated sample is extended over a period of 27,414h. In this case, we compute the theoretical number of exceedences of a 99% tail quantile over a time period of 27,414h as $0.01 \cdot 27,414 = 274.14$.

In the similar way we compute the theoretical exceedences of 95%, 99.5%, 99.9%, 99.95%, and 99.99% (extreme quantiles).

In Table 3.5 we present the number of exceedences of the AR-GARCH-based tail quantiles for different tail probabilities. If a particular method to calculate marginal quantiles works well, then the empirically observed number of exceedences should be close to the theoretically expected. By comparing the AR-GARCH with normal or t-innovations we conclude that the shape of the conditional error term distribution has an important impact on the tail-quantile estimates. Neither the conditional normal distribution nor the conditional t-distribution captures the behavior of the positive tail in a more realistic way. We observe that the performance of the normal model gets worse, the higher the probability for the extreme tail is chosen, while for the t-distributed model we observe a slightly better fit. Similar results where found by [6] in an analysis of Nord Pool electricity prices. The interpretation is that AR-GARCH models describe the entire distribution of returns, not only the most extreme ones. Therefore they are not very successful in capturing extremely large price movements in electricity prices. Similar discussions can be found in [14].

The number of exceedences obtained with the POT method (conditional GPD) are close to the empirical ones for all quantiles. The unconditional EVT-based risk estimator has the advantage of treating the tails separately and thus more efficiently. Extending their analysis with out-of-sample tests and price forecasts, [6] concludes a high performance of the POT method in fitting extreme tails of hourly electricity prices. This is of significant importance for risk managers in determining accurate portfolio "value-at-risk." Additionally, a realistic approach for modeling the tails helps power portfolio managers for the estimation of worst-case scenarios in the context of stress testing.

Table 3.5 In-sample evaluation of estimated tail quantiles at different probabilities (number of exceedences)

Probability	Expected	AR-GARCH	AR-GARCH-t	Conditional GPD
0.95	1,371	1,086	1,232	1,577
0.99	274	603	509	300
0.995	137	464	370	156
0.999	27	291	167	24
0.9995	14	253	116	13
0.9999	3	193	38	1

3.6 Summary

In this paper we present different modeling approaches for electricity prices. An overview of the main deseasonalization techniques is given. The price characteristics reflected by one or another model are discussed. Additionally the performance of some popular financial models is assessed in parallel: Brownian motions, Ornstein–Uhlenbeck processes, and the Pilipovic model. We reveal the importance of modeling the mean reversion and show that it makes sense to separate the short- from the long-term dynamics of electricity prices.

We further assess the performance of EVT to model extremely large price movements. This procedure gives much more realistic tail estimates than the classical time-series models. Realistic tail-quantile estimates for the electricity prices are of high interest for both risk managers and portfolio managers in the high volatile electricity markets.

In this respect, the choice of the modeling approach depends on the different research questions or planning tasks.

Appendix

See Figs. 3.5–3.11.

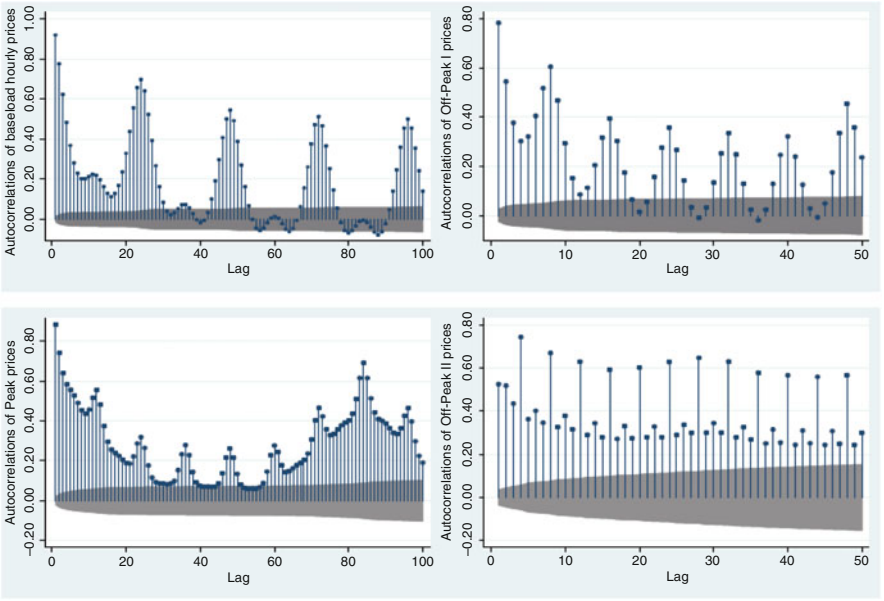

Fig. 3.5 Autocorrelations of day-ahead baseload, off-peak I, peak, and off-peak II hourly prices (source [10], p. 35)

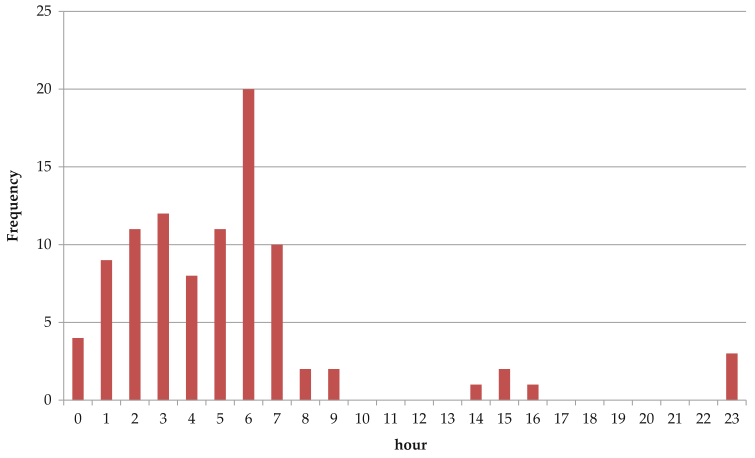

Fig. 3.6 Occurrence of negative prices Sept 2008–Dec 2011 on different hours

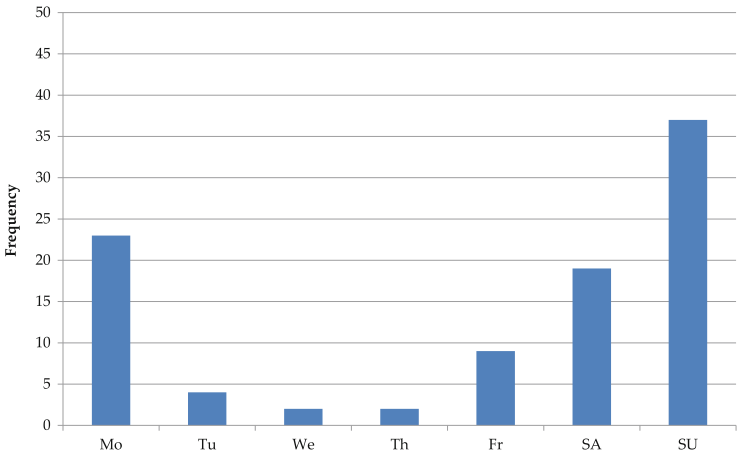

Fig. 3.7 Occurrence of negative prices Sept 2008–Dec 2011 on different days

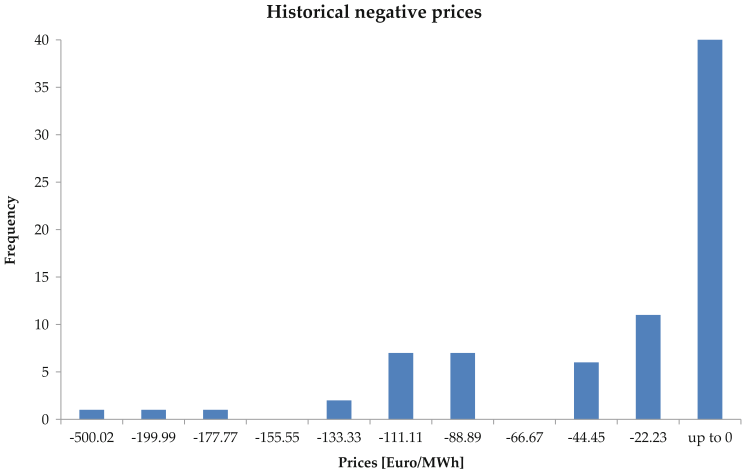

Fig. 3.8 Histogram of historical negative prices at EEX

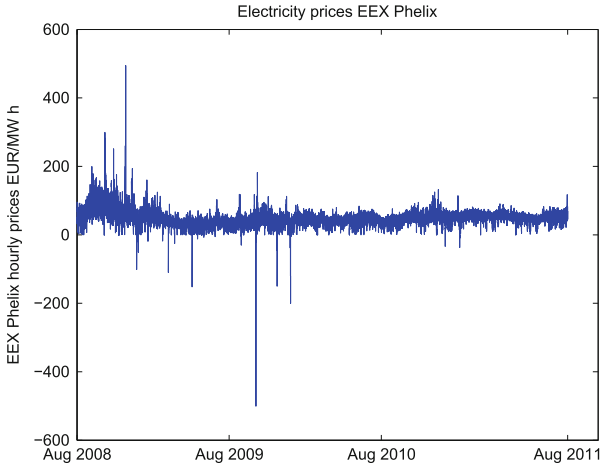

Fig. 3.9 EEX hourly prices

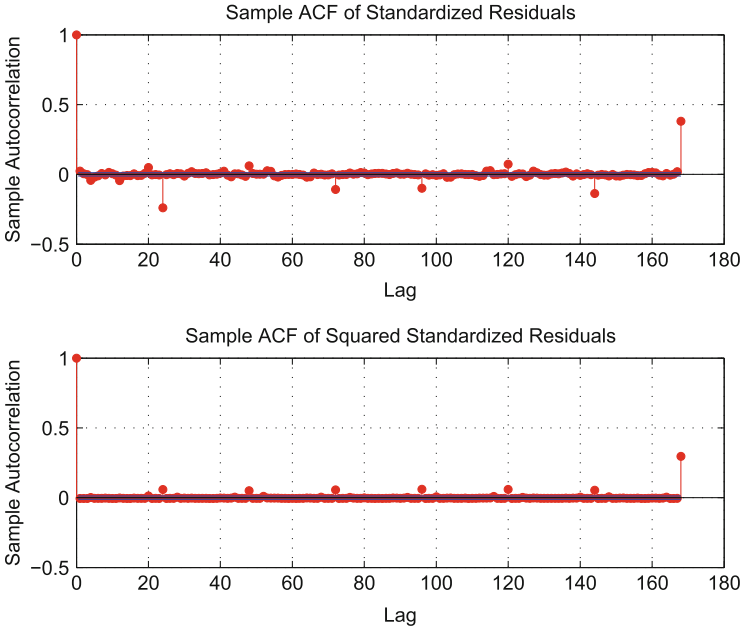

Fig. 3.10 Autocorrelation function of standardized residual

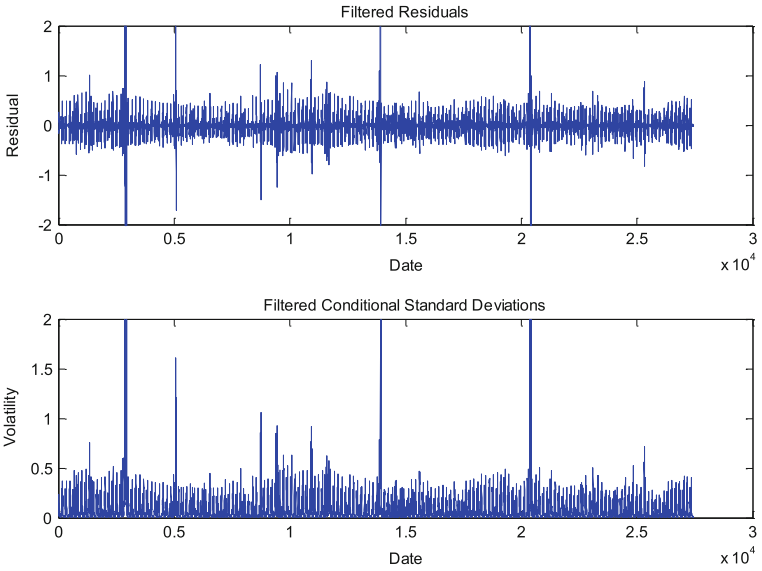

Fig. 3.11 Filtered residuals and filtered conditional standard deviation

References

1. Balkema A, de Haan L (1974) Residual life time at great age. Ann Probab 2:792–804
2. Blöchlinger L (2008) Power prices - a regime-switching spot/forward price model with Kim filter estimation, vol 3442. Dissertation, University of St. Gallen
3. Bollerslev T (1986) Generalized autoregressive conditional heteroskedasticity. J Econ 31(3):307–327
4. Borchert J, Schemm R, Korth S (2006) Stromhandel: institutionen, marktmodelle, pricing und risikomanagement. Schäffer-Poeschel, Stuttgart
5. Burger M, Klar B, Müller A, Schindlmayr G (2003) A spot market model for pricing derivatives in electricity markets. J Quant Financ 4(1):109–122
6. Byström NE (2005) Extreme value theory and extremely large electricity price changes. Int Rev Econ Financ 14:41–55
7. Cartea A, Figueroa MG (2005) Pricing electricity markets: a mean-reverting jump diffusion model with seasonality. Appl Math Financ 12(4):313–335
8. Embrechts P, Kluppelberg C, Mikosh T (1997) Modeling extremal events. Springer, Berlin
9. Engle RF (1982) Autoregressive conditional heteroschedasticity with estimates of the variance of United Kingdom Inflation. Econometrica 50(4):987–1007
10. Geninasca D (2012) Wind power feed-in-effects on prices. Master thesis, University of St. Gallen

11. Gibson R, Schwartz ES (1990) Stochastic convenience yield and the pricing of oil contingent claims. J Financ 45(3):959–976
12. Hoobs BF (2001) Linear complementarity models of Nash-Cournot competition in bilateral and POOLCO power markets. IEEE Trans Power Syst 16(2):194–202
13. Janczura J, Weron R (2010a) An empirical comparison of alternate regime-switching models for electricity spot prices. Energ Econ 32:1059–1073
14. Keles D, Genoese M, Möst D, Fichtner W (2011) Comparison of extended mean-reversion and time series models for electricity spot price simulation considering negative prices. Energ Econ 34:1012–1032
15. Knittel CR, Roberts MR (2001) An empirical examination of deregulated electricity prices. POWER working paper, University of California
16. Lise W, Lindeerhof V, Kuik O, Kemfert C, Ostling R, Heinzow T (2006) A game theoretic model of the Northwestern European electricity market - market power and the environment. Energ Policy 34:2123–2136
17. Lucia JJ, Schwartz ES (2002) Electricity prices and power derivatives: evidence from the nordic power exchange. Rev Derivatives Res 5(1):5–50
18. Manoliu M, Tompaidis S (2002) Energy futures prices: term structure models with Kalman filter estimation. Appl Math Financ 9(1):21–43
19. Nowicka-Zagrajek J, Weron R (2002) Modeling electricity loads in California: ARMA models with hyperbolic noise. HSC Research Report 02/02, Wroclaw University of Technology, Wroclaw (Poland)
20. Pickands J (1975) Statistical inference using extreme order statistics. Ann Stat 3:119–131
21. Pilipovic D (1997) Energy risk: valuing and managing energy derivatives. McGraw-Hill, New York
22. Pilipovic D (2007) Energy risk: valuing and managing energy derivatives, 2nd edn. McGraw-Hill, New York
23. Schwartz E, Smith J (2000) Short-term variations and long-term dynamics in commodity prices. Manage Sci 46(7):893–911
24. Simonsen I, Weron R, Mo B (2004) Structure and stylized facts of a deregulated power market, Working paper. http://mpra.ub.uni-muenchen.de/1443
25. Viehmann J (2011) Risk premiums in the German day-ahead electricity market. Energ Policy 39:386–394
26. Weber C (2005) Uncertainty in the electric power industry. Springer, New York
27. Weron R (2006) Modeling and forecasting electricity loads and prices: a statistical approach. Wiley, Chichester
28. Weron R, Bierbauer M, Trück S (2004) Modeling electricity prices: jump diffusion and regime switching. Physica A 336:39–48

Part II
Optimal Decisions in Managing Energy Systems

Chapter 4
Price-Driven Hydropower Dispatch Under Uncertainty

Martin Densing

Abstract After a review of hydropower optimization models, we focus on price-driven hydropower dispatch models under uncertainty of the electricity price. We present two modeling approaches for pumped-storage plants. In the first model, the water level is constrained in expectation. We discuss the marginal price of water, which is obtained analytically, and influences of price variances. The second model is a multistage stochastic linear program on a scenario tree. Financial risk is constrained by a time-consistent extension of CVaR (conditional-value-at-risk). The model has two time scales: The short-term dispatch decision is separated from the long-term planning by aggregating electricity prices into occupation times at price levels. The risk constraint is tested in a case study.

4.1 Introduction

Risk management for energy production and trading affects short- and long-term decision making. A long-term decision is how to invest in energy generation and transmission facilities. A shorter-term decision is how to operate facilities and to trade energy. The decisions are linked: The expected profit of operation and trading influences the investment decision for a facility, and the uncertainty in operation and trading increases the uncertainty of the return of the investment.

In this chapter, we focus on the operation of hydropower plants under uncertainty. We assume a liberalized market environment, where the flexible dispatch capability of hydropower is used for trading (*price-driven* operation), rather than for meeting a demand (*demand-driven* operation). Uncertainties are market prices and a nat-

M. Densing (✉)
Energy Economics Group, Paul Scherrer Institute, 5232 Villigen PSI, Switzerland
e-mail: martin.densing@psi.ch

R.M. Kovacevic et al. (eds.), *Handbook of Risk Management in Energy Production and Trading*, International Series in Operations Research & Management Science 199,
DOI 10.1007/978-1-4614-9035-7_4, © Springer Science+Business Media New York 2013

ural water inflow into a storage reservoir. The objective is to maximize expected cumulative profit under financial risk. The goal is a mid- to long-term valuation and dispatch planning for the plant by optimizing the short-term dispatch decision.

In the next section, we give an overview of price- and demand-driven hydropower models under uncertainty, and we introduce concepts for the following sections. In Sect. 4.3, we discuss price-driven hydropower models over single and multiple time periods that are exactly solvable. In Sect. 4.4, we consider a long-term planning model. This model involves two time scales by applying occupation times of electricity prices. The financial risk is measured by a multi-period extension of conditional-value-at-risk (CVaR). The mean-risk model is formulated as a multi-stage stochastic linear program on a scenario tree. We provide results of a case study, focusing on the risk constraint. Eventually, we conclude in Sect. 4.5.

4.2 Review of Stochastic Hydropower Dispatch Models

The variety of hydropower models may be classified as follows:

- **Time horizon:** Short-, medium-, or long-term models (hours to several years), single- or multi-period models.
- **Technical detail:** Models may include nonlinear water-head effects, turbine start-up and shutdown costs and delays, several turbines or interconnected reservoirs with delays of flows (or aggregated to a single power plant or single reservoir), minimal discharge or minimal generation requirements, and evaporation from the reservoir; different plant types: storage, pumped-storage, or run-of-the-river.
- **Demand or market view:** The objective may be meeting of demand or profit optimization from power trade, which may include spot, futures, or other derivatives markets; the economic agent may be a price-taker or may have market power; the auction for the market clearing price may be modeled in detail.
- **Uncertainty:** The water inflow, market prices, or the demand may be modeled stochastically.
- **Portfolio:** The model may be for a single hydropower plant or for a portfolio of power plants and contracts, which may include financial and physical delivery contracts; the electricity grid with transmission constraints may be included.
- **Objective:** Models may maximize operational profit, for example, by financial mean or mean-risk optimization; other objectives may include to minimize deviation between generation and demand, to minimize the probability of an empty reservoir or of a spill over, or to optimize irrigation.

The previous classification excludes solution methods; because hydropower optimization problems are numerically demanding, the availability of solution methods influences the range of possible models.In the following, we give a limited overview

of some of the models and methods, focusing on long-term planning, demand and market view, and on uncertainty and risk, and less on portfolio view and technical detail.

For an overview on stochastic programming in energy with emphasis on hydropower, we recommend [86].

4.2.1 Demand-Driven Hydropower Dispatch

In demand-driven models, the predominant source of uncertainty is usually the stochastic water inflow into one or several basins. The objective is to maximize hydropower generation to satisfy demand and to minimize costly alternative generation, for example, by thermal plants.

Historically, a long-term hydropower dispatch model under stochastic inflow was first considered in [58]: An optimal policy is constructed recursively backward in time; the risk of hitting low financial values and an extension to stochastic demand is shortly discussed, too. The recursive method is now known as dynamic programming (based on Bellman's principle [6]) and was applied to hydropower with stochastic inflow in [56]. Dynamic programming is still widely applied; for example, see a recent extension by the dual of the stochastic dynamic programming problem [73].

If a model has several time periods, then uncertainty is usually represented by a scenario tree. The tree is usually non-recombining because a node of the tree should represent a unique state which usually depends on the historical path. If such a tree branches in each nonterminal node, then the number of nodes grows exponentially in the number of time stages. This "curse of dimensionality" limits the number of stages in numerically tractable models. The limitation holds for dynamic programming as well as for stochastic programming, which does not require a recursive model formulation.

For demand-driven hydropower models, decomposition methods are used to solve models with moderate number of stages. The nested Benders' decomposition method for multistage stochastic programs can be suitably extended to hydropower problems [46, 60, 64]; a comparison of nested Benders' decomposition with dynamic programming for hydropower is in [2].

Nested Benders' decomposition can be combined with sampling methods: The widely applied stochastic dual dynamic programming (SDDP) algorithm combines cut sharing and sampling paths over the scenario trees; see the seminal work [65] and the overview [78]. Convergence properties of SDDP were recently addressed in [80], and a related sampling method is in [15].

An alternative to models on a general scenario tree are models on a fan of scenarios, that is, a scenario tree that branches only in the root node. Such models are numerically tractable over many stages [23, 42]; as a trade-off, a successive gain of information over time for decision making is not possible.

Demand-driven hydropower may be considered in a portfolio with other power generation. Models with additional thermal units are usually formulated as a unit commitment problem, which requires a stochastic mixed-integer formulation; for further references—also for multistage unit commitment—see [77, p. 204]. Alternatively, demand-driven hydropower can be combined with wind power [49].

Hydropower models may address the technical details: For example, the varying water level in a reservoir changes the efficiency of the turbines (head effect), and start-up costs of turbines may be relevant. The problem usually becomes nonlinear or includes integer variables; an example that includes also risk management is the run-of-the-river model [14].

Future demand is usually fairly known based on historical evidence and temperature forecast for the required heating and cooling energy. The demand may still be uncertain because of forecast errors, generator failures, or uncertain dispatch from other producers; see the model in [70], which is a short-term unit commitment problem for a chain of reservoirs, with stochastic water inflow and stochastic demand.

4.2.2 Price-Driven Hydropower Dispatch

The advent of electricity markets changed the operational environment of power producers. Without markets, dispatch decisions were mainly demand-driven; the producer of electricity had to cover domestic electricity demand by own production. With markets, demand can be satisfied by buying electricity on the market, and own surplus production can be sold on the market at the market price if we assume the market is liquid and sufficient transmission capacity is available. Empirically, market liberalization results in a different, more intense use of hydropower water resources [87].

A hydropower planning tool for the early deregulated market of Norway is discussed in [35]: Short-, medium-, and long-term planning is separated into three models with different levels of detail. Different time scales of the models allow them to cover long-term planning as well as short-term dispatch. A long-term model represents the aggregated hydropower system, a mid-term model is for each reservoir separately and incorporates trading on markets, and a short-term model, which includes technical constraints (e.g., head effects), is for the immediate dispatch on an hourly time scale. The output of the long-term model is target reservoir levels, which are input to the medium-term model. The output of the medium-term model is marginal values of water which are input to the short-term model.

More integrated dispatch models with the flexibility of a spot and futures market include, for example, the multistage stochastic programming models in [32, 36]. Modeling high-frequency dispatch (including trade) over long-time horizons is problematic because the number of stages in the scenario tree is limited by numerical tractability, and the assumption that market prices are constant or averaged between relatively widely spaced time stages may not be sufficient. A moderate number of stages is possible by a combination of dynamic programming with the

SDDP algorithm, which is originally suited only for demand-driven dispatch; for a corresponding model with trade and a demand, see [38].

Long-term market-driven hydropower models may incorporate two time scales: The dispatch (and trade) is on a short time scale (e.g., hourly), and the long-term reservoir management is on a long time scale (e.g., weekly or longer); the stages of the scenario tree are on the long time scale.

Such a *dual-scale* model is presented in [72], where the profit optimization is decoupled in an intra- and interstage problem: For each stage, a short-term problem optimizes the offer stack (i.e., the offered supply as a function of price; bidding curve), whereas the long-term optimization by dynamic programming takes the uncertainty of water inflow into account. The authors recognize that rather than the time series of prices, the *price–duration curve* (i.e., the fraction of prices below given levels) is relevant for optimizing the offer stack.

A similar dual-scale model is the stochastic multistage mixed-integer problem in [33]: On the short time scale, the chronological order of prices is not important. Hence, prices are divided into *price segments*; the reservoir balance is enforced only on a longer time scale (at the stages). Similar *load periods* of the electricity price are used in the model in [38]. The model in Sect. 4.4 uses similarly the *occupation times* of the electricity price; see also [21].

Another multi-scale long-term model is in [71]. This energy system model encompasses several energy sectors including hydro energy storage. The model includes investment decisions on a long time scale as well as dispatch decisions on an hourly scale.

After fixing a model, we may determine an optimal dispatch decision. For a general class of linear control problems, optimality is achieved by a *bang-bang* control [55]. The mid- to long-term hydropower optimization problem in [61] has optimal bang-bang dispatch decisions, which are functions based on information from forward contracts and inflow estimates. Similarly, the models of a pumped-storage hydropower plant in Sect. 4.3 have optimal bang-bang solutions.

We mention shortly other research directions for price-driven dispatch. The category of short-term models considers mainly the (one) day-ahead market; for example, see [30, 31]. Especially, the day-ahead market bidding mechanism may be modeled in detail; for example, see the day-ahead models in [17, 29]. An overview of market-oriented short-term models with emphasis on hydropower and stochastic programming methods is in [53]. Another category of models encompasses a larger portfolio of power generation units, for example, wind power [18, 83].

4.2.3 Risk-Averse Hydropower Dispatch

In this section, we give an overview of risk measures for hydropower models. For a general overview of risk management in energy, see [34].

Financial risk control is relevant mainly for mid- and long-term hydropower models, because dispatch decisions over a short time horizon may not drastically

change the cumulative profit, which is usually reported only weekly or several times per year, and the purpose of trade on spot and derivatives markets is usually assumed to reduce mid- and long-term risk, for example, by hedging the profit from own production.

Financial risk can be constrained by setting a profit target for each time period and a *penalty term* for not meeting the target. Penalty terms are used in the hydropower model in [59]; this medium-term model includes trading on a market with stochastic spot and future prices, and the contract portfolio comprises also load factor contracts, which are modeled as hydropower plants having reservoirs without inflow. Similar target values are used in the models in [36, 38]. Similarly, a penalty term for each stage is considered in [32]; the model has a stochastic inflow and includes trading with stochastic prices in spot, forward, option, and load factor contracts.

In financial portfolio theory, a widely used mean-risk model is *mean–variance* portfolio optimization. Mean–variance optimization is considered in the pumped-storage model in [62]; the short-term model includes trading in a market with price uncertainty represented by price variances.

Alternatively, financial risk can be constrained by a bound on the coherent risk measure *CVaR*; CVaR is defined in our setting in Sect. 4.4.1. A mean-risk model with CVaR for a power portfolio that includes pumped-storage hydropower plants is investigated in [23]; the long-term model encompasses several reservoirs with uncertain inflow, several markets with uncertain spot and future prices, and additional generation from thermal plants. The risk measure CVaR may also be applied in short-term models; see [14] for a technically detailed model with stochastic prices and deterministic inflow.

The risk measure CVaR fulfills the properties of coherency: For example, the risk of two combined positions does not surpass the sum of the risk of the individual positions (risk reduction by a portfolio); for details, see [3]. Conveniently, the value of CVaR can be calculated by a stochastic linear optimization problem [76]. In contrast, the risk measure value-at-risk (VaR) is not coherent and leads to non-convex optimization problems; given a probability level, the VaR of a continuous random variable is defined as the (sign-reversed) quantile at that level. For a multi-stage hydropower model with trade, the risk measure VaR is compared with CVaR and with worst-scenario values in [44]; CVaR is found conceptually and computationally superior.

CVaR is defined for a single random variable and thus suitable for single-period models; multistage models may include multi-period risk measures that take the gain of information in the scenario tree into account. Multi-period risk measures that are defined backward recursively by dynamic programming are usually *time consistent*. Time consistency ensures that if a decision maker will accept a random variable in terms of risk at an arbitrary future point in time in every scenario, then she accepts the random variable today, too; for more details, see [4].

For a hydrothermal system with stochastic demand and inflow, a time-consistent rolling-horizon policy with risk aversion is presented in [40]. In this model, the

risk of violating reservoir bounds and of unmet demand is confined by chance constraints; in a variant, the CVaR of water levels is constrained.

A storage facility for a general commodity is considered in a multi-period model in [37]; the commodity is traded on a market with random spot and forward prices, and satisfies a random demand. The objective is a recursive risk measure of the profit over time.

A multi-period risk measure is also considered in the long-term hydrothermal model in [69]. The demand-driven model encompasses several plants and a random water inflow. A mean-risk criterion is applied recursively backward from final time. The criterion is a convex combination of the mean and the CVaR of costs in each node of the scenario tree, and the optimization problem can be solved by an extension of the SDDP algorithm.

Similar to the previous models, we discuss in Sect. 4.4 a mean-risk model with a recursively defined risk measure based on CVaR.

Risk measures may also be applied to general power portfolios that encompass a variety of energy contracts and several generation capacities; hydropower is not modeled in large detail. We briefly mention some examples as follows. A portfolio consisting of market-traded contracts, of own generation, and of power-delivery contracts is considered in [50]; the risk measure is a multi-period extension of VaR. The portfolio optimization problem in [79] has a penalty term for a profit target on each stage. The portfolio problem in [43] uses the risk measure Average-VaR, that is, a sign-reversed CVaR. A mean–variance portfolio optimization is considered in [74].

A multi-period, *polyhedral* risk measure is applied in [28]; the mean-risk model is for a combined heat-and-power plant with trade on spot and future markets.

4.3 Price-Driven Dispatch Models with Exact Solutions

In the previous section, we reviewed price- and demand-driven risk-averse hydropower models. In this section, we present price-driven models with trading on a spot market. We assume a risk-neutral decision maker who maximizes expected profit without considering risk; alternatively, we assume that the probabilities of the scenarios are changed in the model to risk-neutral probability weights. Note that a risk constraint is considered in the more detailed mean-risk model in Sect. 4.4.

The decision maker of a hydropower plant has to decide how much water to release immediately and how much water to keep in the reservoir for later release. An optimal decision is given in [58, p. 63]: If the marginal utility x of immediate release is higher than the marginal utility y of the expected remaining water, then we release; if $x \leq y$, then we keep the water (*bang-bang* control). In the following models, the water level is weakly constrained in expectation, which allows us to determine the marginal utilities.

4.3.1 Single-Period Dispatch Model

We consider a *pumped-storage* hydropower plant, which has the additional dispatch flexibility of purchasing energy to store in a reservoir when prices are low; the following results can be specialized to a dispatch without pumping. The plant can sell and buy electricity in a market over a single period, for example, a month. The objective is to maximize the expected profit at the end of the period. The model can be viewed as a steady-state model (in expectation) over a series of time periods.

The electricity price during the period is assumed to be a random variable S; the plant is a price-taker. A head effect of the reservoir is neglected: We assume a constant efficiency of pumping $c \in (0,1)$ for the conversion of purchased electricity into energy (unit of reservoir level is energy). We neglect turbine start-up and shutdown constraints. The maximal capacities of production and pumping over a period are constants $u^+_{\max}$ and $u^-_{\max}$, respectively, and minimal levels are assumed to be zero. We assume that a lower reservoir is sufficiently large to allow maximal pumping over a period.

The dispatch decision how much to produce and to pump over a period is assumed to be a function of the electricity price, $S \mapsto u^\pm(S)$, with sign convention $u^\pm \geq 0$. The dispatch $u^\pm$ depends on the reservoir level only implicitly through the bounds on the reservoir level.

We assume a lower bound on the expected reservoir level. A constraint in expectation is useful in the following cases; see also [54, 62]. The bound may be a target value, which may be an empirical average over several time periods; a constraint in expectation allows the bound to be violated in some scenarios, such that the model can account for statistical estimation errors or for modeling bias. A lower bound that holds in every scenario may be too conservative, and the decision maker may take risk by violating bounds in some scenarios.

We consider only a lower bound $l_{\min}$ of the reservoir level; we assume that a violation of an upper limit has no severe consequences (apart from a spill over) or that the reservoir is sufficiently large. For the single-period model, a water inflow is not explicitly modeled: The initial value l_0 can be interpreted as a sum of the initial reservoir level and expected inflow during a period. To exclude trivial solutions, we prevent selling always at full capacity over a period, and we assume there is usable water, $u^+_{\max} > l := l_0 - l_{\min} > 0$.

The stochastic (steady-state) optimization problem maximizes the expected profit,

$$\max_{u^\pm} \mathbb{E}\Big[S u^+(S) - \frac{1}{c} S u^-(S)\Big], \qquad (P_1)$$
$$\text{s.t.} \begin{cases} l_0 - \mathbb{E}\big[u^+(S) - u^-(S)\big] \geq l_{\min}, \\ 0 \leq u^\pm(S) \leq u^\pm_{\max}, \\ u^\pm \colon \mathbb{R}_+ \to \mathbb{R}_+. \end{cases}$$

The box constraint for the random variable $u^{\pm}(S)$ in (P_1) is understood to hold almost surely. The set of scenarios is denoted by Ω. The probability of an event $A \subset \Omega$ is denoted by $\mathbb{P}[A]$, and $1_A \colon \Omega \to \{0,1\}$ is the indicator function of A. An optimal solution of problem (P_1) can be given by a bang-bang control as follows.

Proposition 1. *Let the price S be nonnegative with continuous distribution function. Then an optimal production and pumping in problem (P_1) is*

$$\hat{u}^{+}(S) = u^{+}_{\max} 1_{\{S \geq \hat{q}\}} , \qquad \hat{u}^{-}(S) = u^{-}_{\max} 1_{\{S \leq c\hat{q}\}} , \tag{4.1}$$

where $\hat{q} \geq 0$ is given by a solution of

$$u^{+}_{\max} \mathbb{P}[S \geq \hat{q}] - u^{-}_{\max} \mathbb{P}[S \leq c\hat{q}] = l. \tag{4.2}$$

For a proof, see [21].

By the optimal bang-bang control (4.1), we produce at full capacity if the electricity price is higher or equal to $\hat{q}$, and we pump at full capacity if the price is lower or equal to $c\hat{q}$, where $c \in (0,1)$ is the efficiency of pumping. A related hydropower model with bang-bang control is in [61] (see Sect. 4.2.2). In our simple model, the production threshold $\hat{q}$ is given by (4.2). The threshold $\hat{q}$ corresponds to the marginal value of water because $\hat{q}$ is the optimal Lagrange multiplier of the water constraint [21].

The marginal value $\hat{q}$ of water is related to the available water l by (4.2). A numerical solution of (4.2) shows that the marginal value decreases if the available water increases; see Fig. 4.1, and for a similar result, see [35, Fig. 3].

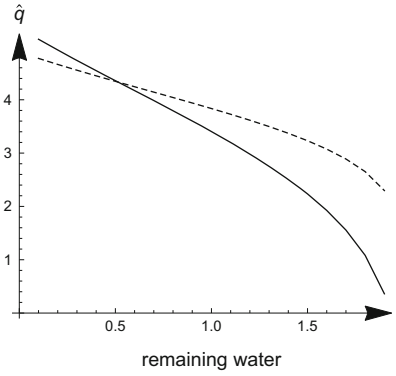

Fig. 4.1 Marginal value of water as a function of remaining water. Electricity price standard deviation = 1 (*dashed line*), = 2 (*solid line*); normally distributed

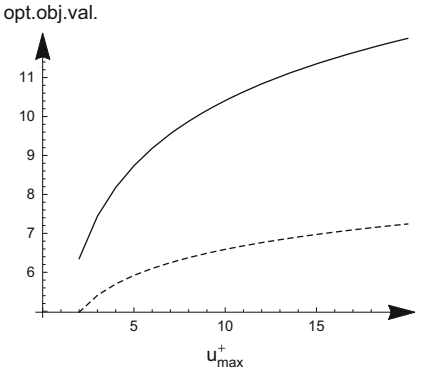

Fig. 4.2 Optimal objective value as a function of maximal production capacity. Electricity price standard deviation = 1 (*dashed line*), = 2 (*solid line*); normally distributed

Figure 4.1 shows also that a higher volatility of the electricity price results in a higher marginal value of water if water is scarce, but a lower marginal value if water is more abundant. Hence, there is a water level where the volatility has no influence on the marginal value.

The objective function of the dispatch problem (P_1) is the expected value of the operation of the plant over a period. The numerical example in Fig. 4.2 shows that the marginal expected value decreases when the capacity of production $u^+_{\max}$ increases (concave curve), and higher volatility yields higher values. Similar results are obtained for values of swing options, which can be considered as values of simplified hydropower plants [24, Fig. 3]; see also the discussion following the next problem.

Price-driven dispatch planning can be formally related to inventory theory, which considers a stochastic demand: A special case of the price-driven dispatch problem (P_1) is a dual of the newsvendor problem, which is a demand-driven inventory problem, and the stochastic price in the hydropower problem is the stochastic demand in the newsvendor problem; for details, see [21].

As a side remark, inventory theory can also be generally related to hydropower dispatch as follows. Inventory theory considers a stock of items for supplying a stochastic exogenous demand. By contrast, in hydropower, power demand is usually fairly known in advance, whereas the reservoir may have a stochastic exogenous inflow of water. If we neglect the sign of the inflow, then we consider in both cases an exogenous stochastic process that changes an inventory level. The analogy is pronounced for inventory models that allow to procure items on a market; for example, see [9, 45]. If we neglect the sign of the procurement, then inventory theory as well as price-driven hydropower considers the optimal trade on a market under a lower bound on the inventory level. The analogy depends on the provided details: If head effects are considered in hydropower, then equally sized amounts of water can have different energy content, whereas equal items in a warehouse have equal values.

4.3.2 Multi-Period Dispatch Model

The previous single-period model (P_1) can be extended without many changes to multiple periods, $t = 0, \dots, T$. The exogenous electricity spot price is now a non-negative stochastic process, $(S_t)_{t=0,\dots,T}$. The reservoir level starts again at l_0 and is bounded from below in expectation by $l_{\min}$ for all t; we assume $l_0 - l_{\min} > 0$. The dispatch decision at time t is assumed to be a function of the prices up to time t, $u^{\pm}_t(S_0, \dots, S_t)$. A water inflow with expectation $w_t \geq 0$ increases the expected reservoir level at each time step. The stochastic optimization problem for the dispatch decision is

$$\max_{(u^{\pm}_t)} \sum_{t=0}^{T} \mathbb{E}\left[S_t u^+_t(S_0, \dots, S_t) - \frac{1}{c} S_t u^-_t(S_0, \dots, S_t) \right], \tag{P_2}$$

$$\text{s.t.} \begin{cases} l_0 + \sum_{t=0}^{s} \left(\mathbb{E}\left[u_t^-(S_0,\dots,S_t) - u_t^+(S_0,\dots,S_t) \right] + w_t \right) \geq l_{\min}, \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad s = 0,\dots,T, \\ 0 \leq u_t^{\pm}(S_0,\dots,S_t) \leq u_{\max}^{\pm}, \quad t = 0,\dots,T, \\ u_t^{\pm} \colon \mathbb{R}_+^{t+1} \to \mathbb{R}_+, \qquad\qquad t = 0,\dots,T. \end{cases}$$

Similar to the single-period model (P_1), an optimal solution of (P_2) is a bang-bang dispatch as follows.

Proposition 2. *Let the distribution function of S_t be continuous for all t. Then an optimal solution of problem (P_2) is*

$$\begin{aligned} \hat{u}_t^+(S_0,\dots,S_t) &= u_{\max}^+ 1_{\{S_t \geq \sum_{s=t}^T \hat{q}_s\}}, \quad t = 0,\dots,T, \\ \hat{u}_t^-(S_0,\dots,S_t) &= u_{\max}^- 1_{\{S_t \leq c \sum_{s=t}^T \hat{q}_s\}}, \quad t = 0,\dots,T, \end{aligned} \tag{4.3}$$

where $\hat{\mathbf{q}} = (\hat{q}_0,\dots,\hat{q}_T)^\top \in \mathbb{R}^{T+1}$ and $\hat{\mathbf{v}} = (\hat{v}_0,\dots,\hat{v}_T)^\top \in \mathbb{R}^{T+1}$ are given by a solution of the following equation and complementary relation:

$$\begin{cases} \sum_{t=0}^{s} \left(u_{\max}^- \mathbb{P}\left[S_t \leq c \sum_{s'=t}^T \hat{q}_{s'} \right] - u_{\max}^+ \mathbb{P}\left[S_t \geq \sum_{s'=t}^T \hat{q}_{s'} \right] + w_t \right) + l_0 - l_{\min} = \hat{v}_s, \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad s = 0,\dots,T, \\ \hat{\mathbf{v}}^\top \hat{\mathbf{q}} = 0, \quad \hat{\mathbf{v}} \geq 0, \ \hat{\mathbf{q}} \geq 0. \end{cases} \tag{4.4}$$

For a proof, see [21].

The optimal threshold for production and pumping in (4.3) decreases over time because $\hat{q}_t \geq 0$ for all t. According to (4.3), we produce at maximal capacity at time t when $\sum_{s=t}^T \hat{q}_s$ is lower or equal the spot price; the sum can be interpreted as the marginal value of remaining water at time t. If the water inflow until a time s is sufficiently large, then $\hat{v}_s > 0$; hence, $\hat{q}_s = 0$ by (4.4), with the interpretation that the marginal value of water at time s is zero. If this holds for all times beyond s, then the amount of remaining water is sufficiently large such that full production at any nonnegative price is optimal.

Let us consider problem (P_2) in the case without pumping, $u_{\max}^- = 0$, and we neglect the intermediate constraints on water at $s = 0,\dots,T-1$. In this case, the problem is similar to the optimization problem for a buyer of a *swing option* contract: At each time step, the buyer can draw energy in the range $[0, u_{\max}^+]$, and the cumulative drawn amount is limited, too. More precisely, problem (P_2) is in this case a relaxation of the problem for a swing option, because the constraint on water (cumulative amount) holds merely in expectation. The optimal exercise policy of a swing option is under some circumstances also a bang-bang control [5]. For an overview on swing options, see [52] and Chap. 14 in this book.

Generally, formulating decisions using option theory is appropriate if the optionality in the decision model can be clearly identified, which may not be possible for more detailed hydropower models (e.g., for the following multistage mean-risk model). Indeed, the following multistage stochastic programming approach takes by its decision variables over time all implicit options automatically into account; see also [85].

4.4 Multistage Mean-Risk Model

Exact solutions as for the previous problems are not known for feature-rich hydropower models. In this section, a price-driven long-term model is formulated as a stochastic multistage linear program (MSLP) on a scenario tree; for MSLPs in general, see the monographs [10, 48]. In stochastic programming, the scenario tree is fixed and determined by realizations of the exogenous random variables (observations). By contrast, in optimal control theory, the control (decision) can in general influence the random observations. The fixed probability structure of stochastic programming facilitates numerical tractability.

We present a mean-risk model where the financial risk is measured by a multiperiod extension of the coherent risk measure CVaR. To overcome the curse of dimensionality, the model has two time scales and uses the occupation times of the electricity prices; see Sect. 4.2.2 for similar approaches. In a case study, we discuss the influence of risk on model results.

4.4.1 Time-Consistent Coherent Risk Measurement

We may define the *acceptability* of an uncertain financial position in terms of risk by a threshold on a so-called coherent risk measure [3]. Coherent risk measurement ensures for example that a portfolio is accepted in terms of risk if the individual positions are accepted, and adding an amount of cash to a portfolio reduces its risk by the same amount.

A widely applied coherent risk measure is CVaR [66, 75]. CVaR can be calculated by a stochastic linear program [76]. The dual problem is known as TailVaR [3], and similar notions (having different sign convention) are AVaR [67] and Expected Shortfall [1]. We define CVaR as a *risk-adjusted value*, which is a sign-reversed risk measure [4], as follows.

Let X be a real-valued random variable that represents an uncertain financial position. We assume a finite scenario space, $\Omega = \{\omega_1, \ldots, \omega_k\}$, $k < \infty$. The probability of scenario ω_i is p_i, $i = 1, \ldots, k$, and the realizations of X are denoted by x_i, $i = 1, \ldots, k$. Let $\alpha \in (0, 1)$. The risk-adjusted value CVaR is defined as

$$\mathrm{CVaR}^{\alpha}[X] = \min_{(q_i)} \sum_{i=1}^{k} q_i x_i \quad \text{s.t.} \left\{ q_i : 0 \le q_i \le \frac{1}{\alpha} p_i, \ \sum_{i=1}^{k} q_i = 1 \right\}. \tag{4.5}$$

The objective function of (4.5) is the expectation of X under probability weights $q_1, \ldots, q_k$. A feasible weight q_i is higher than the original weight p_i at most by a factor $1/\alpha$. The minimization shifts feasible weights to low values of X as much as allowed by the constraints. If X had a continuous distribution function, then CVaR is the expectation of X conditional on values below or equal the α-quantile of X [66].

The risk-adjusted value CVaR can be extended to multiple time periods $t = 0,\dots,T$; see the extensions in [27, 67]. Our extension is based on [4], which is suitable for a sequence of random values over time, $X_0,\dots,X_T$.

In a finite scenario space, realizations of random values over time are represented by a scenario tree. The set of nodes of the tree at time $t = 0,\dots,T$ is denoted by $\mathcal{N}_t$. The root node represents all scenarios (first stage); the terminal nodes at final time T represent single scenarios. The realization of X_t in node $n \in \mathcal{N}_t$ is denoted by x_{tn}. The transition probability p_{nm} is the probability of the set of scenarios going through node m conditional on those going through node n.

The following definition of a multi-period risk-adjusted value is based on the assumption that (i) in each scenario, the risk-adjusted value is lower or equal to the value itself, and (ii) the acceptability increases over time because information increases: The risk-adjusted value at time t is lower or equal to the expected risk-adjusted value at time $t+1$ for all feasible probability weights.

The risk-adjusted value r_{tn} at time t in node $n \in \mathcal{N}_t$ is defined recursively backward in time,

$$r_{tn} = \begin{cases} x_{tn}, & t = T, \\ x_{tn} \wedge \left(\min\limits_{(q_{nm})} \sum\limits_{m \in \mathcal{N}_{t+1}} q_{nm} r_{tm} \right), & t = 0,\dots,T-1, \end{cases} \tag{4.6}$$

$$\text{s.t.} \left\{ q_{nm} : 0 \le q_{nm} \le \frac{1}{\alpha} p_{nm}, \sum_{m \in \mathcal{N}_{t+1}} q_{nm} = 1 \right\},$$

where $y \wedge z := \min(y,z)$ for numbers y, z. Given a nonterminal node n, the set of feasible probabilities q_{nm} in (4.6) is the feasible set in (4.5) for CVaR. It can be shown that the recursive risk-adjusted value (4.6) is coherent and time consistent [4, 22].

In a mean-risk model, the acceptability of a sequence of values $X_0,\dots,X_T$ can be represented by a constraint on the risk-adjusted value (4.6) in root node n_0 at $t = 0$, $r_{0n_0} \ge \rho_{\min}$. The stringency of the risk constraint is given by $\rho_{\min}$ and the level α. Because the recursion (4.6) has a multiplicative effect, the stringency of a level α is comparable to a level α^T of a risk constraint with CVaR (4.5) [22].

4.4.2 Dual-Scale Modeling by Occupation Times

We want to model the high-frequency dispatch of a hydropower plant over a long-time horizon by a stochastic program on a scenario tree. Because the number of stages is numerically limited, we formulate the model with two time scales (*dual-scale model*). In particular, the price process on the short time scale will be converted to occupation times at price levels on the long time scale, and a subsequent principal component analysis reduces the dimensionality. The model is based on [21]; see Sect. 4.2.2 for a discussion of similar approaches [33, 38, 72].

In the dual-scale model, we assume that the *seasonal* variations of water inflow and of prices are relevant for the state of the plant, whereas the *short-term* variations are relevant for the immediate dispatch decision and do not drastically change the state.

Short time scale (e.g., hourly):

- Dispatch decision of production and pumping
- Short-term variation of prices
- Water inflow rate depending on weather

Long time scale (e.g., monthly):

- State of the plant: variations of
  - Cumulative profit
  - Water level of reservoir
- Seasonal variation of prices
- Seasonal water inflow (e.g., summer/winter)

The separation of time scales is applicable, for example, to water reservoirs that can not be emptied quickly.

The long time scale of the state variables is denoted by $t = 0, 1, \dots, T$ with time horizon T. The dispatch decision is on a short time scale $t + \frac{h}{H}$, $h = 0, 1, \dots, H-1$, $t = 0, \dots, T-1$. For simplicity, we assume that the constant H is the number of hours in $[t, t+1]$ for all t.

For a pumped-storage hydropower plant, the change in cumulative profit-and-loss between t and $t+1$ is

$$P_{t+1} - P_t = \sum_{h=1}^{H} S_{t+\frac{h}{H}} \left(U^+_{t+\frac{h}{H}} - \frac{1}{c} U^-_{t+\frac{h}{H}} \right), \tag{4.7}$$

where the random variable $S_{t+h/H}$ is the electricity spot price in the hth hour in $[t, t+1]$, the $U^{\pm}_{t+h/H}$ are the dispatch amounts of production and pumping, and c is the pumping efficiency. The dispatch should take advantage of hourly price changes, that is, the dispatch as a control function should depend on the current price, $U^{\pm}_{t+h/H} = u^{\pm}_t(S_{t+h/H}, \dots)$.

Let us generalize (4.7) to continuous time by replacing the sum by an integral. By taking into account that the dispatch depends on the current price, the right-hand side of (4.7) is for a suitable $f\colon \mathbb{R} \to \mathbb{R}$ the left-hand-side of

$$\int_t^{t+1} f(S_{t'})\, dt' = \int_0^{\infty} f(s)\, d\tilde{F}_{t+1}(s), \quad \tilde{F}_{t+1}(s) := \int_t^{t+1} 1_{\{S_{t'} \le s\}} dt', \tag{4.8}$$

where we applied in (4.8) a transformation to a Stieltjes integral with respect to $\tilde{F}_{t+1}$. A formal proof of (4.8) is in [21, Appendix], or see [11, (4.1)]. In our discrete time setting, the transformation (4.8) is an approximation

$$\sum_{h=1}^{H} f\left(S_{t+\frac{h}{H}}\right) \approx H \sum_{i=1}^{N} f(\bar{s}_i) \left(F_{t+1}(s_i) - F_{t+1}(s_{i-1}) \right), \tag{4.9}$$

where $s_0 < s_1 < \cdots < s_N$ are given price levels with intermediate levels $\bar{s}_i \in (s_{i-1}, s_i)$, $i = 1, \ldots, N$, and we define for the electricity price process in interval $[t, t+1]$ the *occupation time at level s* as

$$F_{t+1}(s) := \frac{1}{H} \sum_{h=1}^{H} 1_{\{S_{t+\frac{h}{H}} \le s\}} \left(= \begin{array}{l}\text{fraction of hours where} \\ \text{price} \le s\end{array} \right). \tag{4.10}$$

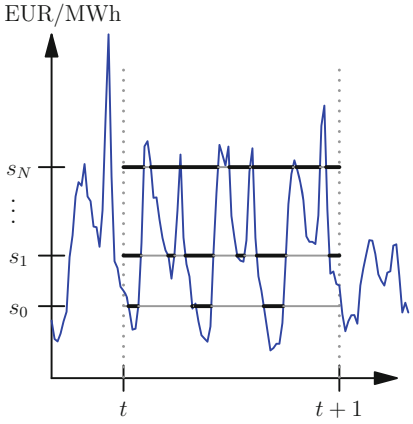

Fig. 4.3 Occupation times of the electricity price at $s_0, \ldots, s_N$ (occupation time at $s_i \triangleq$ time when price is below s_i during $[t, t+1]$). Three-day time step, EEX spot price data (Oct 2010)

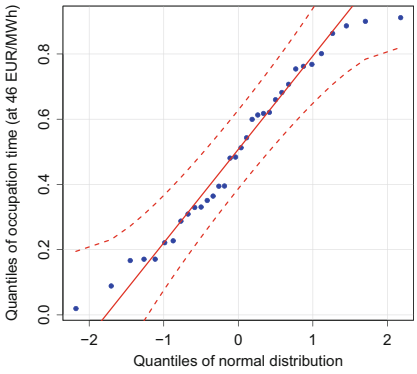

Fig. 4.4 Quantile–quantile plot of occupation time at 46 EUR/MWh. Monthly time steps; EEX spot price data, Jan 2008–Oct 2011, 46 EUR/MWh $\triangleq$ 50 % price quantile; 95 % confidence envelope

The occupation times are visualized in Fig. 4.3. Given a level s, the occupation time $F_{t+1}(s)$ is a random variable. In contrast, the so-called price–duration curve is usually deterministic, see for example [81, Sect. 4.5], and may be defined as an expected occupation time [82, (1)].

If the electricity prices over time were independent and identically distributed, then $F_{t+1}(s)$ in (4.10) is the empirical distribution function, which is asymptotically normally distributed by the central limit theorem. Empirically, the occupation times are similar to a normal distribution; see the quantile–quantile plot in Fig. 4.4.

For a dual-scale model, we apply the transformation (4.9) to the state equation (4.7) of profit and also to similar state equations of the hydropower plant, for example, of the water level. The transformation removes the short time scale of hourly price variations, and the dynamics is captured by a time series of occupation times at different price levels, $\mathbf{F}_t = (F_t(s_0), \ldots, F_t(s_N))^\top$, $t = 1, \ldots, T$. For the time series of occupation times we assume a linear principal-component factor model:

$$\mathbf{F}_t = \mu + B\mathbf{G}_t + \varepsilon_t, \quad t = 1, 2 \ldots, \tag{4.11}$$

where $\mu \in \mathbb{R}^{N+1}$ is an intercept, $\mathbf{G}_t$ is a vector of independent factors with dimension $K \ll N+1$, $B \in \mathbb{R}^{(N+1)\times K}$, and ε_t is the residual [88]. Eventually, we choose suitable models for the time series of factors $(\mathbf{G}_t)_{t=1,2,\ldots}$, for example, autoregressive processes.

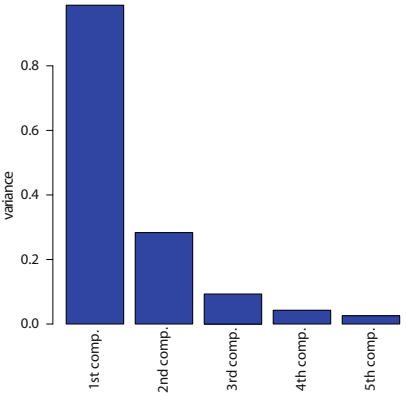

Fig. 4.5 Eigenvalues (variances) of the first 5 principal components of weekly occupation times. Nord Pool spot price data, year 2008–2010

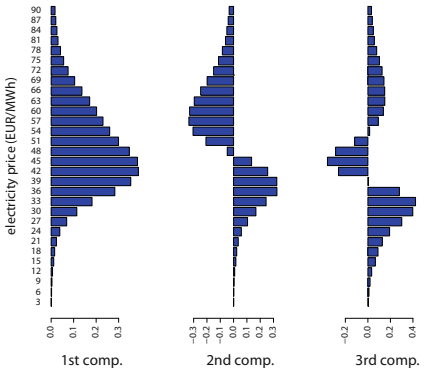

Fig. 4.6 Coefficients of the first 3 principal components of weekly occupation times. Nord Pool spot price data, year 2008–2010

The factor model (4.11) was estimated for different time periods, for different sets of spot price levels, and for the Nord Pool and EEX market places [26, 63]. For example, principal components of occupation times over weekly time steps from Nord Pool prices are in Figs. 4.5 and 4.6. A common empirical result is that the first few principal components of the occupation times explain most of the variance. Hence, we may consider $K \leq 3$ in (4.11) as sufficient, which is usually much smaller than the dimension $N+1$ of $\mathbf{F}_t$.

Note that occupation times of Ornstein–Uhlenbeck processes have a similar pattern of principal components [20]; Ornstein–Uhlenbeck processes are mean-reverting Gaussian processes which are used for electricity price models [8, 57].

4.4.3 Stochastic Program on the Scenario Tree

For the mean-risk model of a hydropower plant, we presented two modeling parts: a constraint on multi-period risk (Sect. 4.4.1) and the dual-scale modeling with occupation times (Sect. 4.4.2). We proceed with the objective function, then we give the linear programming formulation.

The objective function is assumed to be the expected final value, $\mathbb{E}[X_T]$. At time $t = 0, \ldots, T$, the retrospective part of the value X_t is the cumulative profit P_t at time t,

and the prospective part is assumed—as perceived by a decision maker—to be independent of future dispatch decisions and values the expected usable water by a salvage parameter v:

$$X_t := P_t + v\Big(\mathbb{E}\big[\sum_{s=t+1}^{T} W_s \big| \mathscr{F}_t\big] + L_t - l_{\min}\Big), \qquad t = 0,\dots,T, \tag{4.12}$$

where L_t is the reservoir level at time t, $l_{\min}$ is the lower bound on the level, and W_t is the random inflow of water into the reservoir during $[t-1,t]$. The expectation $\mathbb{E}[\cdot|\mathscr{F}_t]$ is conditional on the information which node in the scenario tree is attained at time t (σ-algebra $\mathscr{F}_t$). Discounting of values is neglected for simplicity.

The salvage value v for future water may be determined by historical estimation [59]. Alternatively, the problem is iteratively solved until numerical convergence: The marginal value of the water constraint at final time is used to update v for the next iteration [35, p. 78]. A salvage value can be avoided if the bounds on the reservoir level at final time match recurring seasonal initial conditions [23, 33, 61], or the time horizon is extended to diminish the influence of a finite horizon [40, 58]. A comparison of methods to deal with end-of-horizon effects is in [39].

The scenario tree is generated by realizations of the water inflow $(W_t)_{t=1,\dots,T}$ and of the factors $(\mathbf{G}_t)_{t=1,\dots,T}$ in the factor model (4.11) of the occupation times of the electricity prices. More precisely, the discretized innovations of the factor model and of the inflow model generate the scenario tree. The innovations are assumed to be normally distributed and are discretized by binomial distributions [47]. For simplicity, we assume that the inflow is stochastically independent from prices; if electricity supply is mainly by hydropower, then inflow and price can be correlated [32]. Inflow distributions are idiosyncratic to each reservoir; several models are presented in [84, Chap. 3]. In our model, the inflow is a sum of a deterministic time series with seasonal variation and of a normally distributed i.i.d. innovation. For more details on the scenario tree generation method, see [21]. For simplicity, we do not use additional scenario tree reduction methods; see [41] and references within.

The mean-risk model of the pumped-storage hydropower plant is formulated as a multistage stochastic linear program as follows (explanations are below):

$$\max \sum_{n\in\mathscr{N}_T} p_{n_0 n} x_{Tn} \quad \text{s.t.} \tag{P_3}$$

$$x_{tn} = p_{tn} + v\Big(l_{tn} - l_T + \sum_{s=t+1}^{T} \sum_{m\in\mathscr{N}_s} p_{nm} w_{sm}\Big),\ t = 0,\dots,T,\ \forall n \in \mathscr{N}_t,$$

$$\text{(I)} \begin{cases} p_{tn} = H \sum_{s=1}^{t} \sum_{i=1}^{N} \bar{s}_i \Big(u^{+}_{i(s-1)n} - \dfrac{1}{c} u^{-}_{i(s-1)n}\Big)\big(f_{isn} - f_{(i-1)sn}\big), \\ l_{tn} \le l_0 + H \sum_{s=1}^{t} \sum_{i=1}^{N} \Big(u^{-}_{i(s-1)n} - u^{+}_{i(s-1)n}\Big)\big(f_{isn} - f_{(i-1)sn}\big) \\ \qquad\qquad + \sum_{s=1}^{t} w_{sn},\ t = 1,\dots,T,\ \forall n \in \mathscr{N}_t, \end{cases}$$

$$\text{(II)}\quad \begin{cases} l_{\min} \le l_{tn} \le l_{\max}, & t=0,\dots,T,\ \forall n \in \mathscr{N}_t, \\ 0 \le u^{\pm}_{itn} \le u^{\pm}_{\max}, & i=1,\dots,N, \quad t=0,\dots,T-1,\ \forall n \in \mathscr{N}_t, \\ u^{+}_{itn} \le u^{+}_{(i+1)tn}, & i=1,\dots,N-1,\ t=0,\dots,T-1,\ \forall n \in \mathscr{N}_t, \\ u^{-}_{itn} \ge u^{-}_{(i+1)tn}, & i=1,\dots,N-1,\ t=0,\dots,T-1,\ \forall n \in \mathscr{N}_t, \end{cases}$$

$$\text{(III)}\quad \begin{cases} r_{0n_0} \ge \rho_{\min}, & \\ r_{tn} \le x_{tn}, & t=0,\dots,T,\ \forall n \in \mathscr{N}_t, \\ r_{tn} \le q_{tn} + \dfrac{1}{\alpha} \displaystyle\sum_{m \in \mathscr{N}_{t+1}} p_{nm} z_{(t+1)m}, & t=0,\dots,T-1,\ \forall n \in \mathscr{N}_t, \\ z_{tn} \ge 0, & t=1,\dots,T,\ \forall n \in \mathscr{N}_t, \\ z_{tn} \ge q_{(t-1)n_-} - r_{tn}, & t=1,\dots,T,\ \forall n \in \mathscr{N}_t, \end{cases}$$

Variables (time t, node $n \in \mathscr{N}_t$):

p_{tn}, l_{tn}: cumulative profit-and-loss (zero at $t=0$), water level
x_{tn}, r_{tn}: value of the plant, risk-adjusted value (lower bound)
r_{0n_0}: risk-adjusted value in root n_0 (if risk constraint is binding)
q_{tn}, z_{tn}: auxiliary variables for the risk-adjusted value
u^{+}_{itn}: production rate in spot price interval $[s_{i-1}, s_i]$
u^{-}_{itn}: pumping rate in spot price interval $[s_{i-1}, s_i]$

Scenario tree, exogenous variables (time t, node $n \in \mathscr{N}_t$):

$n \in \mathscr{N}_t$: node in set of nodes $\mathscr{N}_t$ at stage t
n_0, n_-: root node, parent node
p_{nm}: transition probability from n to node m ($=0$ if n not in history of m)
$s_0,\dots,s_N$: price levels; $\bar{s}_i \in (s_{i-1}, s_i)$ intermediate levels
f_{itn}: occupation time of spot price at level s_i during $[t-1,t]$ in node n
w_{tn}: water inflow during $[t-1,t]$ in node n
H: number of hours in time interval $[t-1,t]$

Parameters:

l_0, $l_{\min}$, $l_{\max}$: initial, minimal, and maximal allowed water level
$u^{\pm}_{\max}$: maximal production and pumping capacity
v: salvage value of remaining water
$c \in (0,1)$: efficiency of pumping
$\alpha \in (0,1)$: confidence level of the risk constraint
$\rho_{\min}$: lower bound of risk-adjusted value

Constraints (I) represent the state equations of profit and of water inflow; for example, the original equation (4.7) of cumulative profit is transformed by (4.9) into a sum of dispatch decisions over price levels $s_0,\dots,s_N$. The water level is given by an inequality because of a possible spill-over. **Constraints (II)** enforce minimal and maximal reservoir levels, and minimal and maximal technical dispatch capacities.

Production and pumping are monotonously increasing and decreasing in the electricity price level, respectively. **Constraints (III)** represent the risk constraint on the risk-adjusted value r_{0n_0} at time $t = 0$. The recursive formula (4.6) is reformulated as a set of linear constraints; for details, see [22].

The dispatch decision $u^{\pm}_{itn}$ at time t depends on the exogenous electricity price level i and on the node n in the scenario tree; the scenario tree is generated by the exogenous variables of occupation times and of water inflow. Therefore, the dispatch decision is not explicitly depending on the state of the plant at time t (i.e., the cumulative profit-and-loss and the water level). However, a node n determines a path (a set of scenarios) from time zero to t. The path determines the realizations of the exogenous variables and of the dispatch decisions up to time t. Therefore, a node determines a state.

4.4.4 Futures and Demand in a Dual-Scale Model

Electricity is dispatched in the model (P_3) only in the spot market. The set of decisions can be extended by buying and selling electricity futures contracts; see also the discussion in [86]. Usually, electricity futures exchange a fixed price against a floating price; hence, they are similar to financial swaps [7].

In our multistage model, the floating price is the spot price during consecutive time stages, and the futures are cash settled: The payoff is the hourly difference between the fixed futures price and the spot price. Comparable futures at the EEX market place are so-called base-month futures, which are monthly futures on the hourly spot price [25].

We assume that the futures are used for static hedging, such that the position amount p_t (MW) of futures for time period $[t, t+1]$, $t = 0, \ldots, T-1$, is fixed at time zero. The profit-and-loss of the futures (receive fix, play float) from t to $t+1$ is

$$P^{\text{fut}}_{t+1} - P^{\text{fut}}_{t} = p_t \sum_{h=1}^{H} \left(y_t - S_{t+\frac{h}{H}} \right), \tag{4.13}$$

where y_t is the fixed futures price (EUR/MWh) contracted at time zero, and H is the number of hours in a time period; we neglect discounting and some details of the cash flow exchange, for example, the so-called margin calls.

Applying to (4.13) the transformation (4.9) into occupation times at price levels, we get for the profit-and-loss of futures:

$$P^{\text{fut}}_{t+1} - P^{\text{fut}}_{t} = p_t H \left(y_t - \sum_{i=1}^{N} \bar{s}_i \left(F_{t+1}(s_i) - F_{t+1}(s_{i-1}) \right) \right), \tag{4.14}$$

which can be added to the profit-and-loss of model (P_3).

The model (P_3) can also be extended by a demand of electricity. Let the demand be given by an exogenous hourly stochastic process $(D_{t+\frac{h}{H}})$, $h = 0, \ldots, H-1$,

$t = 0, \dots, T-1$. We assume that the demand can be supplied from a liquid market with sufficient transmission capacity. In this case, the demand is merely an additional profit-and-loss term in the objective function. The profit-and-loss from t to $t+1$ is

$$P_{t+1}^{\text{dem}} - P_t^{\text{dem}} = \sum_{h=1}^{H} D_{t+\frac{h}{H}} \left(d_t - S_{t+\frac{h}{H}} \right), \tag{4.15}$$

where d_t is the fixed retail selling price in period $[t, t+1]$; we neglect discounting. Empirically, demand may be correlated with the electricity spot price; a possible dependency is $D_t = f(S_t)$, where the function $f\colon \mathbb{R} \to \mathbb{R}$ is empirically estimated [12]. Applying to (4.15) the transformation (4.9) into occupation times at price levels, we get

$$P_{t+1}^{\text{dem}} - P_t^{\text{dem}} = H \sum_{i=1}^{N} f(\bar{s}_i)(d_t - \bar{s}_i) \Big(F_{t+1}(s_i) - F_{t+1}(s_{i-1}) \Big).$$

Alternatively, we may use the same approach as for the electricity price in (4.9) as follows. We consider levels of demand, $d_0 < d_1 < \cdots < d_M$, and the approximation

$$D_{t+\frac{h}{H}} \approx \tilde{D}_{t+\frac{h}{H}} := \sum_{i=1}^{M} \bar{d}_i 1_{\{d_{i-1} < D_{t+\frac{h}{H}} \le d_i\}}, \tag{4.16}$$

with intermediate levels $\bar{d}_i \in (d_{i-1}, d_i)$. Substituting (4.16) in (4.15) and using (4.9), we get after some calculation

$$\begin{aligned} P_{t+1}^{\text{dem}} - P_t^{\text{dem}} = H \sum_{i,j=1}^{N,M} \bar{d}_j (d_t - \bar{s}_i) \Big(& F_{t+1}(s_{i+1}, d_{j+1}) - F_{t+1}(s_{i+1}, d_j) \\ & - F_{t+1}(s_i, d_{j+1}) + F_{t+1}(s_i, d_j) \Big), \end{aligned}$$

where

$$F_{t+1}(s,d) := \frac{1}{H} \sum_{h=1}^{H} 1_{\{S_{t+\frac{h}{H}} \le s,\ D_{t+\frac{h}{H}} \le d\}}$$

is the joint *price-demand occupation time* at levels of price s and demand d during interval $[t, t+1]$. The principal component factor model (4.11) for the occupation times of price may be extended to the joint occupation times.

4.4.5 Case Study: Risk Constraint and Futures

In this case study, we consider a pumped-storage plant in Switzerland. The water reservoir is sized such that it can be emptied in approximately one month by full production with no pumping and no natural water inflow; the reservoir is chosen initially nearly full (95 %). The pumping capacity is 25 % of the production capacity.

Electricity spot prices are from the EEX market place (year 2008–2010) [26]. For more details on parameters, see [19].

The mean-risk model (P_3) is solved over four time periods of one month ($T = 4$). As discussed in Sect. 4.4.3, the scenario tree is generated by monthly innovations of occupation times of electricity prices and of water inflows. By numerical experiments, a sufficient discretization of the innovations is 4, 2, and 2 for the first factor in the principal-component model (4.11) of occupation times, for the second factor, and for the inflow, respectively. Hence, the tree branches in every nonterminal node into $4 \times 2 \times 2 = 16$ child nodes (we used in the case study even $4 \times 1 \times 2$ for reasons of speed). A 2-point distribution for inflow is sufficient because the reservoir buffers the water; influences of higher moments of the inflow distribution on the operability of the plant are small.

Futures are added to the model (P_3) by the additional profit-and-loss (4.14). For each stage, a different hedging position in futures can be entered at initial time. The fixed side of the futures is chosen to be the same for all maturities and equals the averaged historical spot price. In the next section, some of the base assumptions are temporarily altered (e.g., futures are not allowed).

The model with four time steps leads to relatively small problem sizes. The problems are solved by the commercial CPLEX simplex solver on a personal computer (1 GB RAM) in several minutes.

4.4.5.1 Mean-Risk Frontier

The mean-risk frontier of model (P_3) is the optimal objective value as a function of the bound $\rho_{\min}$ of the risk constraint. The function is continuous piecewise linear and concave because the lower bound $\rho_{\min}$ is on the right-hand-side of a linear maximization problem; see for example [16, Theorem 6.6]. Figure 4.7 shows the mean-risk frontier in four cases: with futures, without futures, with stochastic inflow, and with deterministic inflow. If we can trade electricity only on the spot market, then the possibility to reduce risk is limited; either the risk constraint is relaxed (flat curve) or the model becomes quickly infeasible. Given a level of risk, futures enable higher means (expected value of the plant), and some further risk reduction is possible; see Fig. 4.7. Even with futures the risk reduction is limited because the only underlying asset is electricity. In contrast, a portfolio in finance depends on several underlyings, and one of them is usually an (uncorrelated) risk-free asset.

Contrary to risk reduction, Fig. 4.7 also shows that futures enable high expected values if we continuously relax the risk constraint. Correspondingly, positions in futures were growing as observed in the numerical experiments. The ability of futures to increase or to decrease risk is also discussed in [86, Sect. 3.5].

Furthermore, Fig. 4.7 shows the effect of a stochastic water inflow: If we assume a deterministic inflow, then the mean-risk frontiers improve considerably.

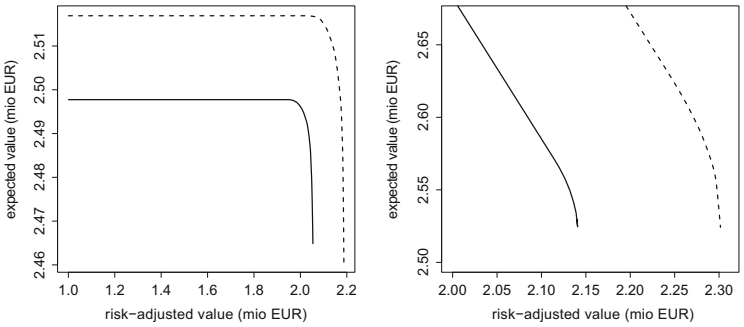

Fig. 4.7 Mean-risk frontier. Vertical axis: expected final value of the hydropower plant; horizontal axis: risk-adjusted value. Deterministic inflow (*dashed line*); stochastic inflow (*solid line*). *Left* without futures. *Right* with futures

4.4.5.2 Optimal Dispatch Decision

Figure 4.8 shows the optimal dispatch (decision rule) of the mean-risk hydropower model (P_3): We pump and produce either at zero or at full capacity, while intermediate pumping and production is restricted to a narrow band of adjacent electricity price levels, which points to a price threshold at a level in between. We have already encountered such bang-bang controls for the models (P_1) and (P_2) (Sect. 4.3).

Furthermore, Fig. 4.8 shows that if the risk constraint becomes tighter, then the threshold for pumping is lowered, and the threshold for production is elevated. The decision to pump less under risk at initial time $t = 0$ depends on the model parameters of water inflow and initial water level; for example, for a model without natural water inflow, the optimal decision may be to initially pump more under risk [62, Fig. 6].

4.4.5.3 Optimal Profit-and-Loss

To test the effect of multi-period risk measurement, we replaced in the mean-risk model (P_3) the multi-period risk measure by the single-period risk measure CVaR (4.5); the risk constraint bounds the CVaR of the values of the plant at final stage. Figure 4.9 shows corresponding distributions of profit-and-loss at the time stages: The multi-period risk constraint can enforce higher intermediate profits, while not deteriorating profits at final stage. Similar results for multi-period risk constraints are in [28, 37].

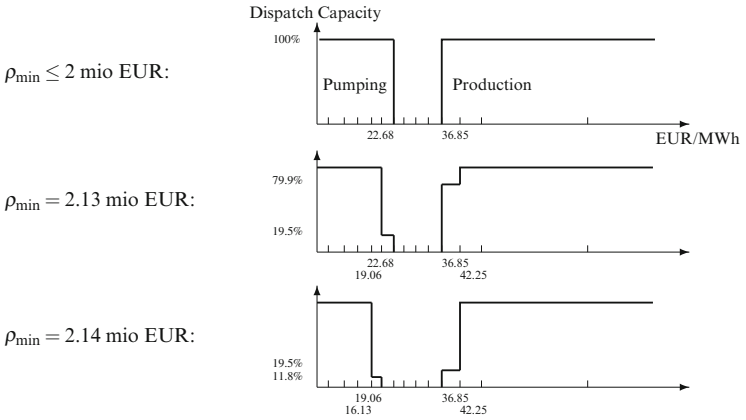

Fig. 4.8 Optimal decision rule for pumping and production at $t = 0$ (root node of scenario tree). *Top* loose bound on risk. *Center* medium bound on risk. *Bottom* tight bound on risk

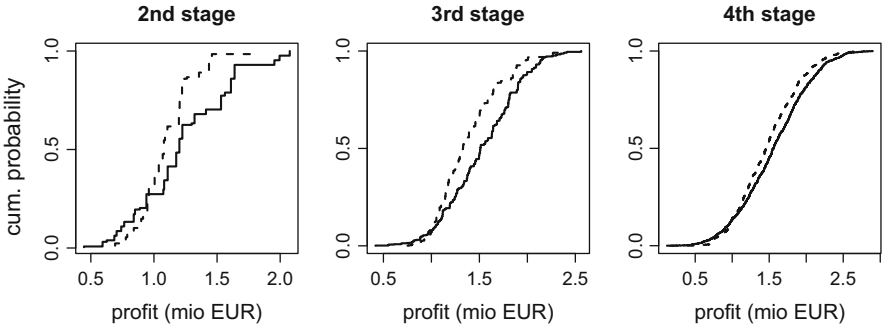

Fig. 4.9 Distribution function of cumulative profit at 2nd, 3rd, and 4th (final) stage. CVaR risk constraint on final values (*dashed line*); multi-period risk constraint (*solid line*)

4.4.5.4 Futures in Dependence of the Risk Constraint

So far, the price of the fixed side of the futures was assumed to be the average of historical spot prices. At what price should the decision maker of the hydropower plant purchase futures positions? The decision for which price, how many, and which kind of new assets to purchase depends on the randomness in the already existing portfolio of assets and on the attitude of the decision maker against risk. If the new asset (minus the purchase price) can increase the acceptability of the portfolio in terms of profit and of risk, then the decision maker should buy [24].

Figure 4.10 shows for the mean-risk model (P_3) [with futures (4.14)] the optimal futures amounts (receive fix, pay float) in dependence on price; for simplicity, we enforced in this example equal amounts over the stages. Obviously, optimal amounts decline when the fixed price, which we receive, is lowered. The lowest acceptable price depends on the risk constraint: If the risk constraint is moderate, then the minimal accepted price is in our example approximately 48 EUR/MWh (below this

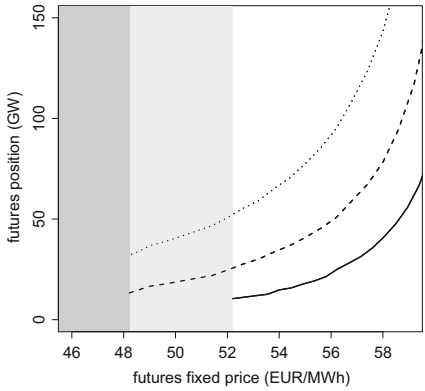

Fig. 4.10 Optimal positions of futures as a function of receiving price. Tight bound on risk (*solid line*); medium and relaxed bound (*dashed and dotted line*). Tight bound: prices below approx. 52 EUR/MWh cause model infeasibility. Medium and relaxed bound: prices below approx. 48 EUR/MWh yield negative positions

threshold, amounts are negative); if the risk constraint is tight, then the amounts of futures are smaller, and the minimal accepted price is higher at 52 EUR/MWh (below this threshold, the mean-risk model is infeasible).

The method to value assets by an acceptance criterion can be generalized: We may evaluate the possible range of prices for purchasing a whole hydropower plant by an acceptance criterion; for example, the criterion may be a threshold on the coherent risk measure CVaR [24]. An accepted price can be considered as the (subjective) value of the plant.

A related notion is the indifference price: Given a position amount of a contract, the indifference price is the minimal price for which the contract is accepted; see [13, 51] and Chap. 14 in this book.

4.4.5.5 Wait-and-See and Scenario-Independent Solutions

In model (P_3), the dispatch decisions $u^{\pm}_{itn}$ at time t depend on the price level s_i and on the node $n \in \mathcal{N}_t$ at time t in the scenario tree. The decision is *non-anticipative* because it depends on a node at current time, that is, it depends on past and on present information (*here-and-now* solution).

In contrast, an *anticipative* decision at time $t < T$ is $\tilde{u}^{\pm}_{itm}$, where the node $m \in \mathcal{N}_T$ is at final time T (perfect foresight, *wait-and-see* solution). The difference in objective value between the wait-and-see and here-and-now solution is the *expected-value-of-perfect-information* (EVPI) [48]. In model (P_3) on the scenario tree, the EVPI is the value a decision maker gains in return for knowledge of the future values of water inflow and of the occupation times of spot prices.

To calculate the EVPI of our model, we assume that positions in futures are not allowed; if the future is known, futures give unlimited profit (unless positions are bounded). For simplicity, we relax also the risk constraint. The resulting EVPI is strongly influenced by the constraints on the state, for example, the water level: If we introduce an additional constraint that forces the reservoir at the time horizon to be nearly full (initial level), then the EVPI relative to the here-and-now objective

value is 80 %, whereas it is merely 7 % with no additional constraint (the reservoir can be emptied).

The opposite of a wait-and-see solution is a *scenario-independent* solution, where a feasible decision at time $t > 0$ does not depend on the node: $u^{\pm}_{itn} = u^{\pm}_{itm}$, for all nodes $n, m \in \mathcal{N}_t$. Scenario-independent decisions (as well as the foregoing wait-and-see decisions) allow a simpler, path-wise problem formulation on a fan of scenarios; for example, see [23].

We test the effect of a scenario-independent solution first under a relaxed risk constraint. If the reservoir can be emptied, then the here-and-now optimal objective value is approximately 10 % higher than with scenario-independent decisions. In contrast, if the reservoir is forced at the time horizon to be nearly full (initial level), then the optimal objective value is approximately 40 times higher. Tightening bounds of risk aggravates the differences: There are levels of the risk constraint where the model on the full scenario tree is still feasible, whereas the model on the fan is already infeasible. Hence, flexibility of decisions is important if constraints are restrictive; see also the discussion of stochastic versus deterministic hydropower modeling in [68].

4.5 Conclusion

After an extensive review of stochastic hydropower optimization, we considered price-driven dispatch models for a pumped-storage hydropower plant, that is, the operation of the plant is optimized against an electricity market. We obtained exact solutions for single- and for multi-period problems that have constraints on the water level in expectation. The marginal value of water is given by relatively simple, implicit equations.

A second price-driven dispatch model was formulated as a multistage stochastic program on a scenario tree. The model is a mean-risk optimization problem, where the risk constraint is given by a time-consistent extension of the coherent risk measure CVaR. Modeling on two time scales (dual-scale model) allows to obtain numerical solutions for a moderate number of time stages: The short time scale for trading is removed by using the occupation times of the electricity price at different price levels. Subsequently, a principal component analysis of the occupation times allows to reduce the dimension of the scenario tree. The dual-scale approach was tested with empirical data from spot markets. We evaluated the influence of the risk constraint on the optimal dispatch and the mean-risk frontier.

The dual-scale model allows to incorporate electricity futures and evaluate their optimal contract sizes; an electricity demand can in principle be incorporated, too.

Hydropower optimization is explored by several research communities; we highlighted the relation to the bang-bang solutions of optimal control theory, and pointed to similarities with swing options and inventory theory.

Acknowledgements I thank H. Turton for valuable suggestions on a draft version.

References

1. Acerbi C, Tasche D (2002) On the coherence of expected shortfall. J Bank Financ 26(7):1487–1503. doi:10.1016/S0378-4266(02)00283-2
2. Archibald T, Buchanan C, McKinnon K, Thomas L (1999) Nested Benders decomposition and dynamic programming for reservoir optimisation. J Oper Res Soc 50(5):468–479. doi:10.1057/palgrave.jors.2600727
3. Artzner P, Delbaen F, Eber JM, Heath D (1999) Coherent measures of risk. Math Financ 9(3):203–228. doi:10.1111/1467-9965.00068
4. Artzner P, Delbaen F, Eber JM, Heath D, Ku H (2007) Coherent multiperiod risk adjusted values and Bellman's principle. Ann Oper Res 152:5–22. doi:10.1007/s10479-006-0132-6
5. Barrera-Esteve C, Bergeret F, Dossal C, Gobet E, Meziou A, Munos R, Reboul-Salze D (2006) Numerical methods for the pricing of swing options: A stochastic control approach. Methodol Comput Appl Probab 8(4):517–540. doi:10.1007/s11009-006-0427-8
6. Bellman R (1954) Some applications of the theory of dynamic programming—a review. Oper Res 2(3):275–288. doi:10.1287/opre.2.3.275
7. Benth FE, Koekebakker S (2008) Stochastic modeling of financial electricity contracts. Energ Econ 30(3):1116–1157. doi:10.1016/j.eneco.2007.06.005
8. Benth FE, Kiesel R, Nazarova A (2012) A critical empirical study of three electricity spot price models. Energ Econ 34(5):1589–1616. doi:10.1016/j.eneco.2011.11.012
9. Berling P, de Albéniz VM (2011) Optimal inventory policies when purchase price and demand are stochastic. Oper Res 59(1):109–124. doi:10.1287/opre.1100.0862
10. Birge JR, Louveaux F (1997) Introduction to stochastic programming. In: Springer series in operations research and financial engineering. Springer, New York. doi:10.1007/b97617
11. Borodin AN, Ibragimov I (1995) Limit theorems for functionals of random walks. In: Proceedings of the Steklov institute of mathematics, vol 195. American Mathematical Society
12. Burger M, Klar B, Müller A, Schindlmayr G (2004) A spot market model for pricing derivatives in electricity markets. Quant Financ 4(1):109–122. doi:10.1088/1469-7688/4/1/010
13. Carmona R (ed) (2009) Indifference pricing: theory and applications. In: Princeton series in financial engineering. Princeton University Press, Princeton
14. Catalão J, Pousinho H, Contreras J (2012) Optimal hydro scheduling and offering strategies considering price uncertainty and risk management. Energy 37(1):237–244. doi:10.1016/j.energy.2011.11.041
15. Dantzig GB, Infanger G (1997) Intelligent control and optimization under uncertainty with application to hydro power. Eur J Oper Res 97(2):396–407. doi:10.1016/S0377-2217(96)00206-8

16. Dantzig GB, Thapa MN (2003) Linear programming – 2: theory and extensions. In: Springer series in operations research and financial engineering. Springer, New York. doi:10.1007/b97283
17. De Ladurantaye D, Gendreau M, Potvin JY (2007) Strategic bidding for price-taker hydroelectricity producers. IEEE Trans Power Syst 22(4), 2187–2203. doi:10.1109/TPWRS.2007.907457
18. Deane JP, McKeogh EJ, Gallachóir BPÓ (2013) Derivation of intertemporal targets for large pumped hydro energy storage with stochastic optimization. IEEE Trans Power Syst. doi:10.1109/TPWRS.2012.2236111
19. Densing M (2007) Hydro-electric power plant dispatch-planning – multi-stage stochastic programming with time-consistent constraints on risk. Ph.D. thesis, ETH, Zürich. doi:10.3929/ethz-a-005464814. Nr. 17244
20. Densing M (2012) Occupation times of the Ornstein–Uhlenbeck process: Functional PCA and evidence from electricity prices. Physica A 391(23):5818–5826. doi:10.1016/j.physa.2012.07.040
21. Densing M (2013) Dispatch planning using newsvendor dual problems and occupation times: Application to hydropower. Eur J Oper Res 228(2):321–330. doi:10.1016/j.ejor.2013.01.033
22. Densing M, Mayer J (2012) Multiperiod stochastic optimization problems with time-consistent risk constraints. In: Klatte D, Lüthi HJ, Schmedders K (eds) Operations research proceedings 2011, Operations research proceedings. Springer, Berlin, pp 521–526. doi:10.1007/978-3-642-29210-1-83
23. Doege J, Schiltknecht P, Lüthi HJ (2006) Risk management of power portfolios and valuation of flexibility. OR Spectrum 28(2):267–287. doi:10.1007/s00291-005-0005-4
24. Doege J, Fehr M, Hinz J, Lüthi HJ, Wilhelm M (2009) Risk management in power markets: the hedging value of production flexibility. Eur J Oper Res 199(3):936–943. doi:10.1016/j.ejor.2009.01.067
25. EEX: European Energy Exchange (2011) Connecting markets - power, natural gas, emissions, coal, Leipzig
26. EEX: European Energy Exchange (2011) Historical market data – power spot market. www.eex.com
27. Eichhorn A, Römisch W (2005) Polyhedral risk measures in stochastic programming. SIAM J Optimiz 16(1):69–95. doi:10.1137/040605217
28. Eichhorn A, Heitsch H, Römisch W (2010) Stochastic optimization of electricity portfolios: Scenario tree modeling and risk management. In: Rebennack S, Pardalos PM, Pereira MVF, Iliadis NA (eds) Handbook of power systems II, Energy systems. Springer, Berlin, pp 405–432. doi:10.1007/978-3-642-12686-4
29. Faria E, Fleten SE (2011) Day-ahead market bidding for a Nordic hydropower producer: taking the Elbas market into account. Comput Manage Sci 8:75–101. doi:10.1007/s10287-009-0108-5
30. Fleten SE, Kristoffersen TK (2007) Stochastic programming for optimizing bidding strategies of a Nordic hydropower producer. Eur J Oper Res 181(2):916–928. doi:10.1016/j.ejor.2006.08.023

31. Fleten SE, Kristoffersen TK (2008) Short-term hydropower production planning by stochastic programming. Comput Oper Res 35(8):2656–2671. doi:10.1016/j.cor.2006.12.022
32. Fleten SE, Wallace SW, Ziemba WT (2002) Hedging electricity portfolios via stochastic programming. In: Greengard C, Ruszczyński A (eds) Decision making under uncertainty: energy and power. The IMA volumes in mathematics and its applications, vol 128. Springer, New York, pp 71–94. doi:10.1007/978-1-4684-9256-9
33. Fleten SE, Haugstvedt D, Steinsbø JA, Belsnes M, Fleischmann F (2011) Bidding hydropower generation: integrating short- and long-term scheduling. In: 17th power system computation conference. http://mpra.ub.uni-muenchen.de/id/eprint/44450
34. Fleten SE, Keppo J, Näsäkkälä E (2012) Risk management in electric utilities. In: Kouvelis P, Dong L, Boyabatli O, Li R (eds) Handbook of integrated risk management in global supply chains, Chap 18, 1st edn. Wiley, Hoboken, pp 495–513. doi:10.1002/9781118115800.ch18
35. Fosso OB, Gjelsvik A, Haugstad A, Mo B, Wangensteen I (1999) Generation scheduling in a deregulated system - the Norwegian case. IEEE Trans Power Syst 14(1):75–81. doi:10.1109/59.744487
36. Frauendorfer K, Güssow J (2002) Stochastic multistage programming in the operation and management of a power system. In: Marti K (ed) Stochastic optimization techniques – numerical methods and technical applications. Lecture notes in economics and mathematical system, vol 513. Springer, Berlin, pp 199–222
37. Geman H, Ohana S (2008) Time-consistency in managing a commodity portfolio: a dynamic risk measure approach. J Bank Financ 32(10):1991–2005. doi:10.1016/j.jbankfin.2007.05.020
38. Gjelsvik A, Mo B, Haugstad A (2010) Long- and medium-term operations planning and stochastic modelling in hydro-dominated power systems based on stochastic dual dynamic programming. In: Rebennack S, Pardalos PM, Pereira MVF, Iliadis NA (eds) Handbook of power systems I, Energy systems. Springer, Berlin, pp 33–55. doi:10.1007/978-3-642-02493-1
39. Grinold RC (1983) Model building techniques for the correction of end effects in multistage convex programs. Oper Res 31(3):407–431. doi:10.1287/opre.31.3.407
40. Guigues V, Sagastizábal C (2012) The value of rolling-horizon policies for risk-averse hydro-thermal planning. Eur J Oper Res 217(1):129–140. doi:10.1016/j.ejor.2011.08.017
41. Heitsch H, Römisch W (2009) Scenario tree modeling for multistage stochastic programs. Math Program 118(2):371–406. doi:10.1007/s10107-007-0197-2
42. Helseth A, Mo B, Warland G (2010) Long-term scheduling of hydro-thermal power systems using scenario fans. Energ Syst 1:377–391. doi:10.1007/s12667-010-0020-7

43. Hochreiter R, Wozabal D (2010) A multi-stage stochastic programming model for managing risk optimal electricity portfolios. In: Rebennack S, Pardalos PM, Pereira MVF, Iliadis NA (eds) Handbook of power systems II, Energy systems. Springer, Berlin, pp 383–404. doi:10.1007/978-3-642-12686-4
44. Iliadis NA, Pereira MVF, Granville S, Chabar R, Barroso LA (2007) Portfolio optimization of hydroelectric assets subject to financial indicators. In: 2007 IEEE power engineering society general meeting. Institute of Electrical and Electronics Engineers, pp 1–8. doi:10.1109/PES.2007.385726
45. Inderfurth K, Kelle P, Kleber R (2013) Dual sourcing using capacity reservation and spot market: optimal procurement policy and heuristic parameter determination. Eur J Oper Res 225(2):298–309. doi:10.1016/j.ejor.2012.08.025
46. Jacobs J, Freeman G, Grygier J, Morton D, Schultz G, Staschus K, Stedinger J (1995) SOCRATES: a system for scheduling hydroelectric generation under uncertainty. Ann Oper Res 59(1):99–133. doi:10.1007/BF02031745
47. Jamshidian F, Zhu Y (1997) Scenario simulation: theory and methodology. Financ Stochast 1(1):43–67. doi:10.1007/s007800050016
48. Kall P, Mayer J (2011) Stochastic linear programming: models, theory, and computation. In: International series in operations research & management science, vol 156, 2nd edn. Springer, New York. doi:10.1007/978-1-4419-7729-8
49. Kim JH, Powell WB (2011) Optimal energy commitments with storage and intermittent supply. Oper Res 59(6):1347–1360. doi:10.1287/opre.1110.0971
50. Kleindorfer PR, Li L (2005) Multi-period VaR-constrained portfolio optimization with application to the electric power sector. Energ J 26(1):1–26. doi:10.5547/ISSN0195-6574-EJ-Vol26-No1-1
51. Kovacevic RM, Paraschiv F (2013) Medium-term planning for thermal electricity production. OR Spectrum. doi:10.1007/s00291-013-0340-9
52. Kovacevic RM, Pflug GC (2013) Electricity swing option pricing by stochastic bilevel optimization: a survey and new approaches. Stochastic programming e-print series (SPES). http://nbn-resolving.de/urn:nbn:de:kobv:11-100209377
53. Kristoffersen TK, Fleten SE (2010) Stochastic programming models for short-term power generation scheduling and bidding. In: Bjørndal E, Bjørndal M, Pardalos PM, Rönnqvist M (eds) Energy, natural resources and environmental economics, energy systems. Springer, Berlin, pp 187–200. doi:10.1007/978-3-642-12067-1
54. Kuhn D (2009) Convergent bounds for stochastic programs with expected value constraints. J Optimiz Theory Appl 141(3):597–618. doi:10.1007/s10957-008-9476-1
55. LaSalle JP (1959) Time optimal control systems. Proc Nat Acad Sci USA 45(4):573–577. http://www.pnas.org/content/45/4
56. Little JDC (1955) The use of storage water in a hydroelectric system. Oper Res 3(2):187–197. doi:10.1287/opre.3.2.187

57. Lucia JJ, Schwartz ES (2002) Electricity prices and power derivatives: evidence from the Nordic power exchange. Rev Derivatives Res 5:5–50
58. Massé P (1946) Les réserves et la régulation de l'avenir dans la vie économique-avenir aléatoire. In: Actualités Scientifiques et Industrielles, vol 1008. Hermann and Cie, Paris
59. Mo B, Gjelsvik A, Grundt A (2001) Integrated risk management of hydro power scheduling and contract management. IEEE Trans Power Syst 16(2):216–221. doi:10.1109/59.918289
60. Morton DP (1996) An enhanced decomposition algorithm for multistage stochastic hydroelectric scheduling. Ann Oper Res 64(1):211–235. doi:10.1007/BF02187647
61. Näsäkkälä E, Keppo J (2008) Hydropower with financial information. Appl Math Financ 15(5-6):503–529. doi:10.1080/13504860701852494
62. Ni E, Luh PB, Rourke S (2004) Optimal integrated generation bidding and scheduling with risk management under a deregulated power market. IEEE Trans Power Syst 19(1):600–609. doi:10.1109/TPWRS.2003.818695
63. Nord Pool (2011) Elspot prices. http://www.nordpoolspot.com
64. Pereira MVF, Pinto LMVG (1985) Stochastic optimization of a multireservoir hydroelectric system: a decomposition approach. Water Resour Res 21(6):779–792. doi:10.1029/WR021i006p00779
65. Pereira MVF, Pinto LMVG (1991) Multi-stage stochastic optimization applied to energy planning. Math Program 52(1–3):359–375. doi:10.1007/BF01582895
66. Pflug GC (2000) Some remarks on the value-at-risk and the conditional value-at-risk. In: Uryasev SP (ed) Pobabilistic constrained optimization. Nonconvex optimization and its application, vol 49, Chap 1. Kluwer Academic, Dordrecht, pp 272–281
67. Pflug GC, Römisch W (2007) Modeling, measuring, and managing risk. World Scientific, New Jersey. doi:10.1142/9789812708724
68. Philbrick CRJ, Kitanidis PK (1999) Limitations of deterministic optimization applied to reservoir operations. J Water Resour Plan Manage 125(3):135–142. doi:10.1061/(ASCE)0733-9496(1999)125:3(135)
69. Philpott AB, de Matos VL (2012) Dynamic sampling algorithms for multi-stage stochastic programs with risk aversion. Eur J Oper Res 218(2):470–483. doi:10.1016/j.ejor.2011.10.056
70. Philpott AB, Craddock M, Waterer H (2000) Hydro-electric unit commitment subject to uncertain demand. Eur J Oper Res 125(2):410–424. doi:10.1016/S0377-2217(99)00172-1
71. Powell WB, George A, Simao H, Scott W, Lamont A, Stewart J (2012) Smart: a stochastic multiscale model for the analysis of energy resources, technology, and policy. INFORMS J Comput 24(4):665–682. doi:10.1287/ijoc.1110.0470
72. Pritchard G, Philpott AB, Neame PJ (2005) Hydroelectric reservoir optimization in a pool market. Math Program 103(3):445–461. doi:10.1007/s10107-004-0565-0

73. Read EG, Hindsberger M (2010) Constructive dual DP for reservoir optimization. In: Rebennack S, Pardalos PM, Pereira MVF, Iliadis NA (eds) Handbook of power systems I, Energy systems. Springer, Berlin, pp 3–32. doi:10.1007/978-3-642-02493-1
74. Rocha P, Kuhn D (2012) Multistage stochastic portfolio optimisation in deregulated electricity markets using linear decision rules. Eur J Oper Res 216(2): 397–408. doi:10.1016/j.ejor.2011.08.001
75. Rockafellar RT, Uryasev S (2000) Optimization of conditional value-at-risk. J Risk 2:21–41
76. Rockafellar RT, Uryasev S (2002) Conditional value-at-risk for general loss distributions. J Bank Financ 26(7):1443–1471. doi:10.1016/S0378-4266(02)00271-6
77. Römisch W, Vigerske S (2010) Recent progress in two-stage mixed-integer stochastic programming with applications to power production planning. In: Rebennack S, Pardalos PM, Pereira MVF, Iliadis NA (eds) Handbook of power systems I, Energy systems. Springer, Berlin, pp 177–208. doi:10.1007/978-3-642-02493-1
78. Sagastizábal C (2012) Divide to conquer: decomposition methods for energy optimization. Math Program 134(1):187–222. doi:10.1007/s10107-012-0570-7
79. Sen S, Yu L, Genc T (2006) A stochastic programming approach to power portfolio optimization. Oper Res 54(1):55–72. doi:10.1287/opre.1050.0264
80. Shapiro A (2011) Analysis of stochastic dual dynamic programming method. Eur J Oper Res 209(1):63–72. doi:10.1016/j.ejor.2010.08.007
81. Tipping J, Read EG (2010) Hybrid bottom-up/to-down modeling of prices. In: Rebennack S, Pardalos PM, Pereira MVF, Iliadis NA (eds) Handbook of power systems II, Energy systems. Springer, Berlin, pp 213–238. doi:10.1007/978-3-642-12686-4
82. Valenzuela J, Mazumdar M (2005) A probability model for the electricity price duration curve under an oligopoly market. IEEE Trans Power Syst 20(3): 1250–1256. doi:10.1109/TPWRS.2005.851966
83. Vespucci MT, Maggioni F, Bertocchi MI, Innorta M (2012) A stochastic model for the daily coordination of pumped storage hydro plants and wind power plants. Ann Oper Res 193(1):91–105. doi:10.1007/s10479-010-0756-4
84. Votruba L, Broža V (1989) Water management in reservoirs, developments in water science, vol 33. Elsevier, Amsterdam. http://www.sciencedirect.com/science/bookseries/01675648/33
85. Wallace SW (2010) Stochastic programming and the option of doing it differently. Ann Oper Res 177(1):3–8. doi:10.1007/s10479-009-0600-x
86. Wallace SW, Fleten SE (2003) Stochastic programming models in energy. In: Ruszczyński A, Shapiro A (eds) Stochastic programming. Handbooks in operations research and management science, vol 10. Elsevier, London, pp 637–677 doi:10.1016/S0927-0507(03)10010-2

87. Wolfgang O, Haugstad A, Mo B, Gjelsvik A, Wangensteen I, Doorman G (2009) Hydro reservoir handling in Norway before and after deregulation. Energy 34(10):1642–1651. doi:10.1016/j.energy.2009.07.025
88. Zivot E, Wang J (2006) Modelling financial time series with S-PLUS, 2nd edn. Springer, New York

Chapter 5
On Cutting Plane Algorithms and Dynamic Programming for Hydroelectricity Generation

Andy Philpott, Anes Dallagi, and Emmanuel Gallet

Abstract We consider *dynamic programming* (DP) approximations to hydroelectric reservoir scheduling problems. The first class of approximate DP methods uses decomposition and multi-modeling heuristics to produce policies that can be expressed as the sum of one-dimensional *Bellman functions*. This heuristic allows us to take into account non-convexities (appearing in models with head effect) by solving a MIP at each time stage. The second class of methods uses cutting planes and sampling. It is able to provide multidimensional policies. We show that the cutting plane methods will produce better policies than the first DP approximation on two convex problem formulations of different types. Modifying the cutting plane method to approximate the effect of reservoir head level on generation also yields better results on problems including these effects. The results are illustrated using tests on two river valley systems.

5.1 Introduction

The mid-term hydrothermal scheduling problem involves determining a policy of releasing water from reservoirs for hydroelectricity generation and generating from thermal plant over some planning horizon of months or years so as to meet the future demand for electricity at the lowest expected fuel cost. The first models (dating back to [8, 9]) for these problems used dynamic programming, a tool that was confined to systems with one or two reservoirs, unless reservoir aggregation heuristics (see, e.g., [2, 13]) are used.

A. Philpott (✉)
University of Auckland, Private Bag 92019, Auckland, New Zealand
e-mail: a.philpott@auckland.ac.nz

A. Dallagi • E. Gallet
EDF R&D, 1 avenue du Général de Gaulle, 92141 Clamart Cedex, France
e-mail: anes.dallagi@edf.fr; emmanuel.gallet@edf.fr

R.M. Kovacevic et al. (eds.), *Handbook of Risk Management in Energy Production and Trading*, International Series in Operations Research & Management Science 199,
DOI 10.1007/978-1-4614-9035-7_5,

An effort to model systems with multiple reservoirs led to the development in the 1980s and 1990s of various multi-stage stochastic linear programming models (see, e.g., [7]) using scenario trees. Stochastic dual dynamic programming (SDDP) [10] was developed as a response to the problem of dealing with a rapidly growing scenario tree. This method approximates the future cost function of dynamic programming using a piecewise linear outer approximation, defined by cutting planes or *cuts* computed by solving linear programs. This avoids the curse of dimensionality that arises from discretizing the state variables. The intractability arising from a branching scenario tree is avoided by essentially assuming stagewise independent uncertainty. This allows cuts to be shared between different states, effectively collapsing the scenario tree.

There has been little published work comparing the SDDP methodology with classical dynamic programming. A relatively old paper by Archibald et al. [1] shows that nested Benders decomposition outperforms classical dynamic programming in some computational tests on models with a small number of stages and scenarios, but becomes intractable as these grow. Our contribution is to demonstrate some advantages of SDDP-type algorithms in comparison with dynamic programming when the problem has many stages, so that nested Benders decomposition is computationally intractable, at least in its standard scenario-tree form.

In some electricity systems hydrothermal scheduling problems are solved using price decomposition (see, e.g., [6]). In a deterministic setting this method gives a subproblem to be solved for each thermal unit and each hydro river-chain. These subproblems can in principle be coordinated by price to yield an overall generation plan that meets demand in every period at minimum total cost. The coordination problem has been less well studied under inflow uncertainty (although see, e.g., [3]). Here subproblems must be coordinated by a random price process to yield an overall generation plan that meets the demand at minimum expected cost. A similar set of (sub)problems arises when river-chains are operated by different agents in a competitive electricity pool market, where the prices over time come from the pool. In both these settings the river-chain optimization subproblem is a challenging problem to solve, since it must handle uncertainties in prices and inflows. Indeed it is a variant of the hydrothermal scheduling problem in which the electricity price is modeled as the (random) marginal cost of an infinitely large thermal unit.

In this paper we compare two approaches to solving this problem and analyze their differences. The first approach uses a stochastic dynamic programming heuristic applied to a low-dimensional approximation of the state space. We discuss the form of the approximation and demonstrate biases in the marginal values of water. The second method is the dynamic outer approximation sampling algorithm (DOASA) described in [11], which is a special version of SDDP. We describe how the DOASA algorithm overcomes some of the biases of the former. The optimal value functions derived from the two methods are tested numerically on two river-chains.

The paper is laid out as follows. In the next section we describe the stochastic control problem that we wish to solve. This problem is illustrated in the following section by two example river-chains (RC1 and RC2) operated by EDF in France. The networks describing these river-chains have different topologies and so one

can use different heuristics to compute approximately optimal release policies for the reservoir control problem, with a view to computing estimates of water values for short-term optimization.[1] The heuristics (called MORGANE) are described in Sect. 5.3. In Sect. 5.4 we give a brief description of the DOASA algorithm and then describe how the forward simulation step of this method can be used to evaluate both the DOASA release policy and the corresponding release policies that arise by applying MORGANE. In Sect. 5.5 we present the results of applying the MORGANE heuristics to the example systems and compare this with the results of applying DOASA.

5.2 The Hydroelectric River-Chain Model

We consider a river-chain represented by a network of n nodes (reservoirs and junctions) and m arcs (canals or river reaches). The topology of the network can be represented by the $n \times m$ incidence matrix A, where

$$a_{ij} = \begin{cases} 1, \text{ if node } i \text{ is the tail of arc } j, \\ -1, \text{ if node } i \text{ is the head of arc } j, \\ 0, \text{ otherwise.} \end{cases} \tag{5.1}$$

By adding dummy nodes if necessary, we can ensure that every pair of nodes is joined by at most one arc. Let $x(t)$ denote a vector of reservoir storages in each node at the beginning of week t and $\omega(t)$ a vector of uncontrolled reservoir inflows (in cubic meters) that have occurred in week t. We let $h(t)$ be a vector of flow rates (cubic meters per hour) in the arcs in the network and $p(t)$ a vector of electricity prices at time t. Here we adopt the convention that these prices are applied to each flow in the network and are adjusted to account for conversion factors. Thus if arc j does not represent a generating station then $p_j(t) \equiv 0$, and if j is a station then $p_j(t)$ is the spot price of electricity multiplied by a scale factor η_j for that station (converting cubic meters of water passing though the station into MWh). Some flows represent spill (with $\eta_j = 0$) from reservoirs to the river reach below a station.

We also need to allow for more than one price period within a week. In our experiments we use $B = 21$ blocks in a week, each of duration $(d_b)_{b=1,\ldots,B}$ hours. These account for variations in price between peak and off-peak times. To model this we assume that $p(t)$ and $h(t)$ are of dimension mB, where the first B components of each vector correspond to arc 1, and so on. We then define the $m \times mB$ matrix

$$D = \begin{bmatrix} d_1 & d_2 & \ldots & d_B & 0 & 0 & \ldots & 0 & \ldots & \ldots & 0 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 & d_1 & d_2 & \ldots & d_B & \ldots & \ldots & \vdots & \vdots & \vdots & \vdots \\ \vdots & & & & & & & & \ldots & \ldots & 0 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 & & & & & & & d_1 & d_2 & \ldots & d_B \end{bmatrix} \tag{5.2}$$

[1] A deterministic optimization is operated on the short-term (within a day) using a more accurate model in order to provide the actual feasible releases to be performed.

so that the total quantity of flow through arc j in week t is $Dh(t)$, and the revenue earned is $p(t)^\top h(t)$, where component $B(j-1)+k$ of $p(t)$ now equals the electricity price/MWh in block k in week t multiplied by both d_b and η_b.

The hydroelectric river-chain problem we wish to solve seeks to construct a policy for generating electricity from the river-chain so as to maximize the expected revenue. We assume in this paper that the uncontrolled reservoir inflows are the only uncertain parameters and that these are stagewise independent. In each realization of uncertainty, the policy we seek will give a set of reservoir releases defined by $h(t)$ that generate electricity. The policy is defined in terms of a dynamic programming Bellman function $\mathbb{E}[V_t(x,\omega(t))]$ where $V_t(x,\omega(t))$ gives the maximum expected revenue that can be earned in weeks $t, t+1, \ldots$ when reservoir storage $x(t)=x$ and week t's inflow is known to be $\omega(t)$. Here

$$\begin{aligned} V_t(x,\omega(t)) = \max\ & p(t)^\top h(t) + \mathbb{E}[V_{t+1}(x(t+1),\omega(t+1))] \\ \text{s.t.}\quad & x(t+1) = x - ADh(t) + \omega(t), \\ & 0 \le h(t) \le b, \quad 0 \le x(t+1) \le r. \end{aligned} \tag{5.3}$$

We place a limit on the time horizon of $t=T$ and specify a future value function $V_{T+1}(x(T+1),\omega(T+1))$ to give a bounded problem.

5.3 River-Chain Optimization and Multi-modeling

In this section we present the MORGANE dynamic programming heuristics as applied to two river-chains. These heuristics are applied in order to keep some advantages of the stochastic dynamic programming method. In fact, they allow us to model non-convexities in the problem without, as far as the stage problems are solved, losing the theoretical validity of the method. Nevertheless dynamic programming is faced with the curse of dimensionality, and so the heuristics seek to avoid this by reducing the dynamic programming calculations to a sequence of one-dimensional problems.

The river-chains are called the RC1 Valley and the RC2 river-chain. Due to confidentiality issues the names of the river-chains and reservoirs are anonymized. Schematic representations of these are shown in Figs. 5.1 and 5.2. The RC1 system has three main reservoirs feeding a chain of two essentially run-of-the-river generating stations. The RC2 system is a chain of stations some of which have headpond reservoirs with significant storage capacity.

The MORGANE model treats these river-chains slightly differently. For the RC1 system, MORGANE computes marginal water values for each of the three main reservoirs by performing a stochastic dynamic programming algorithm for each of these reservoirs considering that the storage levels of the other reservoirs are constrained to be at given levels at the end of each week. In the RC2 system, MORGANE applies a different heuristic that consists of decomposing the river-chains into two parts: an upstream one and a downstream one. Applying the first heuristic

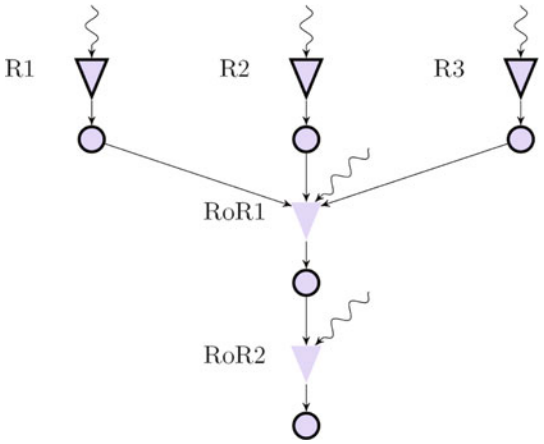

Fig. 5.1 The RC1 river system

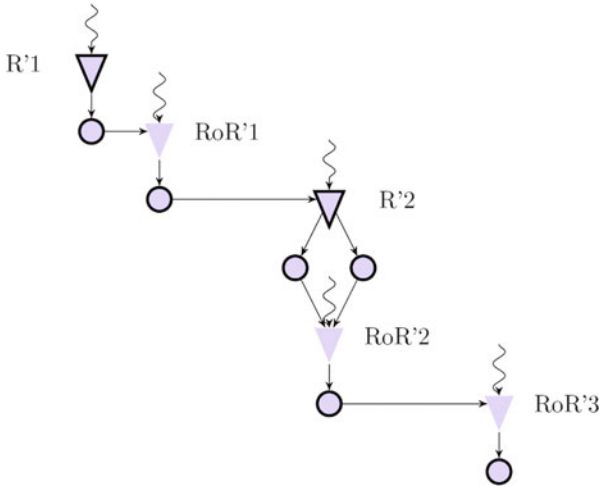

Fig. 5.2 The RC2 river system

to the upstream part generates a controlled flow feeding the downstream part. Again the first heuristic is applied to the downstream section in order to compute water values for the reservoir at the top of this section.

5.3.1 Heuristic 1: Constrained Multi-modeling

The curse of dimensionality means that computing multidimensional policies using dynamic programming is numerically intractable. The idea of heuristic 1 is to compute a multidimensional policy as the sum of one-dimensional ones. In this approach, each reservoir present in the considered river-chain is operated as if the other

reservoirs have storage level targets to reach at the start and end of each stage. This target is an a priori fixed proportion of the maximum storage capacity.

Solving a hydroelectric river-chain model, composed of n interconnected reservoirs, thus amounts to solving n subproblems. Each subproblem is solved using dynamic programming, resulting in n revenue-to-go functions.

We denote by $V_t^{[i]}(x_i, \omega(t))$ the maximum expected revenue that can be earned in weeks $t, t+1, \ldots, T$, when reservoir i storage is $x_i(t) = x_i$ and reservoir $j \neq i$ storage is $x_j(t) = x_j(t+1) = \alpha_j r_j, \quad \alpha_j \in [0,1]$ at the start and end of each time stage. Week t's inflow is known to be $\omega(t)$. Here

$$\begin{aligned} V_t^{[i]}(x_i, \omega(t)) = \max\ & p(t)^\top h(t) + \mathbb{E}[V_{t+1}^{[i]}(x_i(t+1), \omega(t+1))] \\ \text{s.t.}\quad & x_i(t) = x_i, \\ & x(t+1) = x(t) - ADh(t) + \omega(t), \\ & x_j(t) = \alpha_j r_j, \quad \forall j \neq i \\ & x_j(t+1) = \alpha_j r_j, \quad \forall j \neq i \\ & 0 \leq h(t) \leq b, \quad 0 \leq x_i(t+1) \leq r_i. \end{aligned} \tag{5.4}$$

The multidimensional Bellman function that will allow us to recompute an overall policy will be the sum of one-dimensional Bellman functions:

$$V_t(x(t), \omega(t)) = \sum_{i=1}^{n} V_t^{[i]}(x_i(t), \omega(t)). \tag{5.5}$$

It is obvious that this heuristic is suboptimal in the sense that the subproblems have less flexibility than the original problem formulation. Nevertheless, this heuristic allows us to maintain the advantages of classical dynamic programming. In particular, no convexity assumptions are needed in order to perform this heuristic as long as we can solve the (MIP) subproblems with a reasonable amount of computational effort.

5.3.2 Heuristic 2: Geographical Decomposition

The previous heuristic is adapted to parallel reservoirs feeding into the same turbine or run-of-the-river installations. In fact in that kind of configuration the strategies of the different reservoirs defined by the heuristic are optimal if the downstream constraints are never binding. We discuss this issue in more depth in the Appendix.

For a cascade-type river-chain this decoupling is typically not possible, and a geographical decomposition can be applied between downstream and upstream using the special structure of the river network.

For the RC2 river-chain the matrices describing the network can be decomposed as an upstream part containing $p = 2$ reservoirs and a downstream part containing $n - p = 3$ reservoirs, so

$$A = \begin{bmatrix} A^{[1]} & B^{[1]} \\ B^{[2]} & A^{[2]} \end{bmatrix} = \left[\begin{array}{ccc|ccc} -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 \\ \hline 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 \end{array}\right] \tag{5.6}$$

and

$$D = \begin{bmatrix} D^{[1]} & D^{[2]} \end{bmatrix}. \tag{5.7}$$

Heuristic 2 looks for strategies depending only on two reservoir levels: the topmost one (reservoir 1) and the most valuable[2] one in the middle of the chain (reservoir $p+1$).

First, we use heuristic 1 in order to compute a policy for the topmost reservoir while fixing the releases of the other reservoirs in order to meet an a priori fixed target at the start and end of each time stage. We denote by $V_t^{[1]}(x_1, \omega(t))$ the maximum expected revenue that can be earned in weeks $t, t+1, \ldots, T$, when the topmost reservoir storage is $x_1(t) = x_1$ and reservoir $j \neq 1$ storage is $x_j(t) = x_j(t+1) = \alpha_j r_j, \quad \alpha_j \in [0,1]$. Week t's inflow is known to be $\omega(t)$. Here

$$\begin{aligned} V_t^{[1]}(x_1, \omega(t)) = \max\ & p(t)^\top h(t) + \mathbb{E}[V_{t+1}^{[1]}(x_1(t+1), \omega(t+1))] \\ \text{s.t.}\quad & x_1(t) = x_1, \\ & x(t+1) = x(t) - ADh(t) + \omega(t), \\ & x_j(t) = \alpha_j r_j, \quad \forall j \neq 1 \\ & x_j(t+1) = \alpha_j r_j, \quad \forall j \neq 1 \\ & 0 \leq h(t) \leq b, \quad 0 \leq x_1(t+1) \leq r_1. \end{aligned} \tag{5.8}$$

Second, we use the previously computed releases $h^{[1]}(t) \in \mathbb{R}^p$ feeding the downstream part, in order to optimize the second part of the river-chain. Hence, we denote by $V_t^{[2]}(x_{p+1}, \omega(t))$ the maximum expected revenue that can be earned in weeks $t, t+1, \ldots, T$, when the storage of the topmost reservoir (in the downstream section) is $x_{p+1}(t) = x_{p+1}$ and reservoir $j > p$ storage is $x_j(t) = x_j(t+1) = \alpha_j r_j, \quad \alpha_j \in [0,1]$. Week t's inflow in the downstream part of the river is denoted by $\omega^{[2]}(t)$. Here

[2] It could be the biggest one or the most valuable one from the operator's point of view.

$$
\begin{aligned}
V_t^{[2]}(x_{p+1},\omega(t)) = \max\ & p^{[2]}(t)^\top h^{[2]}(t) + \mathbb{E}[V_{t+1}^{[2]}(x_{p+1}(t+1),\omega^{[2]}(t+1))] \\
\text{s.t.}\quad & x_{p+1}(t) = x_{p+1}, \\
& x^{[2]}(t+1) = x^{[2]}(t) - A^{[2]}D^{[2]}h^{[2]}(t) \\
& \qquad\qquad -B^{[2]}D^{[1]}h^{[1]}(t) + \omega^{[2]}(t), \\
& x_j(t) = \alpha_j r_j, \quad \forall j > p \\
& x_j(t+1) = \alpha_j r_j, \quad \forall j > p \\
& 0 \le h^{[2]}(t) \le b, \quad 0 \le x_{p+1}(t) \le r.
\end{aligned}
\tag{5.9}
$$

The Bellman function that defines our policy will then be the sum of the one-dimensional Bellman functions:

$$
V_t(x(t),\omega(t)) = V_t^{[1]}(x_1(t),\omega(t)) + V_t^{[2]}(x_{p+1}(t),\omega(t)). \tag{5.10}
$$

While simulating the generated policy for the remaining reservoirs, we add to the simulator the target constraints at the beginning and end of each stage.

As in the previous case, this heuristic is suboptimal in the sense that the subproblems have less flexibility than the original problem. But, as in the previous heuristic, we are able to keep some of the dynamic programming advantages (dealing with non-convexities).

5.4 Dynamic Outer Approximation Sampling Algorithm

The DOASA code [11] is based on the SDDP technique of Pereira and Pinto [10]. To solve a maximization problem, DOASA approximates $\mathbb{E}[V_t(x,\omega(t))]$ using a piecewise linear outer approximation that is updated using samples of the inflow process. Of course, this approximation is available only if convexity assumptions are made on the considered problem. Weekly prices are represented by a price duration curve with 21 blocks, and $\omega(t)$ is sampled from historical inflow observations.

The DOASA code yields an outer approximation to $\mathbb{E}[V_t(x,\omega(t))]$ at each stage t by solving the single-stage approximating problem:

$$
\begin{aligned}
SP(x,\omega(t))\text{: max}\ & p(t)^\top h(t) + \theta_{t+1} \\
\text{s.t.}\quad & x(t+1) = x - ADh(t) + \omega(t), && [\pi_t] \\
& 0 \le h(t) \le b, \quad 0 \le x(t+1) \le r, \\
& \alpha_{t+1}^k + \left(\beta_{t+1}^k\right)^\top x(t+1) \ge \theta_{t+1},\ k \in \mathscr{C}(t+1).
\end{aligned}
\tag{5.11}
$$

The flow balance constraints at each node i have dual variables π_t. These are used to compute the inequality constraints

$$\alpha_t^k + \left(\beta_t^k\right)^\top x(t) \geq \theta_t, \quad k \in \mathscr{C}(t) \tag{5.12}$$

which are the cuts defining an outer approximation of the future value function. For simplicity of notation, we have represented this construction as applying to all nodes i, but in practice we compute cuts only for the reservoirs with positive storage capacity.

At iteration k of the algorithm the cut coefficients are computed as follows:

1. Solve the stage problem $SP(x^k(1), \omega(1))$ for the known realization $\omega(1)$, giving optimal solution $x^k(2)$ and optimal value $V_1(x^k(1), \omega(1))$.
2. For $t = 2, 3, \ldots, T$, solve the stage problem $SP(x^k(t), \omega(t))$ for a sample realization of ω, recording the solution $x^k(t+1)$.
3. For $t = T, T-1, \ldots, 2$, solve the stage problem $SP(x^k(t), \omega(t))$ for every realization of $\omega(t)$, recording the solution value $V_t(x^k(t), \omega(t))$ and duals $\pi_t(x^k(t), \omega(t))$, and adding the cut

$$\theta_t \leq \alpha_t + \beta_t^\top x(t), \tag{5.13}$$

to every problem at stage $t-1$, where

$$\beta_t = \mathbb{E}[\pi_t(x^k(t), \omega(t))], \tag{5.14}$$

and

$$\alpha_t = \mathbb{E}[V_t(x^k(t), \omega(t))] - \beta_t^\top (x^k(t)). \tag{5.15}$$

The algorithm terminates after little progress is observed in the upper bound estimate $V_1(x^k(1), \omega(1))$. The policy defined by the cuts can then be simulated, and its expected revenue estimated by a sample average. This gives an estimate of a lower bound on the optimal value that can be checked against the upper bound.

In contrast to SDDP that runs many forward passes simultaneously, DOASA performs a single forward pass in each iteration. This can be shown to be more effective at delivering good solutions when stopping the algorithm early [4].

5.5 Experiments

DOASA and MORGANE were applied to data from the two valley systems RC1 and RC2. In both systems, MORGANE was applied subject to the following restrictions[3]:

1. There are no constraints on reservoir levels apart from them ranging between 0 and their full capacity.

[3] These restrictions are not in favor of the MORGANE heuristics. In fact the heuristics are able to take into account stochasticity and time dependency on the prices, end non-convexity constraints.

2. All generating plant are assumed to be available at full capacity throughout the year. Real problems include randomness on the availability of these plants.
3. Electricity prices were assumed to be known in advance, and set to their average over all the scenarios that had been generated. Each river-chain used a different set of average prices. Prices were for 21 blocks of time during the week representing peak, off-peak, and shoulder periods. The number of hours in each block varies with the day of week and the river valley.
4. A set of 41 weekly inflow sequences was used to represent the reservoir inflow distribution, assumed to be stagewise independent.
5. The factors for converting flow to electricity in each station were set to be constant (i.e., we assume that they do not vary with reservoir head level) in a first experiment and vary in a second one.

Under these restrictions MORGANE was solved over a 104-week horizon and marginal water values were computed for each reservoir at the end of each week. The marginal water values for week 52 were then converted into cuts and provided to DOASA as end conditions. DOASA was then applied over the first 52 weeks of the planning horizon.

5.5.1 Converting Marginal Water Values to Cuts

Upon terminating, MORGANE provides marginal water values for each reservoir as a function of their level. This function for reservoir $i = 1,\ldots,n$ is represented as a list of reservoir levels and marginal water values denoted (x_{ik}, β_{ik}), $k = 0,1,\ldots,K$. This can be converted into cuts by choosing an arbitrary future value α_{i0} from an empty reservoir at time t and then computing α_{ik} recursively using

$$\alpha_{ik} = \alpha_{ik-1} + (\beta_{ik-1} - \beta_{ik})x_{ik}, \quad k = 1,2,\ldots,K, \tag{5.16}$$

as shown in Fig. 5.3.

The future value function $\mathbb{E}[V_{T+1}(x(t+1), \omega(t+1))]$ is then represented as the sum of the individual reservoir value functions. Thus

$$\mathbb{E}[V_{t+1}(x(t+1), \omega(t+1))] = \sum_{i=1}^{n} \max_{k=0,1,\ldots,K} \{\alpha_{ik} + \beta_{ik} x_i(t+1)\}. \tag{5.17}$$

We model this in DOASA using a multi-cut representation

$$\theta_{t+1} = \sum_{i=1}^{n} \theta_{t+1}(i) \tag{5.18}$$

$$\alpha_{ik} + \beta_{ik} x_i(t+1) \geq \theta_{t+1}(i), \quad k = 0,1,\ldots,K, \; i = 1,2,\ldots,n. \tag{5.19}$$

Depending on the choice of α_{i0} this representation gives a different level for $\mathbb{E}[V_{t+1}(x(t+1), \omega(t+1))]$, but accurately reproduces the marginal water values for

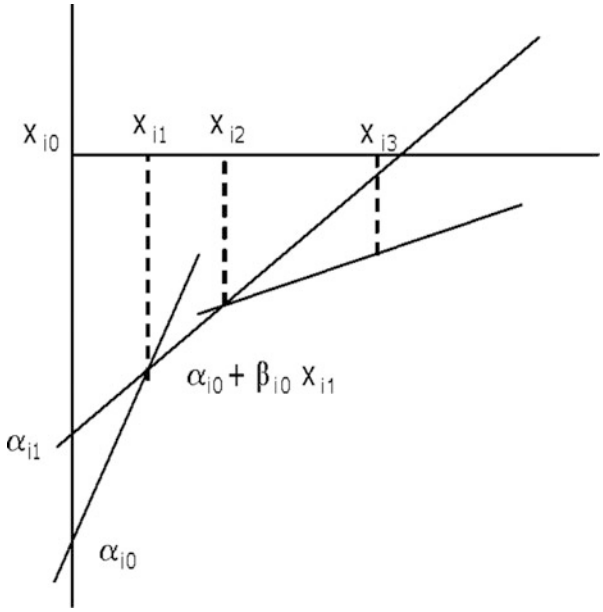

Fig. 5.3 Converting marginal water values to cuts

each storage level at which they are recorded by MORGANE, under the assumption that the MORGANE files of marginal water values give these values at the lowest possible storage level that they apply (i.e., the step function is continuous from the right). We remark that the above construction can be used for representing the MORGANE policy at any stage $t \leq T$. We thus use this both for providing a future cost function to DOASA at $t = 52$ and for simulating the MORGANE policies in our experiments, so that the same code is used to simulate both MORGANE and DOASA policies.

5.5.2 Modeling Variations in Head

Recall that for generating stations $p_j(t)$ denotes the spot price of electricity multiplied by a scale factor η_j for that station (converting cubic meters of water passing through the station into MWh). In practice the scale factor η_j is a function of net head, namely the difference in height of the headwater and tailwater of the turbine generating electricity. A common model (see [12]) assumes that η_j is net head times some efficiency term that depends on flow rate. If we assume that the tailwater height and flow rate is constant then η_j is linear with head level. The headwater height then depends on the volume of water x stored in the reservoir upstream of the station as a concave function since the area of the reservoir increases with head height.

In MORGANE the relationship between conversion factor and head is expressed using a finite set of hydro production functions that depend on reservoir level x. Each production function is modeled using two linear pieces defined by the most efficient flow rate h_e and the maximum flow rate h_m, both of which depend on x. When the reservoir volume is x, the power generated by flow rate h is

$$\begin{aligned} E(h,x) = \max_{h_1,h_2} \ & \eta_e(x)h_1 + \eta_m(x)h_2, \\ \text{s.t. } & h_1 + h_2 = h, \\ & h_1 \le h_e(x), \quad h_2 \le h_m(x) - h_e(x). \end{aligned} \tag{5.20}$$

If values of $\eta_e(x)$ and $\eta_m(x)$ are specified for a range of reservoir volumes then we may compute $E(h,x)$ by linearly interpolating these values and solving the maximization problem above.

The term $p(t)^\top h(t)$ in the objective function of our stage problem becomes $p(t)^\top E(t)$, where $p(t)$ and $E(t)$ are of dimension mK and have components that correspond, respectively, to spot price at time t in each time block and network arc and to $E(h,x)$, where h is the station flow in that time block and network arc, and x is the storage in the upstream reservoir at time t.

The DOASA code now yields an outer approximation to $\mathbb{E}[V_t(x,\omega(t))]$ at each stage t by solving the approximating problem:

$$\begin{aligned} SP(x,\omega(t))\text{: max } & p(t)^\top E(t) + \theta_{,t+1} \\ \text{s.t. } & x(t+1) = x - ADh(t) + \omega(t), && [\pi_t] \\ & 0 \le h(t) \le b, \quad 0 \le x(t+1) \le r, \\ & E(t) = \eta_e(x)h_1 + \eta_m(x)h_2, \\ & h_1 + h_2 = h(t), \\ & h_1 \le h_e(x), && [\rho_t] \\ & h_2 \le h_m(x) - h_e(x), && [\sigma_t] \\ & \alpha_{t+1}^k + \left(\beta_{t+1}^k\right)^\top x(t+1) \ge \theta_{t+1},\ k \in \mathscr{C}(t+1). \end{aligned} \tag{5.21}$$

Even though $SP(x,\omega(t))$ is a convex problem for fixed x, this does not guarantee that the optimal value function $V_t(x,\omega(t))$ is concave as a function of x, even if $\eta_e(x)$ and $\eta_m(x)$ are concave. Indeed it is conceivable that the marginal value of water might increase as reservoirs fill up if the increase in head makes each cubic meter of water more valuable for generation.

A valid outer approximation of $V_t(x, \omega(t))$ by cutting planes requires that it is concave. If we apply the DOASA algorithm to this model then the value of the first-stage problem can no longer be guaranteed to be an upper bound on the value of an optimal policy. Of course it is possible to construct a concave approximation of each production function (see [5]) and use this. However we simply assume that the lack of concavity of these functions is not too severe, so that the cuts computed by DOASA give a reasonably good policy even if the value function is not concave.

To account for head effects in stage 3 of DOASA we now solve $SP(x, \omega(t))$ giving $h_1^k(\omega(t))$ and $h_2^k(\omega(t))$ and duals $\pi_t(x^k(t), \omega(t))$. We can then add the cut

$$\theta_t \leq \alpha_t + \beta_t^\top x(t), \tag{5.22}$$

to every problem at stage $t-1$, where

$$\beta_t = \mathbb{E}[\pi_t(x^k(t), \omega(t)) + p(t)^\top g(x^k(t), \omega(t))], \tag{5.23}$$

and

$$g(x, \omega) = \eta_e'(x) h_1^k(\omega) + \eta_m'(x) h_2^k(\omega). \tag{5.24}$$

This calculation ignores the dependence of h_e and h_m on x which could be included as extra terms in β_t, namely

$$\mathbb{E}[\Big(\rho_t(x^k(t), \omega(t)) - \sigma_t(x^k(t), \omega(t))\Big)^\top h_e'(x^k(t)) + \sigma_t(x^k(t), \omega(t))^\top (h_m'(x^k(t)))], \tag{5.25}$$

where ρ_t and σ_t are the dual variables corresponding to the bounds on h_1 and h_2 as shown in the formulation above.

When head effects are modeled, the forward pass in DOASA must also be changed so as to use the appropriate interpolated production functions corresponding to the state variable that is visited in the forward simulation.

The policies obtained by computing 100 of these cuts in DOASA were simulated against the policies obtained from MORGANE to give the plots shown in Fig. 5.7 for the RC1 and Fig. 5.8 for RC2 systems.

5.6 Numerical Results

We present in this section two different experiments over the two river-chains (RC1 and RC2). In the first experiment we assume that the head is a fixed constant in the production functions. The results concerning this first experiment will allow us to study the proposed MORGANE heuristics versus DOASA. The other experiments are presented to illustrate Sect. 5.5.2 where we take into account head variations.

5.6.1 Experiment 1: Fixed Head

In this experiment DOASA was run for 100 iterations, giving 100 cuts at each stage. Since MORGANE has approximately 40 marginal water values for each of three reservoirs, we claim that this gives a commensurate level of discretization. DOASA takes about one minute per iteration, and so it must be run for nearly two hours on this problem, while MORGANE takes about 10 min. The progress of the DOASA upper bound for a fixed head level model is shown in Fig. 5.4.

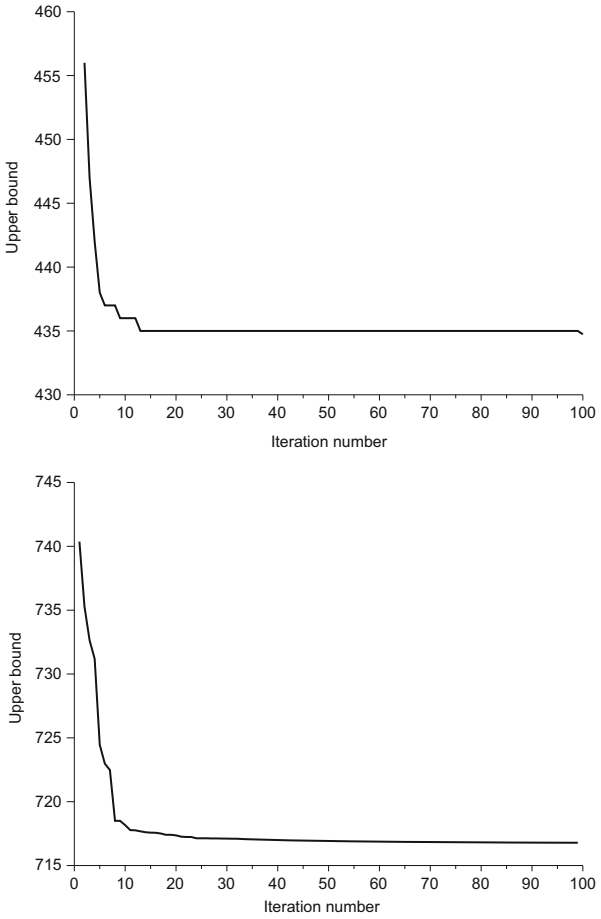

Fig. 5.4 Upper bound for revenue plus true value for RC1 DOASA policy assuming fixed head (RC1 *above*, RC2 *below*)

- For the RC1 system, the minimum (scaled[4]) upper bound is 434.740. The DOASA policy when simulated over 100 out-of-sample stagewise independent inflow sequences gives an expected value of 434.600 with a standard error of 0.139, which indicates that the DOASA policy is close to optimal.
- For the RC2 system,[5] the minimum (scaled) upper bound is 716.785. The DOASA policy when simulated over 100 out-of-sample stagewise independent inflow sequences gives an expected value of 716.454 with a standard error of 0.834, which indicates that the DOASA policy is close to optimal.[6]

The DOASA policy was then simulated over the 41 inflow sequences that were used to construct the policy. Since these sequences display some stagewise dependence, this is a sterner test of DOASA which assumes that inflows are stagewise independent.

- For the RC1 river system, the average optimal value over these scenarios was 434.161 with a standard error of 0.521 for the RC1 river system. This is lower than the out-of-sample expected value of 434.600 but very close. Recall that the upper bound of 434.740 is an upper bound on the value of the DOASA policy that assumes independence, so we cannot deduce that this is a bound on the best policy that took advantage of information about possible persistence in inflows. The policy from MORGANE was also simulated over the same inflow sequences. In 95 % of scenarios the values from DOASA were larger (see Fig. 5.5).
- For the RC2 river system, the average optimal value over these scenarios was 714.068 with a standard error of 3.451. The policy from MORGANE was also simulated over the same inflow sequences. This required some care as MORGANE provides water values only for reservoir R'1 and reservoir R'2. We therefore assumed in computing each week's releases that the terminal values of the reservoir levels of the other reservoirs are set at 50 % of their capacities. The simulation of the MORGANE policy then gives an average optimal value of 712.543 with a standard error of 3.389. In 85 % of scenarios the values from DOASA were larger (see Fig. 5.5).

Some insights into the reason for the difference between the policies can be seen by examining the marginal water values. Since both policies share the same value function at stage T, their stage T optimization problems should deliver the same water value functions at stage $T-1$. The actual values of these are shown for two sets of reservoir storage levels in Tables 5.1 and 5.2.

- For RC1, the marginal values computed by DOASA are lower than those computed by MORGANE. This can be explained by observing that MORGANE

[4] For confidentiality purposes, the units of measurement are omitted.

[5] This has a different network topology, and so it provides a useful comparison of the effect of river-chain topology on the policies from the two methodologies. As before, DOASA was run for 100 iterations, giving 100 cuts at each stage.

[6] An experiment with 200 cuts gave a smaller upper bound of 716.700 and an estimated value of 716.467 with a standard error of 0.833.

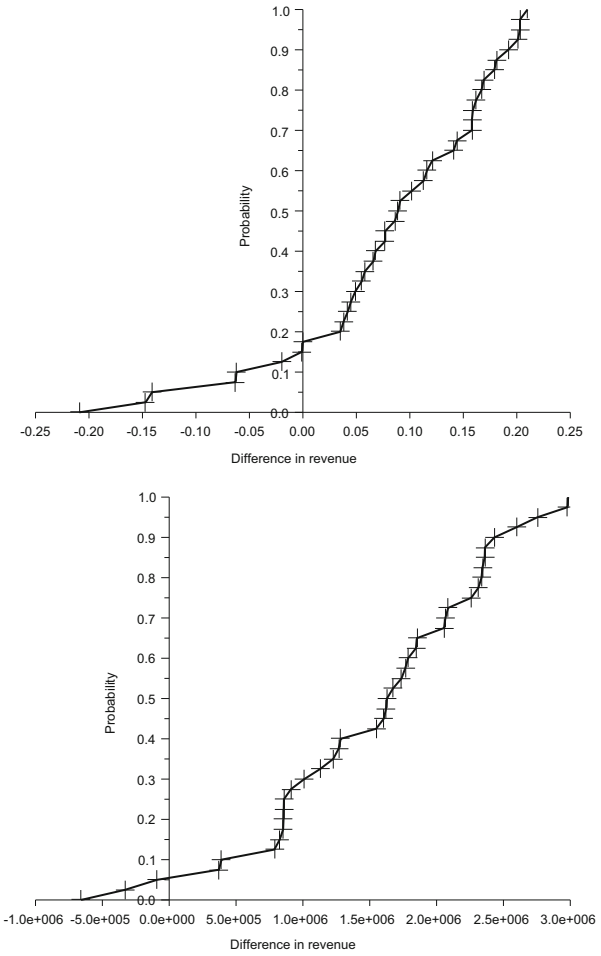

Fig. 5.5 Cumulative plot of the difference in value between the DOASA solution and the MORGANE solution for RC1 (*above*) and RC2 (*below*) over 41 scenarios assuming fixed head

assumes that all other reservoirs terminate at 50 % of capacity when computing marginal values for the reservoir in question. This adds an extra constraint on releases from these reservoirs over the week. Suppose that these constraints result in no other capacity constraints being binding. Then the marginal value of extra water in reservoir R1 (serving the most efficient station) will be the revenue earned by passing this down the river-chain through generating stations. If however the optimal policy means that river reach capacity constraints below the run-of-the-river RoR1 station are binding then the marginal value of extra water in reservoir R1 will be the revenue earned by it minus the loss in revenue from reducing the flow in the other stations that are less efficient, so as to satisfy the constraint. In this way the marginal water values for each reservoir can de-

Table 5.1 Marginal water values (euros per cubic meter) computed from DOASA and MORGANE for end of stage 51 for each of the storage reservoirs assuming fixed head

	Reservoir R1	Reservoir R2	Reservoir R3
Storage (m^3)	13,818,000	3,577,000	14,520,000
DOASA (€/m^3)	0.151	0.176	0.150
MORGANE (€/m^3)	0.159	0.190	0.155
Storage (m^3)	13,818,000	7,114,000	14,520,000
DOASA (€/m^3)	0.154	0.161	0.148
MORGANE (€/m^3)	0.159	0.163	0.155

Table 5.2 Marginal water values (euros per cubic meter) computed from DOASA and MORGANE for end of stage 51 for each of the storage reservoirs assuming fixed head

	Reservoir R'1	Reservoir R'2
Storage (m^3)	203,500,000	79,000,000
DOASA (€/m^3)	0.0359	0.0167
MORGANE (€/m^3)	0.0344	0.0159

pend on the water levels in *other* reservoirs, as well as on the level of their own. In other words, if the other reservoirs are full it is more likely that the capacity constraints will bind and so the water value in reservoir R1 will be lower than the value obtained when the other reservoirs are empty. We examine this nonseparability of the value function in more detail in the Appendix.

- In the RC2 system, the water values for reservoir R'1 under the DOASA policy are higher than those for MORGANE, since if reservoir R'2 is constrained to be at 50 % of its value at the end of the current week, then this precludes transferring water through reservoir R'2 to later weeks. Thus if prices are high now and reservoir R'2 is constrained downstream, then we cannot generate in period 1 without spilling, which could be avoided if water can be transferred by storing it till a later period. We examine the separability of the value function for the cascaded system in the Appendix.

Despite the differences in marginal water values, the policies of DOASA and MORGANE perform similarly. The average storage levels over the 41 scenarios are shown in Fig. 5.6.

5.6.2 Experiment 2: Variation in Head

In this experiment DOASA was run for 100 iterations, giving 100 cuts at each stage on the same reference as the previous section. Head effect was taken into account while optimizing the river-chains and while simulating the obtained policies as

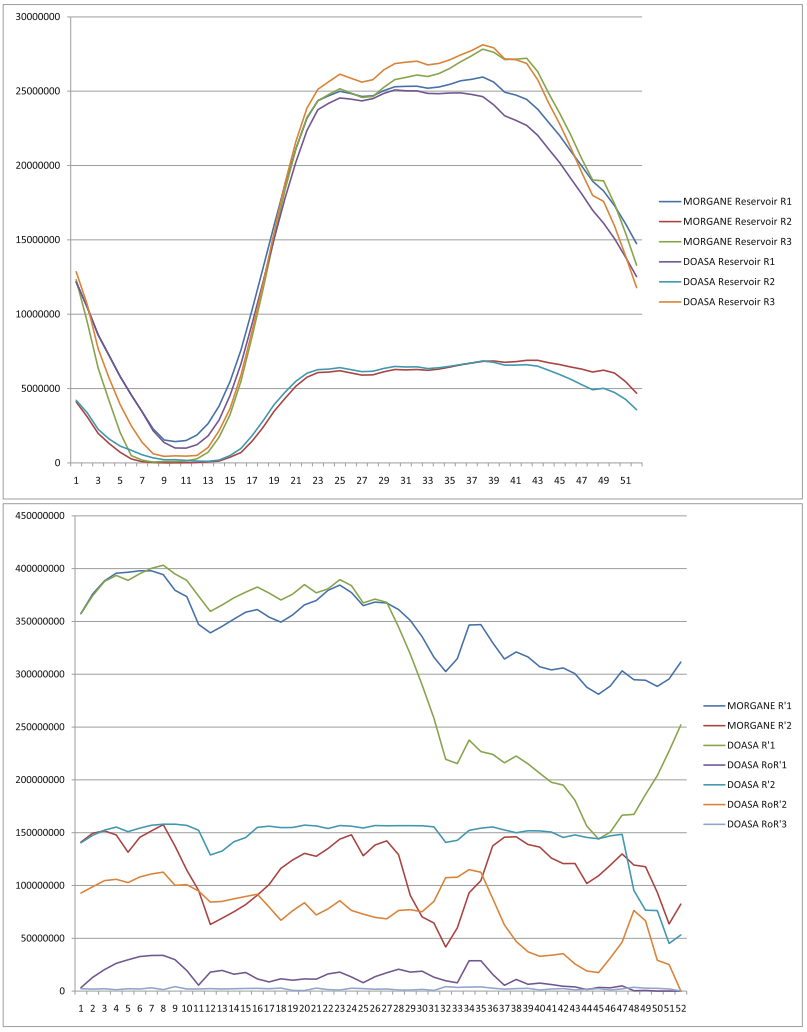

Fig. 5.6 Comparison of average stock levels in each reservoir (over 41 scenarios) for the DOASA and MORGANE policies (RC1 *above* and RC2 *below*) assuming fixed head

described in Sect. 5.5.2. For the RC2, cuts using DOASA are computed assuming that the three smallest reservoirs (i.e., those without storage) can vary their level between bounds. The MORGANE policy assumes that these reservoirs are fixed at their midpoint levels at the end of each week. So for comparison we conducted two computational experiments on the RC2 as follows:

1. We simulate the MORGANE policy, first with free endpoints on the three small reservoirs, and then with fixed weekly endpoints.

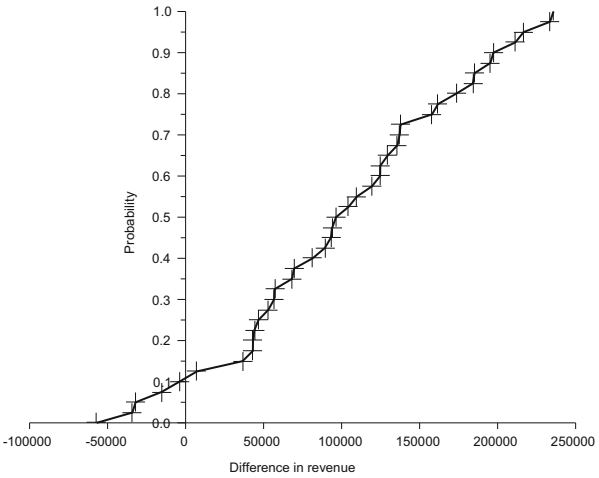

Fig. 5.7 Cumulative plot of the difference in value between the DOASA solution and the MORGANE solution for RC1 with head effect over 41 scenarios

2. We simulate the DOASA policy, first with free endpoints on the three small reservoirs, and then with fixed weekly endpoints.

In these experiments, DOASA behaves better than the MORGANE heuristics, especially if allowed more flexibility in the small reservoirs. However even if this is restricted, the DOASA policy is better than MORGANE showing on average:

- DOASA constrained saves 2M over MORGANE constrained
- DOASA unconstrained saves 3M over MORGANE unconstrained
- DOASA unconstrained saves 3.79M over MORGANE constrained

5.7 Conclusions

The aim of this paper is to compare a stochastic dynamic programming-based heuristic that can handle non-convexities appearing in real problems (MORGANE) to an outer approximation method that needs some convexity assumptions. The set of experiments chosen in this paper demonstrates that constructing DP policies using multivariate Bellman functions gives better results than methods that ignore the cross terms. The source of DOASA's advantage is from using a polyhedral surface that is not a separable sum of one-dimensional curves. When this feature is absent, for example when downstream constraints are never binding, the policies are equivalent. In practice, however there are always periods when these constraints are significant, and these periods cause the policies to diverge. We still need to conduct experiments in order to confirm these results while prices are stochastic and while

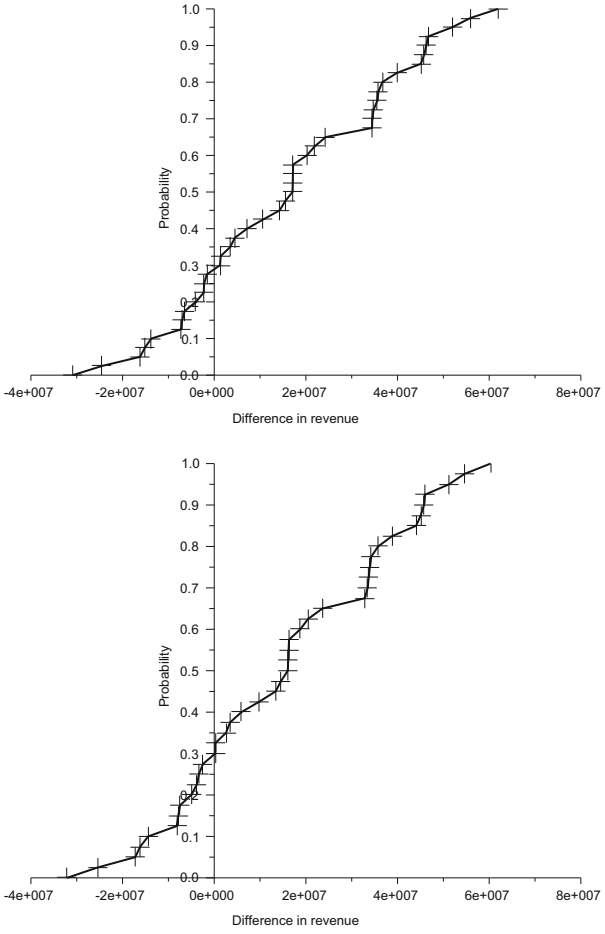

Fig. 5.8 Cumulative plot of the difference in value between the DOASA solution and the MORGANE solution for RC2 with free endpoints (*above*) and constrained endpoints (*below*) over 41 scenarios

the time dependency is taken into account (MORGANE can handle this dependency but not DOASA due to convexity issues). Experiments while we have more complicated constraints (storage constraints, solid water, etc.) are still to be made.

5.8 Appendix: Approximations Made by the Multi-modeling Heuristic

In this appendix we examine the implications of the approximations used by MORGANE for computing marginal water values. The analysis is different for each system, so we discuss each in turn.

5.8.1 RC1 River Chain

The MORGANE approximation of the RC1 system assumes that $V_t(x_1,x_2,x_3)$ is additively separable. To examine the separability of $V_t(x_1,x_2,x_3)$, consider a simplified version of RC1 with two reservoirs having capacity a_1 and a_2 and releases y_1 and y_2 through stations with price per unit of water flow of p_1 and p_2. Like RC1, the tailwater shares a common channel with capacity k. We wish to investigate the form of $V_t(x_1,x_2)$ when

$$V_{t+1}(x_1,x_2) = -(x_1-a_1)^2 - (x_2-a_2)^2 . \tag{5.26}$$

This gives

$$\begin{aligned} V_t(x_1,x_2) = \max_{y_1,y_2} \; & p_1y_1 + p_2y_2 - (x_1-y_1-a_1)^2 - (x_2-y_2-a_2)^2 \\ \text{s.t.} \quad & y_1 \le x_1 \\ & y_2 \le x_2 \\ & y_1+y_2 \le k \\ & y \ge 0 \end{aligned} \tag{5.27}$$

After some algebra, it can be shown that $V_t(x_1,x_2)$ can be separated into a sum $V_t^1(x_1)+V_t^2(x_2)$ for all values of x_1,x_2 except for those satisfying the following three conditions:

$$\begin{bmatrix} \left(x_2-\left(a_2-\frac{1}{2}p_2\right)\right)+\left(x_1-\left(a_1-\frac{1}{2}p_1\right)\right) > k \\ \left(x_1-\left(a_1-\frac{1}{2}p_1\right)\right), \left(x_2-\left(a_2-\frac{1}{2}p_2\right)\right) > 0 \\ \left|\left(x_2-\left(a_2-\frac{1}{2}p_2\right)\right)-\left(x_1-\left(a_1-\frac{1}{2}p_1\right)\right)\right| < k \end{bmatrix} . \tag{5.28}$$

In this exceptional case,

$$\begin{aligned} V_t(x_1,x_2) = & -\frac{1}{2}(x_1+x_2)^2 \\ & + \left(k+a_1+a_2+\frac{1}{2}p_1-\frac{1}{2}p_2\right)x_1 + \left(k+a_1+a_2-\frac{1}{2}p_1+\frac{1}{2}p_2\right)x_2 \\ & + \begin{pmatrix} \frac{1}{2}kp_1-ka_1-ka_2-\frac{1}{2}k^2+\frac{1}{2}kp_2-\frac{1}{2}a_1^2-a_1a_2-\frac{1}{2}a_1p_1 \\ +\frac{1}{2}a_1p_2-\frac{1}{2}a_2^2+\frac{1}{2}a_2p_1-\frac{1}{2}a_2p_2+\frac{1}{8}p_1^2-\frac{1}{4}p_1p_2+\frac{1}{8}p_2^2 \end{pmatrix} \end{aligned} \tag{5.29}$$

which contains the cross term $-x_1x_2$, and so is not separable.

5.8.2 RC2 River Chain

We now look at the water value calculations that MORGANE makes for a system like the RC2. We show by example that fixing the final water levels of other reservoirs (as MORGANE does) in order to compute the marginal water value of a given

reservoir can lead to a smaller number than the true value. Consider a system of two reservoirs in cascade as shown in Fig. 5.2.

Let $x(t) \in \mathbb{R}^2$ be the stock level at the start of period t in each reservoir (having capacity $a \in \mathbb{R}^2$)and let $u(t) \in \mathbb{R}^2$ denote the release from each reservoir, $h(t) \in \mathbb{R}^2$ the stochastic inflow and $s(t) \in \mathbb{R}^2$ the spill. All of these depend on time. Suppose that we control this system using a value of water for reservoir 1 only and setting a target in each stage on the level in reservoir 2. Given an expected future value function $V_{t+1}(x_1)$, the releases $u(\omega)$ are chosen to solve

$$\begin{array}{ll} \max & p_1u_1 + p_2u_2 + V_{t+1}(x_1 - u_1 - s_1 + h_1(\omega)) \\ \text{s.t.} & a_2/2 + h_2(\omega) + u_1 - u_2 - s_2 = a_2/2 \\ & 0 \le u_i \le b_i \\ & 0 \le s_i \le d_i \end{array} \tag{5.30}$$

We then compute

$$V_t(x_1) = \mathbb{E}(p_1u_1(\omega) + p_2u_2(\omega) + V_{t+1}(x_1 - u_1(\omega) - s_1(\omega) + h_1(\omega)) \tag{5.31}$$

We will test this approximation in a deterministic framework and study the marginal water value of reservoir 1. To do this consider two periods $t = 1, 2$ with no residual water value, and consider the additional value of an extra amount δ of water in reservoir 1 at the start of period 1. This can be computed by finding

$$\begin{array}{lll} Q(\delta) = & \max & \sum_{t=1}^{2} \sum_{i=1}^{2} p(t)u_i(t) \\ \text{s.t.} & & x_1 + \delta + h_1(1) - u_1(1) - s_1(1) = x_1(2) \\ & & x_2 + h_2(1) + u_1(1) - u_2(1) - s_2(1) = x_2(2) \\ & & x_1(1) + h_1(2) - u_1(2) - s_1(2) = x_1(3) \\ & & x_2(1) + h_2(2) + u_1(2) - u_2(2) - s_2(2) = x_2(3) \\ & & 0 \le x_i(t) \le a_i \\ & & 0 \le u_i(t) \le b_i \\ & & 0 \le s_i(t) \le d_i \end{array} \tag{5.32}$$

Suppose the problem data are given as

	$p(t)$	$x_1(t)$	$x_2(t)$	$h_1(t)$	$h_2(t)$
$t = 1$	2	1	0	0	0
$t = 2$	1	–	–	0	2

(5.33)

Then

$$\begin{array}{lll} Q(0) = & \max & 2u_1(1) + 2u_2(1) + u_1(2) + u_2(2) \\ \text{s.t.} & & 1 - u_1(1) = x_1(2) + s_1(1) \\ & & 0 + 1 + u_1(1) - u_2(1) = x_2(2) + s_2(1) \\ & & x_1(1) - u_1(2) = x_1(3) + s_1(2) \\ & & x_2(1) + u_1(2) - u_2(2) = x_2(3) + s_2(2) \\ & & 0 \le x_i(t) \le 1 \\ & & 0 \le u_1(t) \le 1, \quad 0 \le u_2(t) \le 2 \\ & & 0 \le s_i(t) \le 3 \end{array} \tag{5.34}$$

This has solution

	$u_1(t)$	$u_2(t)$	$x_1(t+1)$	$x_2(t+1)$	$s_1(t)$	$s_2(t)$
$t=1$	1	2	0	1	0	0
$t=2$	0	1	0	0	0	0

(5.35)

with return 7. Now if we increase x_1 to $x_1+\delta$, then $Q(\delta)=7+3\delta$, so the marginal water value at reservoir 1 is 3. However, if we constrain the storage to be one-half capacity at the end of period 1, then we have $Q(\delta)=7+2\delta$, so the marginal water value at reservoir 1 is 2, which is less than its value in the unconstrained case.

References

1. Archibald T, Buchanan C, McKinnon K, Thomas L (1999) Nested Benders decomposition and dynamic programming for reservoir optimisation. J Oper Res Soc 50(5):468–479
2. Archibald T, McKinnon K, Thomas L (2006) Modeling the operation of multireservoir systems using decomposition and stochastic dynamic programming. Nav Res Log (NRL) 53(3):217–225
3. Barty K, Carpentier P (UMA), Cohen G (CERMICS), Girardeau P (UMA, CERMICS) (2010) Price decomposition in large-scale stochastic optimal control. http://arxiv.org/pdf/1012.2092v2 (submitted)
4. De Matos V, Philpott A, Finardi E, Guan Z (2010) Solving long-term hydrothermal scheduling problems. Technical report, Electric Power Optimization Centre, University of Auckland
5. Diniz AL, Maceira MEP (2008) A four-dimensional model of hydro generation for the short-term hydrothermal dispatch problem considering head and spillage effects. IEEE Trans Power Syst 23(3):1298–1308
6. Dubost L, Gonzalez R, Lemaréchal C (2005) A primal-proximal heuristic applied to the French unit-commitment problem. Math Program (A 104):129–151
7. Jacobs J, Freeman G, Grygier J, Morton D, Schultz G, Staschus K, Stedinger J (1995) SOCRATES: a system for scheduling hydroelectric generation under uncertainty. Ann Oper Res 59:99–133
8. Little, J. (1955), The use of storage water in a hydroelectric system, J Oper Res Soc Amer 3(2):187–197.
9. Masse P (1946) Les réserves et la régulation de l'avenir dans la vie économique. Hermann, Paris
10. Pereira MVF, Pinto LMVG (1991) Multi-stage stochastic optimization applied to energy planning. Math Program 52:359–375
11. Philpott A, Guan Z (2008) On the convergence of stochastic dual dynamic programming and other methods. Oper Res Lett 36:450–455
12. Philpott A, Craddock M, Waterer H (2008) Hydro-electric unit commitment subject to uncertain demand. Eur J Oper Res 125:410–424
13. Turgeon A (1980) Optimal operation of multireservoir power systems with stochastic inflows. Water Resour Res 16(2):275–283

Chapter 6
Medium-Term Operational Planning for Hydrothermal Systems

Raphael E.C. Gonçalves, Michel Gendreau, and Erlon Cristian Finardi

Abstract The planning of operations of hydrothermal systems is, in general, divided into coordinated steps which focus on distinct modeling details of the system for different planning horizons. The medium-term operation planning (MTOP) problem, one of the operation planning steps and the focus of this chapter, aims at defining weekly generation for each power plant with the minimum expected operational cost over a specific planning horizon, with regard especially to the uncertainties related to reservoir inflows. Consequently, it is modeled as a stochastic problem and solving it requires the use of multistage stochastic optimization algorithms. In this sense, the objective of this chapter is to discuss the problem features, its particularities, and its importance in the overall operational planning. The stochastic methods usually used to solve this problem and some applications are also presented.

R.E.C. Gonçalves (✉) • M. Gendreau
NSERC/Hydro-Québec Industrial Research Chair on the Stochastic Optimization of Electricity Generation CIRRELT, Université de Montréal, C.P. 6128, Succ. Centre-ville, Montréal (Québec), Canada

Department of Mathematics and Industrial Engineering, École Polytechnique de Montréal, Canada
e-mail: raphaelecg@gmail.com; michel.gendreau@cirrelt.ca

E.C. Finardi
Universidade Federal de Santa Catarina - UFSC, Campus Universitário, Trindade, Florianópolis, SC, Caixa Postal 476, CEP 88040-900, Brazil

Laboratório de Planejamento de Sistemas de Energia Elétrica - LabPlan, EEL, CTC, UFSC, Florianópolis, SC, Brazil
e-mail: erlon.finardi@ufsc.br

R.M. Kovacevic et al. (eds.), *Handbook of Risk Management in Energy Production and Trading*, International Series in Operations Research & Management Science 199,
DOI 10.1007/978-1-4614-9035-7_6, © Springer Science+Business Media New York 2013

6.1 Introduction

The operation of electric power systems[1] covers a broad spectrum of activities or studies, among which the planning/scheduling of operations stands out (Test, 1797) [1, 2]. In general, this problem [3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21] is divided into several steps (long term, medium term, and short term), which have different planning horizons and which, consequently, each prioritize distinct details of the problem modeling. Briefly, the global problem involves the analysis of the important operational aspects of the system to define the optimal level of energy production to meet demand in an economical and reliable manner.

The medium-term operation planning (MTOP) problem, the focus of this chapter, aims to set weekly generation for each power plant with the minimum expected operation cost over a specific planning horizon, which can vary from months to a year or two, especially taking into account the uncertainties associated with some problem data (inflows and demand, among others). In addition, results from the MTOP can be used to set the spot energy price, depending on the regulatory framework.

Considering the importance of the MTOP, the credibility of the results is essential for the System Operator (SO), which is responsible for settling generation targets, and for the Energy Market (EM) agents, given the economic impacts of the transactions in the market energy environment. In this sense, constant improvements in the MTOP optimization model are required in order to satisfy or update the system and EM participants' requirements. This is the reason why a lot of work has focused on this step of the operation planning problem [6, 15, 22, 23, 24, 25, 26].

As the MTOP is part of a scheduling chain, it can be tightly linked with other steps in this chain in order to obtain the global solution of the operation planning problem, as occurs, for instance, in Brazil. The purpose is to maintain a temporal connection between scheduling chain steps. Briefly, the idea is to exchange information concerning the operational policies, which aim to propitiate coherent global decisions, as illustrated in Fig. 6.1, avoiding, for example, the inefficient use of the generation resources.

Given some problem features, especially those related to data uncertainties, the MTOP is quite complex to solve. As a consequence, solutions obtained by models that do not recognize the uncertainties can produce unsatisfactory results. In other words, the MTOP problem is essentially a stochastic optimization problem [27, 28]. In general, these uncertainties are associated with future inflows into the reservoirs for hydro or hydrothermal systems. Demand and future fuel or energy spot prices can also be modeled as random data, according to the predominance of generation resources of the system or the main objective of the problem.

Like most practical stochastic optimization problems, solving the MTOP problem requires a very significant computational effort, given that the size of the problem increases substantially with the representation of the uncertainties, with the number of stages and with the level of detail in system modeling. Therefore, it

[1] It is important to remark that the systems can be composed of hydro, thermal, or both power plants.

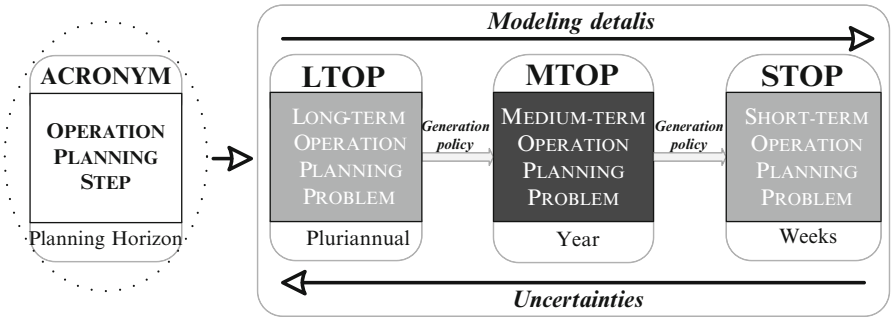

Fig. 6.1 Scheduling chain

is important to balance the level of detail used in modeling uncertainties and system operations by introducing some simplifications to allow the resolution of the problem in reasonable CPU times. For instance, hydro [16, 29] and thermal production and cost functions, which are typically nonlinear and non-convex [1], are usually modeled as linear or a piecewise linear functions to overcome their nonlinearity features. In this way, the resulting optimization problem can be approximated by a linear model.

One therefore ends up normally with a large multistage linear stochastic (MLS) problem [30], for which the use of stochastic decomposition algorithms [31] is essential to reduce the computational burden (in fact, attacking this problem directly keeping its standard structure, known as the deterministic equivalent problem, DEP, is often computationally infeasible). In this context, one must highlight the algorithms based on the Benders Decomposition (BD) principle [32], which display an excellent computational performance when dealing with problem instances with a representative number of scenarios, as shown in [33].

BD-based algorithms, such as the nested decomposition algorithm (NDA) [27] or stochastic dual dynamic programming (SDDP) [4], are broadly used to solve large-scale operation planning problems. The idea underlying NDA is to decompose the DEP into smaller subproblems with a restricted set of variables and constraints, which are easier to solve. These subproblems are solved individually and the coordination among them is performed by means of optimality constraints, which are built and updated iteratively. Although this is an efficient method, some disadvantages are often highlighted for this class of methods [27], as the difficulty to improve the quality of the solution when the iterative process is close to the optimal value. A similar idea can be found behind SDDP, which, however, is more suited to large-scale problems with many stages and scenarios. In this algorithm, scenario sampling, based on the original probability distributions of the random variables [2, 34, 35, 36, 37], is used to reduce the size of the problem, as discussed in another chapter of this book.

Algorithms with different characteristics in relation to the BD-based approach, such as the augmented Lagrangian (AL)-based algorithm [30, 38], have been successfully used to solve MLS problems. In this sense, it is possible to highlight the progressive hedging algorithm (PHA), which has been applied in several fields:

financial market [39], network flow problems [40] and, more recently, operation planning problems [22, 23, 26].

Like other AL-based algorithms, the PHA proceeds by relaxing the nonanticipativity constraints[2] [27, 41] providing an independent set of quadratic subproblems, which are linked via Lagrangian multipliers and penalized by a positive scalar parameter. In this class of methods, the aforementioned disadvantage of BD-based methods can be mitigated because the resulting subproblems are quadratic and they are still sensitive to the use of warm start techniques [25]. Nevertheless, its main disadvantages are associated with the penalty parameter adjustment, which is crucial to algorithm success. Otherwise, it can take a long time to converge.

Based on the aspects mentioned before, the remainder of this chapter aims to describe in detail the main features of the MTOP problem, the mathematical structure of the methods usually used to solve this kind of problem, some application, and important remarks concerning the problem resolution. More precisely, in Sect. 6.2, the stochastic optimization aspects are pointed out, emphasizing the challenges of the MTOP problem resolution. The general problem formulation is presented in Sect. 6.3. In Sect. 6.4, a brief idea of the algorithms and some application results are discussed. Finally, conclusions are presented in Sect. 6.5.

6.2 Stochastic Optimization Aspects

Unlike deterministic problems in which there are no uncertainties with respect to the future data, in stochastic programming problems it is necessary to optimize taking into account the data unpredictability. In order to make the problem resolution computationally viable, it is essential to ensure that the problem horizon has a finite number of stages and, additionally, knowing beforehand the probability distribution of the random variables.

In this context, some important aspects are often addressed in the literature [42], such as the modeling of random variables, the quality solution analysis based on the set of random variables, the problem resolution, solution algorithms, among other aspects.

Once the MTOP problem is essentially a stochastic problem, this section focuses on presenting a brief review associated with the stochastic programming, highlighting some important features which can make the MTOP formulation and challenges more understandable.

[2] The algorithm details and the stochastic theoretical aspects will be discussed in the next sections.

6.2.1 Problem Description

Assuming that the discrete probability distribution of the random variable is known, the classical two-stage linear stochastic programming problem or two-stage equivalent stochastic problem can be defined as follows:

$$\begin{aligned} &\min c_1^{\mathrm{T}} x_1 + \sum_{\omega \in \Omega_2} p_2^{\omega} c_2^{\mathrm{T}} x_2^{\omega} \\ &\text{s.t}: \quad A_1 x_1 = b_1, \\ &A_2^{\omega} x_2^{\omega} + B_2^{\omega} x_1 = b_2^{\omega}, \\ &x_1 \geq 0, x_2^{\omega} \geq 0. \end{aligned} \tag{6.1}$$

where:

T	Total of stages
t	Index of stage, so that $t = 1,T$
Ω_t	Set of realizations (nodes) on stage t
ω_t	Index associated with a specific realization (node) in the stage t, so that $\omega \in \Omega_t$
c_t	Cost vector related to stage t;
x_t	Vector decision of stage t, $x \in \Re^n$
p_t^{ω}	Probability associated to the each node ω, such that $\sum_{\omega \in \Omega_t} p_t^{\omega} = 1$
A_t^{ω}	Coefficient matrix in stage t (m_t x n_t)
B_t^{ω}	The technology matrix at stage t (m_t x n_t)
b_t^{ω}	Right-hand side for a specific realization ω at stage t

Based on this formulation, as called Deterministic Equivalent (ED) [43] of the stochastic problem, notice that uncertainties can be related to A or B matrices, as well as the vector b. Nevertheless, in the MTOP problem, the randomness is usually associated with the vector b, given that the inflows or the demand are the most future unpredictable data.

Solving this stochastic problem requires the use of methods that exploit the matrices structure of the problem, such as simplex method and interior point method [44].

Notice that the uncertainties ω are addressed to the second stage. The objective of problem (6.1) aims to minimize the cost over two stages, being composed by the costs associated with the decisions x_1 plus the expected future value of the second stage decisions. The remaining equations of (6.1) correspond to sets of constraints related to the first and second stages, interconnected by the technology matrix, B_2, beyond the variable bounds.

As it will be discussed later, this is the same idea of the MTOP problem formulation, i.e., an objective function composed of the total operational cost over the planning horizon and a set of constraints that are associated with the operational features or particularities of the system.

6.2.2 Scenario Tree

Considering the beforehand aspects with respect to the discrete probability distribution, the problem uncertainties are represented by means of a scenario tree. Thus, for a finite number of second stage realization Ω_2, the scenario tree can be represented as illustrated in Fig. 6.2.

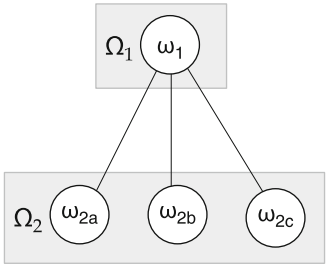

Fig. 6.2 Scenario tree

It is possible to say that the uncertainties representation by means of graphs aims to show the set of the problem realizations in each stage and the link (transition) among the each node decision. Thus, each node of the scenarios tree is associated with a specific realization or a set of random variables. In this context, a specific scenario can be defined as a path from the initial stage to the last stage, regarding a single realization at each stage or, simply, a set of realization from stage 1 onwards. In Fig. 6.2, for instance, a scenario 1 is composed of the realization ω_1 and ω_{2a}, the scenario 2 composed of the realization ω_1 and ω_{2b}, and the scenario 3 composed of ω_1 and ω_{2c}.

According to scenario assumptions, all scenarios share the same first stage realization/decision regardless of the second stage realization (the set of random variables of the first stage is identical in all of them). In this sense, it is possible to define an important concept in stochastic optimization: the nonanticipativity condition [41]. It means that the decisions are not made regarding future expectations, but based on past and current realizations of the random variables. In other words, if two different scenarios have identical path up to stage t, they must have the same decisions until this stage t regardless the next realizations.

Therefore, the scenario tree shown in Fig. 6.2 can also be represented as it is illustrated in Fig. 6.3.

It is possible to say that the nonanticipativity condition is modeled in an implicit way in (6.2) [43], i.e., the first stage decision, equal in all scenarios, is only represented by a unique vector x_1. On the other hand, in Fig. 6.3, the nonanticipativity condition is represented in an explicit way, given that there are nodes associates

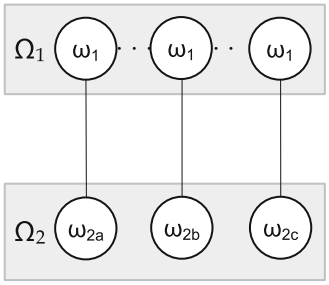

Fig. 6.3 Scenario tree representation—another possibility

with each stage for all scenarios. Therefore, the problem formulation (6.2) can be rewritten as follows:

$$\begin{aligned} f = & \min \sum_{s=1}^{S} p^s (c_1^s x_1^s + c_2^s x_2^s) \\ & \text{s.t.} : A_1^s x_1^s = b_1^s, \\ & A_2^s x_2^s + B_2^s x_1^s = b_2^s, \\ & x_1^s - x_1^{\bar{s}} = 0, \ \forall \bar{s} \in \Psi_1^s, \\ & x_1^s \geq 0, x_2^s \geq 0, \end{aligned} \tag{6.2}$$

where:

S	Total number of scenarios
s	Index of scenario, so that, s=1,...,S
Ψ_t^s	Set of all scenarios related to scenario s at stage t by the nonanticipativity condition, including itself
$\bar{s}$	Index associated with Ψ_t^s
p^s	Probability associated with the scenario s

Observe that, in this case, the nonanticipativity condition becomes a set of additional constraints explicit written which aims to ensure the same decision at stage 1. Thus, in short, there are two ways to model a stochastic optimization problem, depending on the nonanticipativity constraints management.

6.2.3 Data Structure

Based on the aspects aforementioned, the nonanticipativity constraints impacts into the problem formulation are pointed out. Initially, observe the matrix structure of problem (6.2) highlighted in Fig. 6.4.

c_1	$p_2^{\omega_{2a}} c_2$	$p_2^{\omega_{2b}} c_2$	$p_2^{\omega_{2c}} c_2$
A_1			
$B_2^{\omega_{2a}}$	$A_2^{\omega_{2a}}$		
$B_2^{\omega_{2b}}$		$A_2^{\omega_{2b}}$	
$B_2^{\omega_{2c}}$			$A_2^{\omega_{2c}}$

Fig. 6.4 Structure of the DE implicit

According to Fig. 6.4, it is easy to see that the structure of stochastic programming problems is substantially sparse. Thus, solving a stochastic problem considering this formulation, the ED implicit of stochastic problem, can require a high computational effort, especially in a multistage case.

For this reason, decomposition algorithms [45] are most often used to overcome the computational burden. By this structure, the decomposition idea is to solve each node subproblem individually, maintaining the link among them by means of some mathematical strategies. It is possible since for a feasible decision x_1, the remaining node subproblems can be solved recursively with a specific set of constraints and variables. For instance, methods based on BD principle "attack" this ED representation.

In turn, problem (6.2) matrix structure is shown in Fig. 6.5.

$p^1c_1^1$	$p^1c_2^1$	$p^2c_1^2$	$p^2c_2^2$	$p^3c_1^3$	$p^3c_2^3$
A_1^1					
B_2^1	A_2^1				
W_1^1		W_1^2		W_1^3	
		A_1^2			
		B_2^2	A_2^2		
W_2^1		W_2^2		W_1^3	
				A_1^3	
				B_2^3	A_2^3
W_3^1		W_3^2		W_3^3	

Fig. 6.5 Data structure of the DE explicit

Notice that the number of variables and constraints of the problem increases when compared to Fig. 6.4. Thus, the computation burden tends to be higher than the ED implicit modeling and, thereby, the ED explicit is basically used by some

kind of specific decomposition algorithm. It means that, in case of solving the ED of the stochastic problem, the implicit structure is more indicated.

By this approach, only the nonanticipativity constraints, W^s, connect the scenarios decisions. Thus, the ED explicit decomposition algorithm aims to decompose the problem into scenario subproblems in order to obtain smaller subproblems with variables and constraints that belong to a particular scenario s. In general, algorithms based on AL-based methods, such as PHA, use this formulation to decompose the original stochastic problem.

Finally, it is important to remark that there are algorithms that reduce the problem size by means of exploring the sparse structure of the problem without decomposing it [46, 47].

6.3 MTOP Problem

The MTOP problem presents many particularities, which make it a complex problem to be solved. It is possible to highlight the uncertainties related to some data, such as the inflows, besides other operating characteristics intrinsic to each system. In this sense, the purpose of this section is to discuss some operational aspects related to the majority of MTOP problems with predominance of hydro resources and its consequence into the problem modeling. Additionally, an idea of the problem formulation is presented.

6.3.1 Problem Features

As aforementioned, the MTOP problem usually takes part of a scheduling chain which aims to define the optimal dispatch of all system power plants and, depending on the system regulatory framework, it can also be responsible for giving an economic sign of the energy price. For this reason, besides the stochastic aspects, the modeling of the operational characteristic can be crucial to provide a satisfactory operation of the system, making the MTOP model an important tool for all system agents (operator, generator, regulator, among others).

6.3.1.1 Stochastic Aspects

Although there are other uncertainties in the MTOP problem with hydro power plant predominance, the inflows into the reservoirs modeling has received special attention. In recent years, the studies focusing on the representation of the stochastic issues into the MTOP problem are, in general, only associated with this random variable.

The modeling of these uncertainties is, in general, based in the discrete probability distribution of a historical data. Therefore, one of the most important

challenges concerning the MTOP problem is to find a way to represent rightly this original information. In other words, it is necessary to study the best strategy to represent the original infinite scenario tree by means of a representative discrete scenario tree [48], as illustrated in Sect. 6.2.

For that, some issues are extremely important, such as the number of realization in each stage, the precise representation of statistic aspects of the original probability distribution, and the quality solution of the related scenario tree [42].

As introduced before, the demand, the future fuel thermal price and the inflows into the reservoirs can be modeled as a random data in the operation planning problem context. Depending on the MTOP problem features or its main purpose, some of them can be concerned stochastic or simply represented as a deterministic data. For instance, in problems with the hydro power plant predominance, the inflow uncertainties modeling have received special attention in recent years.

6.3.1.2 General Aspects

In addition to the stochastic factors, the MTOP problem has other operational important features which must be modeled in order to represent the physics characteristics and the dynamic of the system. Among them, it is possible to highlight the streamflow balance for each reservoir, the hydro production function [16, 49], the future cost-go function, the load levels, the bounds of the exchange power flow among subsystems or zones, and water travel time among hydro plant located on the same cascade, among others.

Given that the general idea of MTOP problem formulation is detailed in the next subsection, some of these model aspects should be emphasized to make understanding of the modeling easier. For example, the hydro production function and the thermal cost function must be linearized, given that they are essentially nonlinear functions and the MTOP problem must be modeled as a linear programming problem in order to propitiate a computational resolution feasible. It is possible to say that it is not a trivial task, especially in systems where there are many power plants, such as in Brazil, Canada, Colombia, and Norway [17, 18].

Concerning the hydro power plant, the challenge is to obtain a linear function with the same operative characteristics when compared to the original nonlinear function. It can require a high effort, considering that there are hydro plants with production functions neither concave nor convex [50].

Another aspect that adds complexities into the problem formulation is representation of load levels. These levels emulate the load variation over a specific stage. In other words, the purpose consists in divide the demand into the different stage levels, in order to represent the peaks and valleys of the demand. For this reason, the number of variables increases substantially.

Obviously that the modeling depends on the system characteristics and thereby the idea in this section is to show some general features. For instance, in Brazil, the future cost-go function used in the MTOP problem, it is a result from LTOP, which

is not currently used in Canada. Moreover, aspects as evaporation and operational particularities of power plants should be represented to propitiate real-life results.

6.3.2 MTOP Problem Formulation

According to the assumptions presented in the previous subsection, it is possible to present the general idea of the problem formulation highlighting the most important constraints and variables based on the Brazilian and Canadian systems. For that, the ED implicit modeling is used and thereby can be written as follows.

6.3.2.1 Objective Function

The objective function aims to minimize the expected operational cost over a specific horizon. It is composed of the fuel thermal cost over the total MTOP horizon in case of hydrothermal system, the penalty associated with the slack of energy (deficit level cost), and the expected future cost, which depends on the reservoir level at the end of the MTOP horizon, T:

$$\text{Min} \quad F = \sum_{t=1}^{T} \sum_{\omega \in \Omega_t} p_t^{\omega_t} \sum_{u \in U} \sum_{e \in E} \left(\sum_{i \in I_e} ct_{iu} gt_{iut}^{\omega_t} + \sum_{\delta=1}^{\Delta} cd_{\delta eu} d_{\delta uet}^{\omega_t} \right) + \alpha_T, \qquad (6.3)$$

where:

F	Objective function ($)
E	Energy subsystems or operative zones
e	Index of subsystems or operative zones, so that, e=1,...,E
U	Total number of load levels
u	Index of load levels, u=1,...,U
Δ	Total number of deficit levels
δ	Index of deficit level, δ=1,..., Δ
I	Total number of thermal plants
i	Index related to thermal plants, i=1,...,I
ct_{iu}	Thermal incremental cost of the ith thermal plant ($/MWh)
gt_{iut}	Generation of thermal plant i, load level u, and stage t (MWh)
$cd_{\delta eu}$	Deficit incremental cost in δth deficit level, uth load level, and subsystem e ($/MWh)
$d_{\delta eut}$	Deficit in the δth deficit level, uth load level, subsystem e, and stage t (MWh)
α_T	Expected value of operation cost from state T+1 onwards, i.e., a future cost-go function which can be a result from LTOP ($)

6.3.2.2 Supply the Demand

These constraints aim to ensure that the sum of all generation resources is equal to the system demand for all stages, subsystems, and load levels, respecting the operational power plant bounds:

$$\sum_{i=1}^{I} gt_{iut}^{\omega_t} + \sum_{r=1}^{R} gh_{rut}^{\omega_t} + \sum_{l \in \Gamma_e} Int_{leut}^{\omega_t} + \sum_{\delta=1}^{\Delta} d_{\delta uet}^{\omega_t} = L_{uet}, \tag{6.4}$$

where:

R	Total number of hydro plants
r	Index of hydro plants, r=1,…,R
gh_{rut}	Generation of hydro plant r, load level u, and stage t (MWh)
Int	Power interchange from subsystem l to subsystem e, load level u, and stage t (MWh)
L_{uet}	System demand of the uth load level, eth subsystem, and stage t (MWh)
Γ_e	Set of subsystems linked to the subsystem e

6.3.2.3 Stream-Flow Balance

The stream-flow balance constraint ensures that the final volume at the end of a specific stage must be equal to the initial volume plus inflows and minus the total released outflow, regardless casual losses or evaporations:

$$v_{rt}^{\omega_t} - v_{r,t-1}^{\omega_{t-1}} + C\sum_{u}^{U}\left[q_{rut}^{\omega_t} + sp_{rut}^{\omega_t} - \sum_{m \in M_i}\left(q_{mu,t-\tau_{mr}}^{\omega_t} + sp_{mu,t-\tau_{mr}}^{\omega_t}\right)\right] = Cy_{rt}^{\omega_t}, \tag{6.5}$$

where

$v_{rt}^{\omega_t}$	Volume of rth hydro plant reservoir at the end of stage t (hm^3) considering a specific node ω
$q_{rut}^{\omega_t}$	Discharge outflow of hydro plant r, load level u, and stage t (m^3/s)
$sp_{rut}^{\omega_t}$	Spillage of hydro plant r, load level u, and stage t (m^3/s)
$y_{rt}^{\omega_t}$	Incremental inflow of rth hydro plant reservoir and stage t (m^3/s)
C	Conversion factor of water discharge unit (m^3/s) in volume units (hm^3)
M_r	Set of upstream reservoirs from hydro plant r
m	Index of upstream reservoirs, m=1,…,M
τ_{mr}	Number of stages that the total outflow of a hydro plant m takes to reach the downstream hydro plant r

6.3.2.4 Hydro Piecewise Linear Function

$$gh_{rut}^{\omega_t} = \Theta\left(v_{rt}^{\omega_t}, q_{rut}^{\omega_t}, sp_{rut}^{\omega_t}\right). \tag{6.6}$$

The hydro production function is a nonlinear function which related the net head, the generator units efficiency, and the release of the power plant [51]. Nevertheless, in the MTOP problem context, the hydro production is, in general, modeled as piecewise linear function which can depend on the released outflow, the volume, and the spillage, as detailed in [50], or simply depend on the release outflow and the net head. In literature, alternatives approaches are also broadly used to represent the hydro production function, as discussed in [16, 52], because it is definitely not a trivial task to obtain a good hydro production function linearization.

6.3.2.5 Future Cost-Go Function

This function can be given by the LTOP problem, if there is a scheduling chain models or built taking into the future level of the reservoirs. Thus, it estimates the expected future cost. In short, it is a piecewise linear function depending on the volume of water in the reservoirs at the end of the planning horizon, T. In other words, it represents the expected future cost from $T + 1$:

$$\alpha_T - \sum_{\omega \in \Omega_T} \sum_{r \in R} \gamma_{rj} v_{rt}^{\omega_T} \geq \alpha_r^0, \tag{6.7}$$

where:

J	Number of linear constraints used in the piecewise future cost-go function
j	Index related to the piecewise future cost function, with j=1,...,J
γ_{rj}	Slope associated with jth linear segment of the future cost-go function associated with hydro plant r (MW/hm3)

6.3.2.6 Bounds

The individual variable limits must be also considered which aims to determine the physic operational of the system and its generator resources, such as the exchange bounds, released outflow, and reservoir volumes:

$$\begin{aligned} Int_{leut}^{\min} \leq Int_{leut}^{\omega_t} \leq Int_{leut}^{\max}, \\ v_r^{\min} \leq v_{rt}^{\omega_t} \leq v_r^{\max} \\ 0 \leq q_{rut}^{\omega_t} \leq q_r^{\max}, \\ 0 \leq sp_{rut}^{\omega_t} \leq sp_r^{\max}, \\ 0 \leq gt_{iut}^{\omega_t} \leq gt_i^{\max}, \\ 0 \leq gh_{rut}^{\omega_t} \leq gh_r^{\max}. \end{aligned} \tag{6.8}$$

Thus, the (6.3)–(6.8) formulation summarizes the MTOP problem formulation. Obviously that in case of a pure hydro systems or other system peculiarities, some variable considered above must be disregarded or adapted to the system reality.

6.4 Decomposition Algorithms

Based on the features and challenges associated with the MTOP problem resolution presented so far, the study of the stochastic decomposition algorithms becomes mandatory [53, 54]. Depending on the planning horizon and the operational features which are directly related to the problem size, some decomposition algorithms are more appropriated. It is possible to highlight the algorithm based on BD [14, 27, 55, 56, 57, 58], AL [23, 24, 38, 40, 41, 59, 60, 61, 62], and its particularities [63], among others [64].

Thus, in this section, a simplified MTOP problem example is used to show the main features of the two algorithms broadly used to solve this kind of problem. The idea is to discuss the main differences between the algorithms and its properties.

For that, consider the simple hydrothermal system and the inflow scenario tree of the problem illustrated in Fig. 6.6.

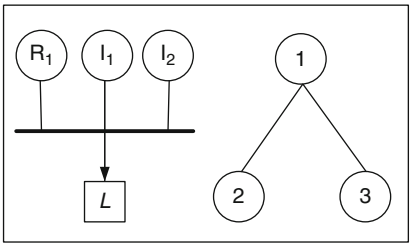

Fig. 6.6 Hydrothermal system and scenario tree

Notice that the hydrothermal system has only one hydro plant, R, and two thermal plants, I_1 and I_2. In this example, the hydro production function is represented by a simple linear function depending on the reservoir released outflow as it is shown in (6.9). With these generator system resources, the objective is to supply a constant demand L with the minimum expected cost over two-stage planning horizon, considering only the inflow uncertainties:

$$gh = q. \tag{6.9}$$

The inflow scenario tree has two possible inflow realizations in the second stage. As a consequence, there are three nodes and two scenarios, given that a scenario can be defined as a complete path from node 1 in the first stage to a specific node in the last (second) stage. The inflows scenario tree data and additional information concerning the hydrothermal system are presented in Tables 6.1 and 6.2, respectively.

Table 6.1 Data of scenario tree?

Stage (t)	Node (ω)	Probability (p)	Inflow [y (hm^3)]	Demand [L (MW)]
1	1	1	300	450
2	2	0.5	100	450
2	3	0.5	400	450

Table 6.2 Additional information of the hydrothermal system

Power plant	Incremental cost [c ($/MW)]	Maximum generation [gt or gh (MW)]	Initial volume [v_0 (hm^3)]
I_1	50	60	–
I_2	200	450	–
H_1	–	450	150

According to the data shown, it is possible to write the general stochastic optimization problem formulation by means of the implicit ED modeling, (6.10)–(6.13), where, for sake of simplification, the variable bounds are omitted:

$$\begin{aligned}\min F = {} & 50gt_{11}^1 + 200gt_{21}^1 + 0.5(50gt_{12}^2 + 200gt_{22}^2) \\ & + 0.5(50gt_{12}^3 + 200gt_{22}^3)\end{aligned} \tag{6.10}$$

$$\begin{aligned}\text{s.t.:} \quad & gt_{11}^1 + gt_{21}^1 + q_{11}^1 = 450 \qquad (node1) \\ & v_{11}^1 + q_{11}^1 = 150 + 300 \qquad (node1)\end{aligned} \tag{6.11}$$

$$\begin{aligned}& gt_{12}^2 + gt_{22}^2 + q_{12}^2 = 450 \qquad (node2) \\ & v_{12}^2 + q_{12}^2 - v_{11}^1 = 100 \qquad (node2)\end{aligned} \tag{6.12}$$

$$\begin{aligned}& gt_{12}^3 + gt_{22}^3 + q_{12}^3 = 450 \qquad (node3) \\ & v_{12}^3 + q_{12}^3 - v_{11}^1 = 400 \qquad (node3)\end{aligned} \tag{6.13}$$

The objective function (6.10) aims to minimize the expected value of thermal production over the horizon taking into account the node probabilities. Additionally, there are set of constraints and variables associated with each node, given by (6.11)–(6.13). The constraints represent the demand supply and the stream-flow balance of the reservoir R_1.

Notice that, in the MTOP problem context, the link between different stages is performed by the reservoir volume, i.e., the initial volume for all nodes in the second stage corresponds to the final volume in the first node. For this reason, it is possible to conclude that the reservoir storage is a state variable of the problem. It is an important MTOP characteristic given that the optimal decision is related to the reservoir storage levels. In other words, once defined the volume storage in each stage, the other variable are consequently determined.

As already mentioned, the (6.10)–(6.13) formulation is the simplest way to represent a stochastic optimization problem. According to Sect. 6.2, it is also possible to write the variables and constraints with respect to scenarios, the explicit ED, as follows:

$$\begin{aligned}\min F = {} & 0.5(50gt_{11}^1 + 200gt_{21}^1 + 50gt_{12}^1 + 200gt_{22}^1) \\ & +0.5(50gt_{11}^2 + 200gt_{21}^2 + 50gt_{12}^2 + 200gt_{22}^2)\end{aligned} \tag{6.14}$$

s.t. :

$$\begin{aligned}&\text{scenario1}: \\ &gt_{11}^1 + gt_{21}^1 + q_{11}^1 = 450, \\ &v_{11}^1 + q_{11}^1 = 150 + 300, \\ &gt_{12}^1 + gt_{22}^1 + q_{12}^1 = 450, \\ &v_{12}^1 + q_{12}^1 - v_{11}^1 = 100,\end{aligned} \tag{6.15}$$

$$\begin{aligned}&\text{scenario2}: \\ &gt_{11}^2 + gt_{21}^2 + q_{11}^2 = 450, \\ &v_{11}^2 + q_{11}^2 = 150 + 300, \\ &gt_{12}^2 + gt_{22}^2 + q_{12}^2 = 450, \\ &v_{12}^2 + q_{12}^2 - v_{11}^2 = 400,\end{aligned} \tag{6.16}$$

$$v_{111} - v_{112} = 0. \tag{6.17}$$

As discussed in the formulation (6.2), the nonanticipativity constraints are added to the problem formulation in order to guarantee the same decision in all scenarios that share the same nodes until stage T-1. In this case, the purpose is to ensure the unique decision for both scenarios that share the same realizations in stage 1. Due to the reservoir volume is a unique state variable of the MTOP problem, using it to represent the nonanticipativity constraints can be an interesting approach to reduce the problem size, as discussed in [22].

It is important to remark that the differences between the formulations are essential for better understanding of the decomposition algorithm strategies presented later in this chapter. In summary, the BD-based algorithm uses the (6.10)–(6.13) representation in order to get node subproblems and, in turn, the AL-based algorithm explores (6.14)–(6.17) structure to obtain scenario subproblems linked by nonanticipativity constraints.

Before discussing the decomposition algorithms, it is convenient to mention that once the MTOP is modeled as a convex optimization problem [65], both formulations of the MTOP problem present the same optimal objective function value, as it is shown in Table 6.3.

Table 6.3 ED solution

Stages	Node	gt_1	gt_2	gh	v
1	1	60	0	390	60
2	2	60	230	160	0
	3	0	0	450	10
Objective function($)27,500					

6.4.1 Benders Decomposition Algorithm

In this subsection, the idea is to present some basic notion of the BD-based algorithms. Given that the test problem illustrated in Fig. 6.6 is a two-stage stochastic problem, the L-shaped algorithm [27, 66] is detailed. Nevertheless, the theoretical concept can be extended to the multistage algorithm, called NDA, which is, for instance, currently used to solve the Brazilian MTOP problem.

In summary, this algorithm solves the first stage subproblem and manages the remaining stages as other subproblems, solving them recursively. Thus, it moves down and up the scenario tree, also called forward and backward recursions, by means of solving each node subproblem passing forward information to immediate successors to form the right-hand side and passing backward to its ancestors in the form of feasibility cuts (cutting planes) [56].

Thus, the L-shaped resulting subproblems can be written to each node as follows:

$$\begin{aligned} \min f_1 &= 50gt_{11}^1 + 200gt_{21}^1 + \alpha \\ \text{s.t.}: \quad & gt_{11}^1 + gt_{21}^1 + q_{11}^1 = 450 \\ & v_{11}^1 + q_{11}^1 = 150 + 300. \end{aligned} \tag{6.18}$$

$$\begin{aligned} \min f_2 &= 50gt_{12}^2 + 200gt_{22}^2 \\ \text{s.t.}: \quad & gt_{12}^2 + gt_{22}^2 + q_{12}^2 = 450 \\ & v_{12}^2 + q_{12}^2 = v_{11}^2 + 100. \end{aligned} \tag{6.19}$$

$$\begin{aligned} \min f_3 &= 50gt_{12}^3 + 200gt_{22}^3 \\ \text{s.t.}: \quad & gt_{12}^3 + gt_{22}^3 + q_{12}^3 = 450 \\ & v_{12}^3 + q_{12}^3 = v_{11}^1 + 400. \end{aligned} \tag{6.20}$$

Notice that the objective function of subproblem (6.18) has a new variable α, which aims to represent the expected value of the second stage according to the first stage decisions. Consequently, it is updated in each algorithms iteration, as detailed below.

Then, once writing the node subproblems, the L-shaped decisions are sequential. It means that the first stage subproblem (6.18) must be solved to obtain the value

of the reservoir volume and, therefore, it uses this information to solve the second stage subproblems: (6.19) and (6.20). This step is called forward and it is briefly illustrated in Fig. 6.7, where the Lagrangian multiplier π associated with the stream-flow balance constraint is also highlighted.

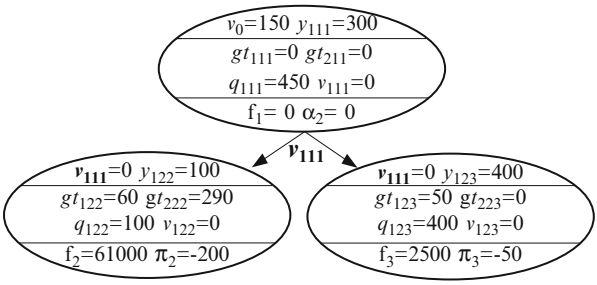

Fig. 6.7 First forward step

After finishing the forward step, by the L-shaped algorithm, it is necessary to compare the total cost of the first stage, the Lower Cost (LC), with the sum of individual cost of each stage f disregarding the future cost, the Upper Cost (UC), as it is detailed in (6.21). If the future cost α accurately represents the second stage cost taking into account a stopping criterion υ, the algorithm is stopped. Based on the results presented in Fig. 6.7, it is easy to notice that the LC, equal to first node cost f_1, is different when compared to the UC showed in (6.21) (in the first iteration, there is no second stage approximation; i.e., α equal to zero):

$$\begin{aligned} &\mathrm{UC} = p_1\,(f_1 - \alpha) + p_2 f_2 + p_3 f_3, \\ &\therefore \mathrm{UC} = 31{,}750. \end{aligned} \tag{6.21}$$

Thus, the algorithms' next step, called backward, aims to build the expected cost of the second stage. For this, the expected value of the Lagrange multipliers related to the stream-flow balance constraints[3] are used to obtain a Benders cut, which represents the lower bound approximation of the second stage expected cost. In this context, it is possible to write the feasibility cut as follows:

$$\begin{aligned} \alpha - \alpha^* &\geq \bar{\pi}\,(v_{111} - v^*_{111})\,, \\ \alpha - (p_2 f_2 + p_3 f_3) &\geq (p_2 pi_2 + p_3 pi_3)(v_{111} - 0), \\ \therefore \alpha + 125 v_{111} &\geq 31{,}750. \end{aligned} \tag{6.22}$$

This constraint is thereby added in the first stage problem (6.18) and, afterwards, a new forward recursion must be initialized. The backward and forward steps should

[3] More precisely, it represents the derivative of dual cost function in relation to volume variable v111, in (R$/hm3).

be continued until the stopping criterion is reached. It is an important remark that the size of the first stage subproblem is increased iteratively, which can cause eventually some impact to the algorithm efficiency.

The convergence process evolution is shown in Fig. 6.8 and the second stage expected cost function α, also called the future cost-go function, is illustrated in Fig. 6.9.

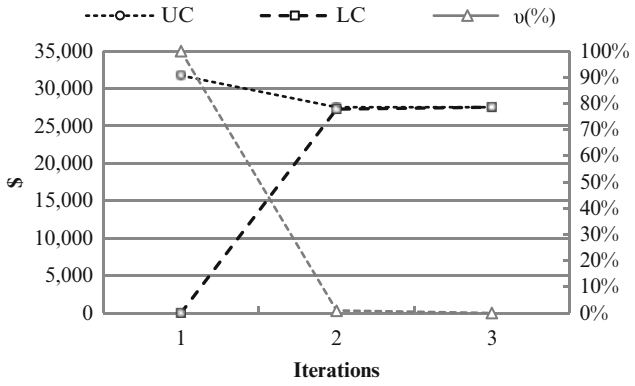

Fig. 6.8 Iterative process

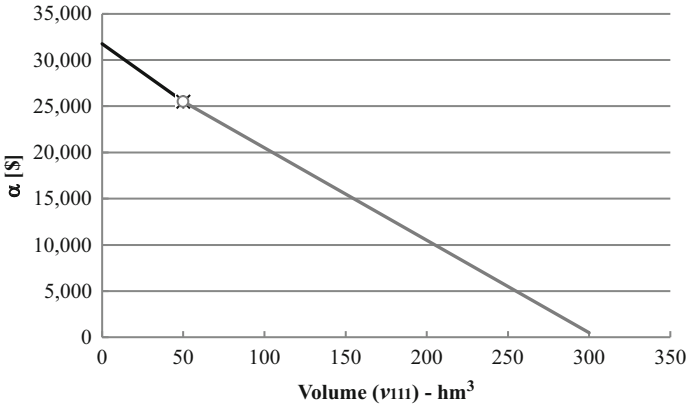

Fig. 6.9 First stage future cost-go function with two linear approximations resulting from two backward steps

Considering that the results are equal to those presented in Table 6.3, two L-shaped aspects should be highlighted: (i) the convergence was attained when the UC and LC were equal (it is an academic example and thereby the stopping criterion υ is equal to zero); (ii) given that three iterations were necessary to reach the stopping criterion (three forward steps), two future cost-go function approximations were built (two backward steps).

6.4.2 Augmented Lagrangian-Based Algorithm

Unlike BD-based algorithms, the AL-based algorithms applied to solve the multistage stochastic problem, such as the PHA, aim to explore the structure of the explicit ED (6.14)–(6.17) in order to decompose it into scenario subproblems. The purpose is to relax the nonanticipativity constraint such that each subproblem presents only variables that belong to a particular scenario.

Thus, following the PHA idea recently applied to solve the Brazilian MTOP problem [22, 23, 24], to obtain the independent scenario subproblems, the first algorithm step is to make the nonanticipativity decisions equal to a constant value corresponding to the expected value of nonanticipativity variables which must be updated iteratively. Thus, regarding the nonanticipativity constraint (6.17), in the PHA, it becomes

$$v_{111} - \bar{v} = 0, \;\; v_{112} - \bar{v} = 0. \tag{6.23}$$

Consequently, problems (6.14)–(6.17) can be rewritten, replacing (6.17) by (6.23), as follows:

$$\begin{aligned} \min F \; = \; & 0.5(50gt_{111} + 200gt_{211} + 50gt_{121} + 200gt_{221}) \\ & +0.5(50gt_{112} + 200gt_{212} + 50gt_{122} + 200gt_{222}) \end{aligned} \tag{6.24}$$

$$\begin{aligned} \text{s.t.}: \quad & \text{scenario1}: \\ & gt_{111} + gt_{211} + q_{111} = 450, \\ & v_{111} + q_{111} = 150 + 300, \\ & gt_{121} + gt_{221} + q_{121} = 450, \\ & v_{121} + q_{121} - v_{111} = 100, \end{aligned} \tag{6.25}$$

$$\begin{aligned} & \text{scenario2}: \\ & gt_{112} + gt_{212} + q_{112} = 450, \\ & v_{112} + q_{112} = 150 + 300, \\ & gt_{122} + gt_{222} + q_{122} = 450, \\ & v_{122} + q_{122} - v_{112} = 400, \end{aligned} \tag{6.26}$$

$$\begin{aligned} & v_{111} - \bar{v} = 0, \\ & v_{112} - \bar{v} = 0. \end{aligned} \tag{6.27}$$

By the PHA, the next step is to relax (6.28) taking into account the AL concept in order to obtain the following separable problem:

$$\begin{aligned} \Theta \; = \; & \min 0.5(50gt_{111} + 200gt_{211} + 50gt_{121} + 200gt_{221}) \\ & +0.5(50gt_{112} + 200gt_{212} + 50gt_{122} + 200gt_{222}) \\ & +\pi_1(v_{111} - \bar{v}) + \pi_1(v_{112} - \bar{v}) \\ & +\frac{\mu}{2}\|v_{111} - \bar{v}\|^2 + \frac{\mu}{2}\|v_{112} - \bar{v}\|^2 \end{aligned} \tag{6.28}$$

$$\text{s.t.}: \quad gt_{111} + gt_{211} + q_{111} = 450,$$

$$v_{111} + q_{111} = 150 + 300, \tag{6.29}$$

$$\begin{aligned} &gt_{121} + gt_{221} + q_{121} = 450,\\ &v_{121} + q_{121} - v_{111} = 100,\\ &gt_{112} + gt_{212} + q_{112} = 450,\\ &v_{112} + q_{112} = 150 + 300,\\ &gt_{122} + gt_{222} + q_{122} = 450,\\ &v_{122} + q_{122} - v_{112} = 400. \end{aligned} \tag{6.30}$$

Therefore, considering fixed values for $\bar{v}$, π_1, π_2, and μ, it is possible to solve problems (6.28)–(6.30) by means of the resolution of each scenario subproblem individually:

$$\Theta = \theta_1 + \theta_2. \tag{6.31}$$

where each scenario subproblem θ can be written as follows:

$$\begin{aligned} \theta_1 = &\min 50gt_{111} + 200gt_{211} + 50gt_{121} + 200gt_{221}\\ &+ \pi_1(v_{111} - \bar{v}) + \frac{\mu}{2}\|v_{111} - \bar{v}\| \end{aligned} \tag{6.32}$$

$$\begin{aligned} \text{s.t.}: \quad &gt_{111} + gt_{211} + q_{111} = 450,\\ &v_{111} + q_{111} = 150 + 300,\\ &gt_{121} + gt_{221} + q_{121} = 450,\\ &v_{121} + q_{121} - v_{111} = 100. \end{aligned} \tag{6.33}$$

$$\begin{aligned} \theta_2 = &\min 50gt_{112} + 200gt_{212} + 50gt_{122} + 200gt_{222}\\ &+ \pi_2(v_{112} - \bar{v}) + \frac{\mu}{2}\|v_{112} - \bar{v}\|^2 \end{aligned} \tag{6.34}$$

$$\begin{aligned} \text{s.t.}: \quad &gt_{112} + gt_{212} + q_{112} = 450,\\ &v_{112} + q_{112} = 150 + 300,\\ &gt_{122} + gt_{222} + q_{122} = 450,\\ &v_{122} + q_{122} - v_{112} = 400. \end{aligned} \tag{6.35}$$

The first iteration solution, considering μ equal to 1, the initial target $\bar{v}$ equal to 0 and the Lagrangian multipliers also equal to 0, is presented in Table 6.4.

Table 6.4 Scenario subproblems solution—first iteration

Stages	Scenario	gt_1	gt_2	gh	v
1	1	60	0	390	60
	2	60	230	160	0
2	1	0	0	450	0
	2	50	0	400	0
Objective function(R$) 27,250					

After solving both scenario subproblems, the next algorithm step consists in calculating the new value of the nonanticipativity condition (6.23), the volume average, and update the Lagrangian multipliers using, for instance, the gradient method [5], as described by (6.36):

$$\pi_s^{iter+1} = \pi_s^{iter} + \mu(v_{11s} - \bar{v}). \tag{6.36}$$

Therefore, the new values of the Lagrangian multipliers are shown in Table 6.5.

Table 6.5 Lagrangian multipliers—first iteration

Scenario	π
1	30
2	-30

Finally, the stopping criterion must be assessed by D_{iter}, as proposed by [22]. It means that while D_{iter} is bigger than error ε, the algorithm process should be continued:

$$D_{iter} = \mathrm{E}\left[\sum_{s=1}^{S=2}\left(\|v_{r1s} - \bar{v}\|^2 + \frac{1}{\mu^2}\left\|\pi_{s,iter+1} - \pi_{s,iter}\right\|^2\right)\right] < \varepsilon, \tag{6.37}$$

where $E[\cdot]$ represents the expected value.

Figure 6.10 illustrates the "D_{iter} track" iteration after iteration until the stopping criterion is satisfied (in this example problem, D_{iter} is less than 0.1). As the L-shaped algorithm, PHA presented the same optimal results described in Table 6.3 after seven iterations.

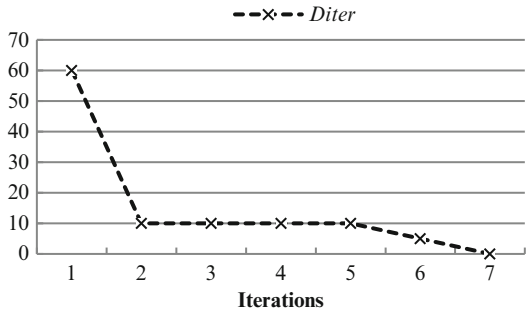

Fig. 6.10 Over the iterative process

Although not mentioned so far, one aspect is essential for the AL-based algorithm success: the choice of the suitable penalty parameter μ. Therefore, instead of using a fixed penalty parameter value during the optimization process as done in this test

example, the following update iteratively strategy (6.38) [24] seems more interesting for large-scale cases:

$$\mu_{iter+1} = \mu_{iter} \left\{ \mathrm{E} \left[\sum_{s}^{S} \sum_{t=1}^{T-1} \left(\frac{\| v_{rts} - \bar{v} \|^2}{\left(v_{rt,iter}^{\max} - v_{rt,iter}^{\min} + 1 \right)} \right) \right] \sigma + 1 \right\}. \tag{6.38}$$

According to the algorithm features described above, it is possible to notice the differences between both algorithms presented here . Both algorithms present its particularities as well as advantages and disadvantages. With respect to the PHA, it can mitigate some BD-based algorithms or cutting planes method disadvantages [27] (initial iteration is often inefficient and iterations may become degenerate at the end of the process), given that it provides quadratic subproblems and it is still sensitive to the use of warm start techniques [25], though some other heuristics are required, such as the choice of the penalty parameter.

Finally, it is possible to say that the scenario decomposition algorithms are easily implemented using parallel processing, once it has a weak link between scenario subproblems.

6.5 Conclusions

This chapter focused on the MTOP problem of hydrothermal systems. In summary, this problem aims to define the dispatch of the power plants and the spot energy price depending on system framework, taking into account the uncertainties associates with some problems data.

Given that it is a stochastic problem, important aspects related to the stochastic programming were addressed in this chapter, such as the random variable representation, the solution algorithms, and other theoretical aspects especially associated with the MTOP problem.

With respect to the random variables, the impact of the scenario tree representation into the problem data structure was emphasized, which makes the solution algorithms steps presentation easier. In addition, the stochastic problem formulation related to each scenario tree structure was discussed.

The MTOP problem formulation was also stressed, highlighting some operational features and its particularities. In this context, the current and future challenges taking into account the MTOP modeling were also discussed.

Finally, an academic example was used to help the description of two different solution algorithms with distinct characteristics: L-shaped and an augmented Lagrangian-based algorithm. Both are broadly used in literature to solve these kinds of problems.

References

1. Wood AJ, Wollenberg BF (1996) Power generation, operation, and control, 2nd edn. Wiley, New York
2. Pereira MVF (1989) Optimal stochastic operations scheduling of large hydroelectric systems. Electr Power Energ Syst 11:161–169
3. Sjelvgren D et al (1989) Large-scale non-linear programming applied to operations planning. Int J Electr Power Energ Syst 11:213–217
4. Pereira MVF, Pinto LMVG (1984) Operation planning of large-scale hydroelectrical systems. In: 8th power system computation conference (PSCC), Helsinki
5. Redondo NJ, Conejo AJ (1999) Short-term hydro-thermal coordination by lagrangian relaxation: solution of the dual problem. IEEE Trans Power Syst 14:89–95
6. Cabero J et al (2005) A medium-term integrated risk management model for a hydrothermal generation company. IEEE Trans Power Syst 20:1379–1388
7. Morton D (1996) An enhanced decomposition algorithm for multistage stochastic hydroelectric scheduling. Ann Oper Res 64:211–235
8. Escudero LF (2000) WARSYP: a robust modeling approach for water resources system planning under uncertainty. Ann Oper Res 95:313–339
9. Watchorn CW (1967) Inside hydrothermal coordination. IEEE Trans Power Apparat Syst PAS-86:106–117
10. Terry LA et al (1986) Coordinating the energy generation of the brazilian national hydrothermal electrical generating system. Interfaces 16:16–38
11. Sherkat VR et al (1988) Modular and flexible software for medium and short-term hydrothermal scheduling. IEEE Trans Power Syst 3:1390–1396
12. Enamorado JC et al (2000) Multi-area decentralized optimal hydro-thermal coordination by the Dantzig–Wolfe decomposition method. In: IEEE in Power Engineering Society Summer Meeting, Seattle, WA, vol 4, pp 2027–2032
13. Zambelli M et al (2006) Deterministic versus stochastic models for long term hydrothermal scheduling. In: IEEE in Power Engineering Society General Meeting, Montreal, QC, pp 7
14. Dias BH et al (2010) Stochastic dynamic programming applied to hydrothermal power systems operation planning based on the convex hull algorithm. Math Probl Eng 2010:20
15. Aouam T, Zuwei Y (2008) Multistage stochastic hydrothermal scheduling. In: IEEE International Conference on Electro/Information Technology, EIT, pp 66–71
16. Bortolossi HJ et al (2002) Optimal hydrothermal scheduling with variable production coefficient. Math Methods Operations Res 55:11–36
17. Gjelsvik A, Wallace SW (1996) Methods for stochastic medium-term scheduling in hydro-dominated power systems. Report EFI TR A 4438, Electric Power Research Institute
18. Johannesen A, Flataboe N (1989) Scheduling methods in operation planning of a hydro-dominated power production system. Int J Electr Power Energ Syst 11:189–199

19. Labadie J (2004) Optimal operation of multireservoir systems: state-of-the-art review. J Water Resour Plann Manag 130:93–111
20. Botnen OJ et al (1992) Modelling of hydropower scheduling in a national/international context. Proceedings Hydropower '92, A.A. Balkema
21. Belsnes MM et al (2003) Quota modeling in hydrothermal systems. In: IEEE bologna power tech conference proceedings, bologna (italy), vol 3. p 6
22. Gonçalves REC et al (2011) Applying different decomposition schemes using the progressive hedging algorithm to the operation planning problem of a hydrothermal system. Electr Power Syst Res 83:19–27
23. Santos MLL et al (2009) Practical aspects in solving the medium-term operation planning problem of hydrothermal power systems by using the progressive hedging method. Int J Electr Power Energ Syst 31:546–552
24. Gonçalves REC et al (2011) Exploring the progressive hedging characteristics in the solution of the medium term operation planning problem. In: 17th power systems computation conference (PSCC), Sweden, 2011
25. Santos MLL et al (2008) Solving the short term operating planning problem of hydrothermal systems by using the progressive hedging method. In: 16th power system computation conference (PSCC), Glasgow, 2011
26. Gonçalves REC et al (2011) Comparing stochastic optimization methods to solve the medium-term operation planning problem. Comput Appl Math 30:289–313
27. Birge JR, Louveaux F (1997) Introduction to stochastic programming. Springer, New York
28. Greengard C, Ruszczynski AP (2002) Decision making under uncertainty: energy and power. Springer, New York
29. Cunha SHF et al (1997) Modelagem da Produtividade Variável de Usinas Hidrelétricas com base na Construção de uma Função de Produção Energética," in XII Simposio Brasileiro de Recursos Hidricos, ABRH, pp 391–397
30. Rosa C, Ruszczyinski A (1996) On augmented lagrangian decomposition methods for multistage stochastic programs. Ann Oper Res 64:289–309
31. Ruszczynski A (1999) Some advances in decomposition methods for stochastic linear programming. Ann Oper Res 85:153–172
32. Benders JF (1962) Partitioning procedures for solving mixed variables programming problems. Numer Math 4:238–252
33. Jacobs J et al (1995) SOCRATES: a system for scheduling hydroelectric generation under uncertainty. Ann Oper Res 59:99–133
34. Anukal Chiralaksanakul BS (2003) Monte carlo methods for multi-stage stochastic programs. Doctorate, Department of Philosophy, University of Texas at Austin, Austin
35. Liu H et al (2009) Portfolio management of hydropower producer via stochastic programming. Energ Convers Manag 50:2593–2599
36. Shapiro A (2011) Analysis of stochastic dual dynamic programming method. Eur J Oper Res 209:63–72

37. da Costa JP et al (2006) Reduced scenario tree generation for mid-term hydrothermal operation planning. In: International conference on probabilistic methods applied to power systems, PMAPS, Stockholm - Sweden, pp 1–7
38. Ruszczyinski A (1992) Augmented lagrangian decomposition for sparse convex optimization. Ann Oper Res
39. Mulvey J, Vladimirou H (1991) Applying the progressive hedging algorithm to stochastic generalized networks. Ann Oper Res 31:399–424
40. Watson J-P, Woodruff D (2011) Progressive hedging innovations for a class of stochastic mixed-integer resource allocation problems. Comput Manag Sci 8(4): 355–370
41. Rockafellar RT, Wets RJB (1991) Scenarios and policy aggregation in optimization under uncertainty. Math Oper Res 16:119–147
42. Shapiro A, Philpott A (2007) A tutorial on stochastic programming. Technical report. INFORMS Tutorials in Operations Research, Technical report 2007
43. Fourer R, Lopes L (2006) A management system for decompositions in stochastic programming. Ann Oper Res 142:99–118
44. Medina J et al (1998) A comparison of interior-point codes for medium-term hydro-thermal coordination. IEEE Trans Power Syst 13:836–843
45. Ruszczyinski A, Shapiro A (2003) HandBooks in operations research and management science. 1st edn. vol 10, Elsevier Science B.V., Netherlands
46. Christoforidis M et al (1996) Long-term/mid-term resource optimization of a hydrodominant power system using interior point method. IEEE Trans Power Syst 11:287–294
47. Fuentes-Loyola R et al (2000) A performance comparison of primal-dual interior point method vs lagrangian relaxation to solve the medium term hydrothermal coordination problem. In: IEEE power engineering society summer, Seattle, pp 2255–2260
48. Dupacova J et al (2000) Scenarios for multistage stochastic programs. Ann Oper Res 100:25–53
49. Garcia-Gonzalez J, Castro GA (2001) Short-term hydro scheduling with cascaded and head-dependent reservoirs based on mixed-integer linear programming. In: IEEE Portugal power tech proceedings, Porto - Portugal, vol 3, p 6
50. Diniz AL, Maceira MEP (2008) A four-dimensional model of hydro generation for the short-term hydrothermal dispatch problem considering head and spillage effects. IEEE Trans Power Syst 23:1298–1308
51. Finardi EC, da Silva EL (2006) Solving the hydro unit commitment via dual decomposition and sequential quadratic programming. IEEE Trans Power Syst 21:835–844
52. Breton M et al (2004) Accounting for losses in the optimization of production of hydroplants. IEEE Trans Energ Convers 19:346–351
53. Ruszczyinski A (1997) Decomposition methods in stochastic programming. Math Program 79:333–353
54. Pennanen T, Kallio M (2006) A splitting method for stochastic programs. Ann Oper Res 142:259–268

55. Gorenstin BG et al (1991) Stochastic optimization of a hydro-thermal system including network constraints. In: Conference proceedings on power industry computer application conference, Baltimore, MD, pp 127–133
56. Pereira MVF, Pinto LMVG (1985) Stochastic optimization of a multireservoir hydroelectric system: a decomposition approach. Water Resour Res 21:779–792
57. Philpott AB, de Matos VL (2012) Dynamic sampling algorithms for multi-stage stochastic programs with risk aversion. Eur J Oper Res 218:470–483
58. Matos VLD, Finardi EC (2012) A computational study of a stochastic optimization model for long term hydrothermal scheduling. Int J Electr Power Energ Syst 43:1443–1452
59. Al-Agtash S (2001) Hydrothermal scheduling by augmented lagrangian: consideration of transmission constraints and pumped-storage units. IEEE Power Eng Rev 21:58–59
60. Cristian Finardi E, Reolon Scuzziato M (2013) Hydro unit commitment and loading problem for day-ahead operation planning problem. Int J Electr Power Energ Syst 44:7–16
61. Diniz AL et al (2007) Assessment of lagrangian relaxation with variable splitting for hydrothermal scheduling. In: IEEE power engineering society general meeting, pp 1–8.
62. Helgason T, Wallace SW (1991) Approximate scenario solutions in the progressive hedging algorithm: a numerical study. Ann. Oper. Res. 31:425–444
63. LeMaréchal C et al (1996) Bundle methods applied to the unit commitment problem. Syst Model Optim, IFIP – The International Federation for Information Processing, 395–402
64. Berger AJ et al (1994) An extension of the DQA algorithm to convex stochastic programs. SIAM J Optim 4:735–753
65. Nocedal J, Wright SJ (2006) Numerical optimization, 2nd edn. Springer, New York
66. Slyke RMV Wets R (1969) L-Shaped linear programs with applications to optimal control and stochastic programming. SIAM J Appl Math 17:638–663

Chapter 7
Stochastic Optimization of Power Generation and Storage Management in a Market Environment

Andreas Eichhorn

Abstract This chapter provides an overview of practically applying mathematical optimization techniques to short-term and medium-term planning of a power generation system in a market environment. The considered power generating system may contain thermal plants (gas or coal fired), hydro power plants, new renewables, as well as dedicated energy storages (e.g., gas storages, hydro reservoirs). We argue that stochastic optimization is an appropriate modeling framework in order to take into account the uncertainty of input data (such as natural hydrologic inflows and energy market prices), market decision structures, as well as the optional character of power generating units and energy storages.

7.1 Introduction

In this chapter we take the perspective of a power producer, i.e., a company running a power generation system consisting of *assets* such as thermal power plants (gas or coal fired), hydro power plants, new renewables (e.g., wind and solar power), as well as dedicated energy storages (e.g., gas storages, hydro reservoirs). We are particularly interested in controllable assets, i.e., assets featuring some degree of *flexibility*. We consider the case that all these assets are located in an area with a liquid electricity market such as the common market area of Germany and Austria. We assume that this electricity market is big and liquid enough and that there is no monopolist who has much more market power than all the other market participants. We do not take into account aspects of power transmission, i.e., we assume a fully *unbundled* market where the problems of transmission and distribution are completely left to the system operator.

A. Eichhorn (✉)
VERBUND Trading AG, Am Hof 6a, 1010 Vienna, Austria, http://www.verbund.com
e-mail: Andreas.Eichhorn@verbund.com

R.M. Kovacevic et al. (eds.), *Handbook of Risk Management in Energy Production and Trading*, International Series in Operations Research & Management Science 199,
DOI 10.1007/978-1-4614-9035-7_7, © Springer Science+Business Media New York 2013

Running power generating assets includes the important task of planning and scheduling them, i.e., to make use of their flexibility. The company can and has to decide permanently which power plant shall generate how much electricity in the respective present and in the future, and how much energy shall be stored into or released from the available energy storages. The flexibility of storages, however, links together decisions for different time periods: stored energy can be used in the present (here and now) *or* at some time in the future. Clearly, since companies typically aim at maximizing revenues and contribution margins, optimization techniques are interesting for all these scheduling decisions.

In a market environment, the major amount of electricity is scheduled according to market prices. In a typical market environment there are several markets where electricity can be traded at; probably the most important ones are the day-ahead and intraday *spot markets*; in addition there are forward and futures markets for the delivery of electricity in *bundles of hours* in the future, such as all hours (*base*) or selected (e.g., *peak*) hours within a week, month, years, etc.; and there is also a market for options (e.g., options on forwards or on spot) where in addition to electricity some kinds of flexibility are traded. Moreover, there may be reserve markets (ancillary service markets) for short-term flexibility; and there is of course the possibility to sell (buy) future electricity (and possibly also flexibility) in any nonstandardized form to (from) other market participants or customers. Typically, the physically dispatch of the assets is mainly driven by the spot markets (for combined heat and power plants it may also be driven by a heat load that must be satisfied), whereas the financial results for the company are also strongly determined by its *hedging*, i.e., by its market activity with respect to forwards, futures, options, etc. For thermal power plants there is also a reverse side: markets for gas, coal, CO_2 emission allowances, etc.

Thus, a power generating company has to make decisions again and again, not only with respect to the scheduling of all available assets but also with respect to all possible activities on various markets. Of course, it would be desirable to have a model at hand that always calculates the optimal decisions such that the (expected) total contribution margin is at maximum. However, many of these decisions are interdependent; physical assets (e.g., storages, plants with minimum up/down times) link together electricity generation in different time periods; the existence of forward and spot markets (where the latter is typically based on an auction) makes it necessary to decide on which market the possible power generation should be sold. Thus, any optimization model in this context should not only take into account the present but also anticipate the future. However, future market prices as well as other important determining factors (such as plant availabilities, natural hydrologic inflows, wind, solar, run-of-river generation) are *uncertain*, i.e., not exactly known when decisions have to be made. Therefore, the evolution of different possibilities has to be anticipated, e.g., via branching *scenarios*; moreover, it is desirable not only to maximize the expected total contribution margin but also to minimize or bound its *downside risk* in order to prevent a too high capital requirement. Thus, technical and financial (stochastic) modeling come together here and make things highly complex. On the other hand, this decision making process is often very time

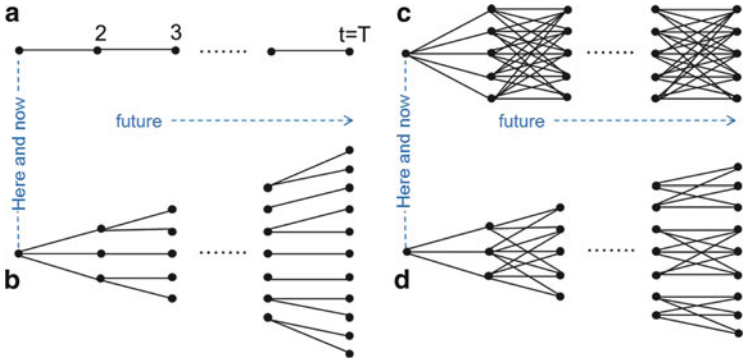

Fig. 7.1 Modeling of asset scheduling and market activities leads to a canonical time structure. If uncertainty shall be anticipated, branching of scenarios has to be introduced; each node carries input data, the arcs represent transition probabilities. (**a**) Discrete time line (typically hours) used in all kinds of (deterministic) power generation planning models; (**b**) scenario tree (non-recombining) for classical Stochastic Programming; cf. Sect. 7.3.1; (**c**) recombining scenario tree representing a Markov input data process for stochastic (dual) dynamic programming; cf. Sects. 7.3.3 and 7.3.4; (**d**) recombining scenario tree with dispersing branches

critical. Therefore, there is no all-in-one model that can cope with all aspects of this decision making process; some simplifications as well as intersections into several models are necessary.

7.2 Modeling the Planning and Scheduling Process

7.2.1 Basic Modeling Aspects

Electricity is a commodity which is not (directly) storable; electricity for a certain time period has nothing to do with electricity in another time period (unless the two time periods overlap). Therefore, market prices exist for consuming or delivering electricity in prescribed time periods in the future, e.g., (day-ahead or intraday) spot market prices for all hours within a day (or even for shorter periods, e.g., quarters of hours). Thus, for optimization models in this context, there is a canonical *time structure* according to the market with the shortest delivery periods (typically hours); cf. Fig. 7.1a.

Time periods will be denoted by $t = 1, \ldots, T$ in the sequel; $t = 1$ represents *here and now*, i.e., the present; important parameters such as hydrologic inflows and (spot market) prices for electricity in future time periods $t > 1$ are mostly *uncertain*, i.e., here and now they can only be estimated statistically. However, at the beginning of each time period t, decisions can be made that do not only affect the electricity delivery in this single time period but also in future time periods; as mentioned above, both physics and markets link together decisions for different time periods.

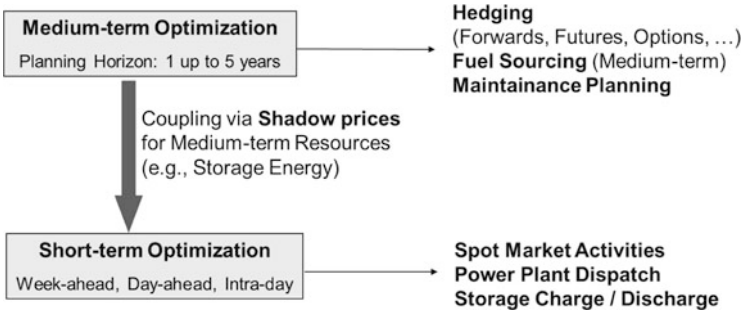

Fig. 7.2 For practical reasons, the optimization in this context is split into (at least) two models

As time proceeds (what was $t = 2, 3, \ldots$ becomes $t = 1$), the estimations of uncertain parameters in the future may change according to new available information. Thus, any optimization model in this context should be recalculated on the basis of the latest available information as often as possible (*rolling horizon*).

However, optimization models that take into account the whole medium-term time horizon of one up to, say, five years often require a computation time that is too long for time-critical short-term decisions. Therefore, in practice, the optimization process is often divided into (at least) two models, one for the short term and one for the medium term; see Fig. 7.2. The former focuses on physical dispatch and short-term market activities and should be very accurate (e.g., with an underlying quarter-hourly time grid), but with fast computation time due to a short time horizon (typically 1 up to 7 days); the latter focuses on medium-term resources such as storage energy and hedging positions (i.e., forwards, futures, options, etc.); longer computation times (e.g., overnight) as well as more simplifications are typically necessary and acceptable here; however, the future short-term operation has to be anticipated in a consistent way in order to balance the usage of medium-term resources between here and now and the future in an optimal way. Vice versa, in order to achieve a consistent overall optimization process, the short-term model has to follow the guidelines from the medium-term model, e.g., via *shadow prices* for medium-term resources.

7.2.2 Technical Modeling

Within most optimization models in practice, it is assumed that there is one hourly (deterministic) price curve for power delivery in the future (as well as one price curve for gas, coal, CO_2); cf. Fig. 7.1a. Even under this simplification, models can become very complicated due to technical aspects of power generation that require nonlinearity, such that the limits of computational power are reached (e.g., thermal plants with minimum power, start-up costs and minimum up and downtimes, or

hydro storage plants where the generation coefficients (MWh power per cubic meter of water) depend nonlinearly on the levels of the upper and the lower reservoirs).

In this chapter, we focus on markets and their decisions structures, thus we do not go further into these technical details. We only mention that most nonlinearities can be approximated reasonably by piecewise linear modeling; if these nonlinearities are even non-convexities, integer decision variables become necessary. Integer variables cause high computation times (theoretically, the computation time of an optimization problem grows exponentially with the number of integer variable to be considered), but in practice this has turned out to be the most efficient way.

In most cases it is also the technical side that forces to have several power plants in one optimization model, e.g., hydro storage plants may be coupled by common underlying run-of-river plants, thermal power plants may be coupled by a heat demand curve that has to be covered due to existing contracts. Clearly, if there is no reason for optimizing two power plants jointly, they should be optimized separately due to gains with respect to computation time.

7.2.3 Various Electricity Markets

As mentioned above, electricity (as well as gas) can be traded on various markets (such as forward and futures markets, day-ahead and intraday spot markets) that differ mainly in the amount of time between *trade date* and *delivery date*; cf. Figs. 7.3 and 7.4. The electricity market as a whole is *incomplete* in the following sense: electricity for the elementary time periods (hours) are not effectively traded before the day preceding them; trading before that day is basically restricted to block products, i.e., forwards for *standardized* bundles of hours (of course, it is possible to sell or buy hourly nonstandard products from other market participants but extra premiums will have to be paid for that). Only on a day-ahead basis it is then possible to adjust the *portfolio* in order to end up with the intended hourly profile. On the intraday market you can then go down even to quarters of hours.

7.2.3.1 Forwards and Futures Versus Spot

In most countries, forwards (with physical delivery) and futures (with financial settlement[1] on the basis of the day-ahead spot market) for standardized delivery periods can be traded continuously in both directions some months/years before the respective delivery period [6], cf. Fig. 7.3; however, there might be bid-ask-spreads and price elasticity since these markets are sometimes not very liquid. A rather liquid market is the day-ahead spot market because many market participants want to

[1] This financial settlement is such that, if the holder of a future buys the energy for all the delivery hours on the day-ahead market, the resulting net costs in the end are the same as for the holder of a forward for the same delivery period; however, the intermediate payments before the end of the delivery period can be very different.

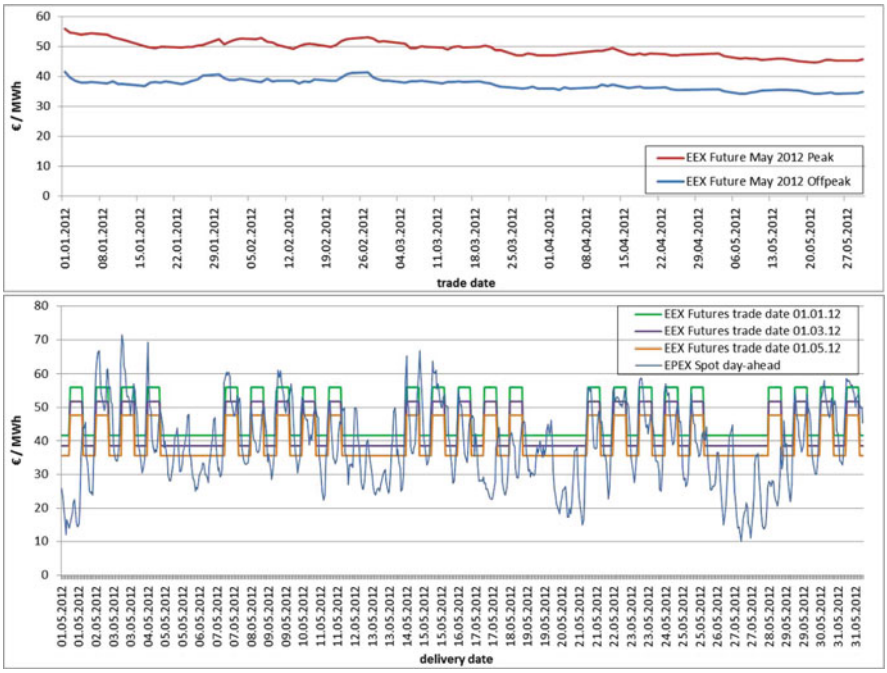

Fig. 7.3 Prices for EEX futures versus EPEX Spot for delivery in May 2012 at the market area Germany/Austria [6, 9];
Top: futures prices (peak and off-peak) for delivery month May 2012 at different trade dates;
Bottom: futures prices (from different trade dates) and spot prices for each delivery hour

readjust their positions to hourly varying schedules. In most European countries, this market is set up as an auction with a unique *clearing price* for each hour of the following day [9, 23]. Price dependent bidding allows to include price limits for each hour or for blocks of hours. The subsequent intraday spot market is typically based on continuous trading; it is less liquid and rather volatile (due to short-term variations of demand and noncontrollable generation such as run-of-river hydro power and new renewables) and, thus, offers additional opportunities for controllable generation, but it would be rather risky to rely on this market only; cf. Fig. 7.4.

We assume that the day-ahead spot market is liquid enough so that the following assumption is justified: the short-term physical dispatch of the assets of a power producer can be optimized on the basis of the spot market prices only, i.e., *independently of existing forward positions* and other (retail) delivery contracts (however, existing reserve market positions have to be taken into account; see Sect. 7.2.3.3). Of course, in the end the resulting asset schedules should be matched with all other existing positions/schedules and only the differences have to be traded on the spot market (*netting*).

Whereas the physical dispatch is primarily determined by spot market prices, the financial result of the assets is strongly determined by the *hedging*, i.e., by the

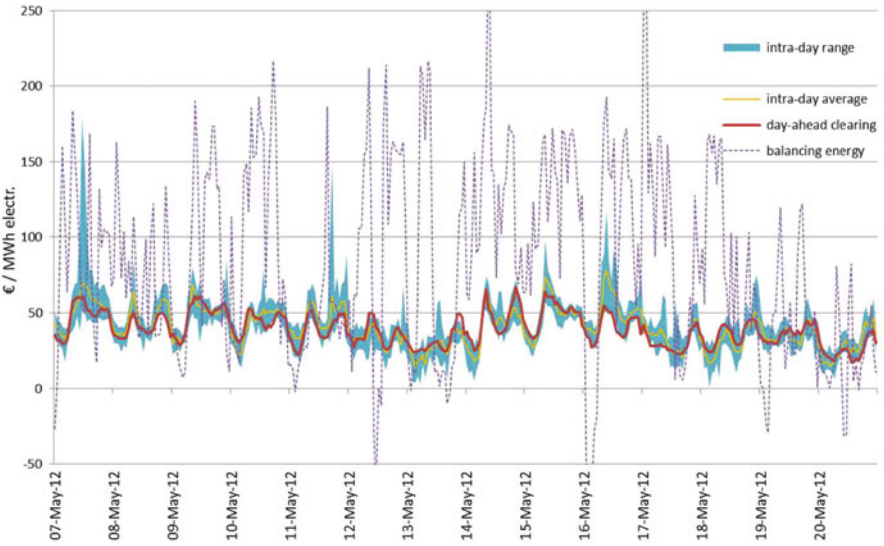

Fig. 7.4 Different market prices for electricity delivery at Austria in various hours of May 2012

question if and when power that can be generated should be sold on the forward market. Hedging of physical production is popular, because the financial results of the assets are much better foreseeable; moreover, in the last years, on average forward prices have been significantly higher than equivalent spot prices. This can be a sign that buyers are more risk-averse than power producers.

7.2.3.2 Optionality and Options

A power producer with flexible assets can make use of this flexibility in various ways. He/she can sell power for some delivery period on the forward market when the price is high (above fuel and other variable costs, fuel costs should be hedged at the same time); he/she then has further opportunities without any market risk: he/she has the possibility to generate this power in this period or to *buy it back* before the delivery period, e.g., if prices have fallen; in the latter case he/she has generated income without having used his/her assets physically. This sort of flexibility can also be hedged by selling call *options* on forwards for which an option premium can be charged (without taking a market risk).

Moreover, flexible assets also provide the possibility to sell power at the day-ahead spot market only at high price hours (and to buy back at low price hours if forwards have been sold before). Last but not least, the resulting schedule after the day-ahead market can additionally be adapted by trading in the intraday market which, despite a lower market depth compared to the day-ahead market, may provide further considerable opportunities due to high volatility; cf. Fig. 7.4. This kind

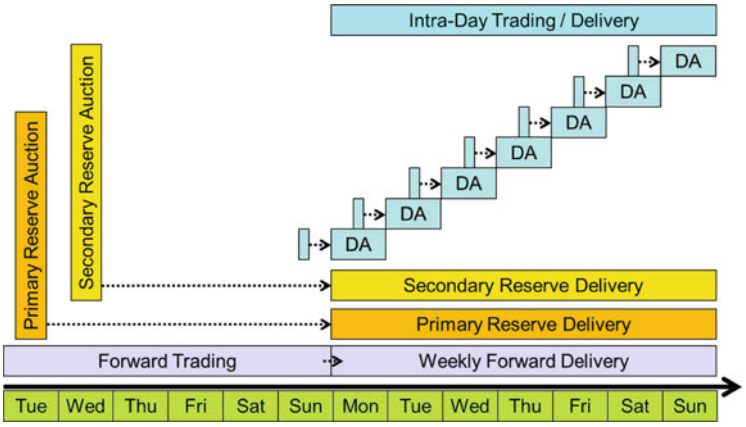

Fig. 7.5 Weekly decision structure for a power producer induced by the German electricity markets: primary control reserve and secondary control reserve are auctioned week-ahead, tertiary control reserve and spot are auctioned day-ahead (*DA*); and there is a continuous forward and intraday trading

of optionality can only be hedged (i.e., sold in advance) on the basis of bilateral, nonstandardized contracts, e.g., fixed hourly schedule deliveries, full supply contracts, swing options, etc.; of course, for such contracts a substantial premium will be charged.

Thus, a flexible asset such as a power plant can be understood as a *bundle of call options* on (quarter-) hourly power prices where the *strike price* is given by the variable generation costs (including fuel and CO_2 emission costs in the case of thermal power plants[2]).

7.2.3.3 Balancing Energy and Reserve Market

If the net physical delivery of a market participant does not exactly match his/her final quarter-hourly net schedule (which is equal to the sum of all positions on all markets for a certain area), the differences are billed by the system operator (ISO/TSO) according to a *balancing energy* regime. The prices for balancing energy depend on regulatory settings and on the mismatch of energy in the whole (or in zones of the) market area in different time periods[3]; they are typically highly volatile (cf. Fig. 7.4) and published only ex post, so it is wise to meet your net schedule as accurately as possible in order to avoid unforeseeable balancing energy costs. Thus, for a power producer with controllable assets, there is an incentive to use the flexibility of these assets to better match his/her final net schedule.

[2] If there is a liquid gas spot market, a gas fired plant (e.g., a CCGT) can also be understood as a bundle of call options on the *clean spark spread*, i.e., on the difference between hourly power price minus appropriately scaled gas and CO_2 prices.

[3] In this chapter we do not consider *nodal pricing* systems which are applied, e.g., in the USA [28].

An additional market, which differs fundamentally from the previous ones, is the *reserve market* (ancillary service market); this is where the system operator gets the energy and the short-term flexibility for compensating the above-mentioned mismatches in the system or in some parts of it. Here, power producers running assets with flexibility (or large consumers with flexible demand) can sell some of their generation capacity for certain time periods and receive a *capacity price* for that; in these time periods, they have to be ready to increase (positive reserve) or decrease (negative reserve) their production (or consume additional energy, e.g., by pumping into hydro reservoirs) on (very) short notice from the system operator. If some power producer has sold some capacity on the reserve market, he/she may be *called* to actually change his/her production; in this case, there is an *energy price* as a compensation for the delivered energy (additionally to the capacity price). In Germany and Austria these markets are set up as an auction (with pay-as-bid prices) that takes place before the day-ahead spot market auction [1, 29] (cf. Fig. 7.5), whereas in Italy [14] and France [31] reserve is scheduled thereafter. Note that in this market a power producer can only sell reserve capacity to the system operator; it is not possible to buy it back once it is sold. Since the reserve market prices usually include a premium for the flexibility, the overall revenues (capacity price and expected energy price) are typically above the (expected) revenues for this capacity put on the spot market instead.

7.3 Stochastic Optimization

In the previous section we have illustrated that the electricity market as a whole has a complicated, sequential decision structure with stochastic prices. It has become clear that, in this environment, the flexibility of assets has a value by itself; flexible assets can be understood as bundles of options. However, physical assets have technical, time-spanning restrictions such as minimum uptimes and storage balances; therefore, methods from Mathematical Finance cannot directly be applied here. Stochastic optimization can take into account both stochastic decision structures and technical aspects.

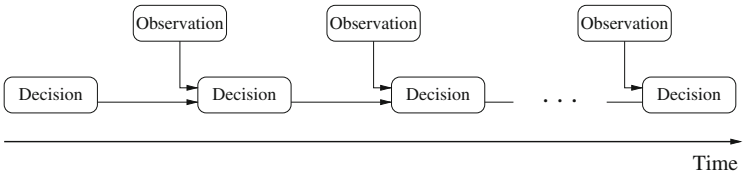

Fig. 7.6 Multistage Stochastic Programming is based on a finite alternating sequence of decisions and observations of previously uncertain data (e.g., energy market prices, hydrologic inflows); thus, this framework covers the situation of a power producer in a market environment (cf. Fig. 7.5)

7.3.1 Multistage Stochastic Programming

Multistage Stochastic Programming is a framework for stochastic optimization that takes into account the real decision structure of planning and scheduling in power generation (which is the same as in many other application areas). This framework is based on a finite alternating sequence of *decisions* and *observations* of previously uncertain data [2, 33]; see Fig. 7.6. The uncertain data is represented by *random variables*. Each random variable is assigned to a time step $t \in \{2,\dots,T\}$ at which its actual outcome can be observed. Before that time step, stochastic information is available. It is assumed that the outcomes of the random variables are not affected by the decisions.

For the (mixed-integer) linear case such a *multistage stochastic program* can be formally stated, e.g., as follows:

$$\min_{x_1,\dots,x_T} \left\{ \mathbb{E}\left[\sum_{t=1}^{T} c_t \cdot x_t\right] \,\middle|\, \begin{array}{l} x_t \in X_t \ (t=1,\dots,T), \\ A_{t,0}x_t + A_{t,1}x_{t-1} = h_t \ (t=2,\dots,T), \\ x_t = x_t((c_\tau, A_{\tau,0}, A_{\tau,1}, h_\tau)_{\tau=2,\dots,t}) \ (t=2,\dots,T) \end{array} \right\} \quad (7.1)$$

with

- decision vectors x_t to be chosen optimally;
- cost vectors $c_t \in \mathbb{R}^n$, for $t \geq 2$ some components can be random variables;
- the *expected value* functional $\mathbb{E}$ that maps the (random) total costs to a real number (note that $\mathbb{E}$ is a *linear* functional);
- sets $X_t \subseteq \mathbb{R}^n$ defining *constraints* for respectively one time step only; typically X_t are polyhedrons, e.g., $X_t = \{x \in \mathbb{R}^n : A_t x_t \leq b_t\}$ or $\{x \in \mathbb{Z}^k \times \mathbb{R}^{n-k} : A_t x_t \leq b_t\}$, if some decisions have to be restricted to *integer* numbers;
- dynamic constraints $A_{t,0}x_t + A_{t,1}x_{t-1} = h_t$ that link together consecutive time steps (e.g., balance constraints for energy storages, constraints for minimum up- and down-times of power plants), some components of the matrices $A_{t,0}, A_{t,1}$ and the right-hand side vectors h_t can be random for $t \geq 2$;
- *non-anticipativity* constraints $x_t = x_t((c_\tau, A_{\tau,0}, A_{\tau,1}, h_\tau)_{\tau=1,\dots,t})$ which ensure that each decision vector x_t can only depend on the actual outcomes of the random variables observed until time t (however, x_t may depend on the conditional distribution of $(c_\tau, A_{\tau,0}, A_{\tau,1}, h_\tau)_{\tau=t+1,\dots,T}$ given the outcomes until time t); note that, formally, the decision vectors $x_2,\dots,x_T$ are also random vectors, whereas x_1 represents the *here and now* decisions that have to be chosen uniquely for all possible outcomes.

The sequence $(c_t, A_{t,0}, A_{t,1}, h_t)_{t=1,\dots,T}$ can be understood as a multivariate, discrete-time *stochastic process*. In order to be able to solve (7.1), we assume in the following that this process is discrete (not only in time) with *only finitely many possible outcomes* for each random variable[4]. Then, clearly, also the stochastic process of decision vectors $x_1,\dots,x_T$ is discrete and finite. Since random components

[4] With regard to (7.1), process with infinitely many possible outcomes can in many cases be suitably approximated by finite ones, e.g., by Monte Carlo sampling and clustering on the basis of

at different time stages may be autocorrelated and since each x_t is allowed to depend not only on the information revealed at time t but on the complete respective history $1,\ldots,t$, the resulting decision structure is a tree which is called the *scenario tree*; see Fig. 7.1b. Thus, by introducing copies of the decision vectors x_t for each scenario tree node belonging to the time step t, (7.1) can be computed numerically by standard optimization solvers such as Xpress or CPLEX (or by special decomposition approaches for special instances of (7.1) [32]).

However, the size of the scenario tree and thus the dimension of (7.1) grows rapidly with the number of branches; if we assume a systematic branching at certain points in time where all scenarios split, then, clearly, the tree size depends exponentially on the number of such branching points (*curse of dimensionality*). Standard solvers (as well as decomposition approaches) will reach their limits. Thus, such non-recombining scenario trees can only be used for short-term models or for medium-term models that focus only on medium-term uncertainty (such as general price trend or inflows for large hydro reservoirs) but not on short-term (day-to-day) uncertainty.

7.3.1.1 Statistical Modeling of Uncertainty

To set up a complete Stochastic Programming model for asset scheduling and planning, a statistical discrete-time model for all the relevant uncertain input parameters is required, e.g., for

- all kind of power market prices (cf. Sect. 7.2.3) including price elasticity[5];
- fuel market prices and market prices for CO_2 emission allowances;
- heat demand (if combined heat and power generation is considered);
- electricity demand (if that has to be considered for some reason);
- generation of noncontrollable generators (run-of-river hydro power, new renewables);
- natural inflows into hydro storages on the basis of rainfall, snowmelt, evaporation;
- reserve calls (if ancillary services are considered);
- unexpected outages of any sort of asset.

Because all these uncertainty factors (except unexpected outages) are correlated in general, a joint model is desirable. The traditional approaches are classical time series models such as AR(I)MA, (G)ARCH, as well as discretized SDEs with jumps, mean reversion, regime switching, etc., inspired from Mathematical Finance

stability theory (cf., e.g., [16, 17, 25]) or via other (Quasi) Monte Carlo methods. Thereby, the discrete approximation of the stochastic process is separated from the solution process; however, there is also work on integrated sampling and solution algorithms. Note that, without any sampling, (7.1) would have to be solved analytically and that is possible only in very special cases.

[5] Note that price elasticity, if it is approximated in a piecewise linear way, can easily be incorporated into (7.1) without inducing additional integer variables.

[4, 20, 37]. To get correlated price curves, often only a subset of price curves is modeled as a time series model; the other ones are deduced therefrom by regression models.

Recently, also so-called *fundamental models* have become very popular, cf., e.g., [3] and [5, Chap. 4]. These models carry out a (simplified) optimization of *all* physical assets within a market area on the basis of an electricity demand curve and fuel prices; electricity prices are given by the marginal costs of the demand satisfaction constraints. This approach has the advantage that there is typically a reasonable correlation between the uncertainty factors and that there is also information about market depth and price elasticity.

In this chapter we cannot go into further details about statistical modeling; we only mention at this point that for a stochastic program such as (7.1) a tree structure with transition probabilities is required; cf. Fig 7.1b. This might be achieved by conditional sampling. If the output of a statistical model consists of individual (multivariate) sample paths only, then it is possible to generate reasonable tree structures out of them by applying special clustering methods on the basis of stability theory (cf., e.g., [16, 17, 25]).

Recall that we assume that there is no monopolist in the electricity market. We also assume that big market participants do not use their market power in a way that would make it necessary to consider game theoretic aspects. If such aspects have to be taken into account, then not only the price model but also the optimization model can become much more complex than (7.1), cf., e.g., [19].

7.3.2 Risk Functionals

Optimization in general tends to go to the limits (given by the constraints) and may thereby produce rather radical solutions which might cause trouble in practice. Using Stochastic Programming counteracts this tendency to some degree, because each decision vector x_t is chosen such that it is good not just for one set of data but for everything that can happen at time $t+1,\dots,T$ within the scenario tree. However, if the expected total costs $\mathbb{E}[\sum_{t=1}^{T} c_t \cdot x_t]$ are minimized, there is no incentive for the dispersion of these costs; it is then not unlikely that, after the random data has been revealed, one might end up with a cash flow that is significantly below what was expected.

Stochastic Programming allows to reduce this *downside risk* by including a *risk functional* ρ (risk measure) or a deviation functional in the objective[6] of (7.1), as it is known from Mathematical Finance [12, 18, 22]:

$$\min_{x_1,\dots,x_T} \left\{ \begin{array}{l} \gamma \cdot \rho(c_1 \cdot x_1, \dots, c_T \cdot x_T) \\ +(1-\gamma) \cdot \mathbb{E}[\sum_{t=1}^{T} c_t \cdot x_t] \end{array} \middle| \begin{array}{l} x_t \in X_t,\ A_{t,0} x_t + A_{t,1} x_{t-1} = h_t, \\ x_t = x_t((c_\tau, A_{\tau,0}, A_{\tau,1}, h_\tau)_{\tau=2,\dots,t}) \end{array} \right\} \tag{7.2}$$

with some fixed number $\gamma \in [0,1]$.

[6] Alternatively, it is possible to include a risk constraint of the form $\rho(c_1 \cdot x_1, \dots, c_T \cdot x_T) \le \beta$ into (7.1) with some fixed real number β.

If a risk functional ρ is incorporated into (7.1), it should be chosen carefully such that

- it represents reasonably the attitude towards financial risk of the power producer or is consistent with it;
- it does not destroy the efficient numerical resolvability of (7.1).

To satisfy the latter point, a minimum requirement for ρ is *convexity*; a polyhedral risk functional [7] such as Conditional-Value-at-Risk [30] can be even more advantageous in this regard since it maintains linearity structures of (7.1) (note that risk functionals are naturally nonlinear). Going into details with respect to the former point is beyond the scope of this chapter; we refer to the monographs [12, 26] for that. We only mention that there are not only one-period risk functionals $\rho(\sum_{t=1}^{T} c_t \cdot x_t)$ (such as the well-known Value-at-Risk, which is not convex, or the variance that has been used as a risk functional some decades ago [22]) but also multi-period risk functionals $\rho(c_1 \cdot x_1, \ldots, c_T \cdot x_T)$ that take into account the evolution of the costs over time (note that very high accumulated cost at some time $t < T$ can mean that the company will not survive until time T). Such approaches have been demonstrated, e.g., in [8].

7.3.3 Stochastic Dynamic Programming

Since the optimization problems (7.1) and (7.2) have a fundamental time structure, it appears attractive to tackle it in a recursive way, in particular for medium- and long-term time horizons. Formally, under some mild regularity assumptions, the (mixed-integer) linear[7] problem (7.1) can equivalently be stated recursively as

$$\begin{aligned}
&\min_{x_1}\{c_1 \cdot x_1 + C_2(x_1, d_1) \,|\, x_1 \in X_1\}\,, \\
&C_t(x_{t-1}, d^{t-1}) = \mathbb{E}_{d_t|d^{t-1}}\left[\min_{x_t}\left\{c_t \cdot x_t + C_{t+1}(x_t, d^t) \,\middle|\, \begin{array}{l} x_t \in X_t, \\ A_{t,0}x_t + A_{t,1}x_{t-1} = h_t \end{array}\right\}\right] \\
&(t = 2, \ldots, T)
\end{aligned} \tag{7.3}$$

and $C_{T+1} \equiv 0$. Here, the short notations $d_t = (c_t, A_{t,0}, A_{t,1}, h_t)$ for the random *data* at time t and $d^t = (d_1, \ldots, d_t) = (d^{t-1}, d_t)$ for the whole respective history have been used. $\mathbb{E}_{d_t|d^{t-1}}$ denotes the conditional expectation which is just a weighted sum if everything is discrete. The functions $C_t(x_{t-1}, d^{t-1})$ are called the *future cost functions*. Note that the minimizations in (7.3) are separate with respect to the time steps and with respect to the scenario realizations d_t, i.e., x_t are no longer random vectors here. Thus, problem (7.1) has been separated into many smaller problems and that appears to be favorable for the solution procedure.

[7] For the risk-averse problem (7.2) such a decomposition is also possible but only for special risk functional such as recursive [34] or polyhedral ones [7].

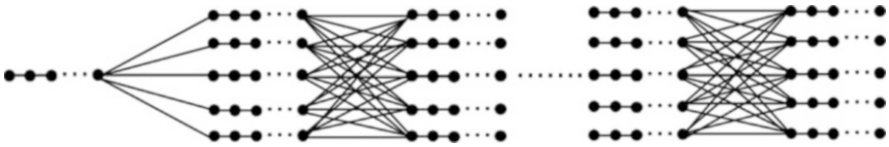

Fig. 7.7 In SDP/SDDP it is not always necessary to consider branching at every time step t (hour); e.g., allowing branching only every day or once every week may be a compromise with regard to computation time. Subproblems in (7.3) will cover 24 or 168 time steps, respectively, instead of just one

But in general, unless there are no time-spanning or dynamic constraints, the decisions x_t are all interdependent and a solution by a simple forward recursion is not possible since at time t the future cost function $C_{t+1}(.,.)$ is required but unknown. The *dynamic programming* principle is therefore a backward recursion in which, starting at $t = T$, the function values $C_t(x_{t-1}, d^{t-1})$ are calculated according to (7.3) for all scenarios d^{t-1} in the scenario tree at time $t-1$ and, if possible, *for all* possible solutions $x_{t-1} \in X_{t-1}$ (since the optimal $x_{t-1}(d^{t-1})$ is yet unknown). If some or all components of x_{t-1} have to be chosen from a continuous region (which is typically the case), it is impossible to calculate anything *for all* x_{t-1} since there are infinitely many possibilities. In this case, the admissible set X_{t-1} for x_{t-1} has to be *discretized* finely.

Therefore, the dynamic programming principle can only be applied if the sets X_{t-1} are bounded for $t = 1, \ldots, T$ and if the dimension $n = \dim(x_{t-1})$ is not too high or, as it is the case in many practical applications, if at least $C_t(x_{t-1}, d^{t-1})$ does in fact not depend on all components of x_{t-1} but only on some (*state variables*) as many columns of $A_{t,1}$ are zero. For example, for a hydro storage system, $C_t(x_{t-1}, d^{t-1})$ only depends on the storage volumes but is independent of the other components of x_{t-1} such as turbined outflow and spill.

After all $C_t(.,.)$ have been (approximately) calculated in a backward recursion, a *forward pass* through the scenario tree starting at $t = 1$ can be carried out in which the solutions $x_t(d^t)$ for all respective scenarios d^t can be calculated; $C_{t+1}(.,d^t)$ has to be interpolated between the discretization points for x_t.

The advantage of solving smaller problems in (7.3) than in (7.1) may however be (over)compensated by the fact that, within the backward pass, $C_t(x_{t-1}, d^{t-1})$ has to be calculated *for all* discretization points of the state space; the number of discretization points and thus the number of (small) problems in (7.3) to be solved (for each t and each scenario) grows exponentially with the number of state components of x_t (*curse of dimensionality*). Thus, only application problems with few state variables (e.g., hydro power systems with not more than, say, four reservoirs) can be solved in this way.

The real advantage of *stochastic dynamic programming* (SDP) (if it can be practically applied due to a moderate number of state variables) comes out if the stochastic input data is given by a discrete *Markov process* (i.e., each d_t depends directly only on d_{t-1} such that $\mathbb{E}_{d_t|d^{t-1}} = \mathbb{E}_{d_t|d_{t-1}}$, respectively); such a process can be represented by a *recombining scenario tree*, cf. Fig. 7.1c. In this case, the future cost functions

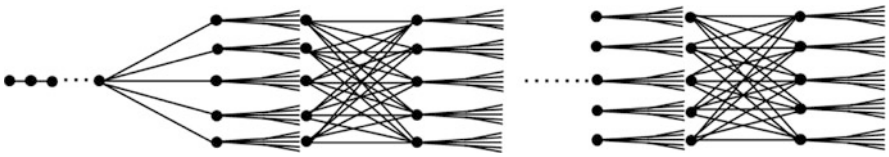

Fig. 7.8 In SDP/SDDP the subproblems in (7.3) for the intraday or intra-week can also be stochastic

C_t do not depend on the complete data history $d_1, \ldots, d_{t-1}$; it suffices to calculate them for each node of the recombining tree only (but of course for all possible states x_{t-1}). Thus, the computation time is linear in the number of time steps, i.e., much lower than for a general scenario tree (cf. Fig. 7.1b), and much more branching can be incorporated; the representation of the real-world uncertainty is much finer.

On the other hand, being restricted to Markov processes for the input data might appear discouraging for many practitioners since ordinary Markov processes have a short memory only and power prices and hydrologic inflow processes are therefore usually modeled in a non-Markovian way. However, there are possibilities to loosen this restriction to some extent. Firstly, since the process may be multivariate, *regime switching* can be included (e.g., for power prices switching between a normal regime and spike regime). Furthermore, it is possible to let the number of nodes of the recombining tree increase with t and establish independent branches by setting some transition probabilities to 0; these independent branches may disperse in completely different directions with respect to some data (e.g., power and gas prices). Last but not least we note that the branching of the scenario tree (*state transition*) does not necessarily need to occur on *every* time step t (which typically refers to a single hour of the planning horizon) but on some selected time steps only, e.g., once a day or once a week, cf. Fig. 7.7. Thereby, the Markov process can integrate longer time periods; this may also be advantageous with respect to the overall computation time as the number of (sub)problems in (7.3) is reduced substantially. This approach also provides the possibility to make even the intraday or intra-week problems stochastic, e.g., to cover the decision structure of day-ahead auction and intraday trading [21].

7.3.4 SDDP

Stochastic dual dynamic programming (SDDP) is, like SDP, based on the recursive formulation (7.3) with $d_1, \ldots, d_T$ being a Markov process [24]; it is assumed that there are no integer requirements for the decisions x_t. In this case, it is known from optimization theory, that the future cost function $C_t(x_{t-1}, d_{t-1})$ is *convex* with respect to[8] x_{t-1}. Therefore, the idea of SDDP is, instead of calculating the values

[8] For certain data processes $d_1, \ldots, d_T$, it can be shown that $C_t(x_{t-1}, d_{t-1})$ is also convex with respect to the right-hand sides h_{t-1} (e.g., hydrologic inflows) and this can be used to save further

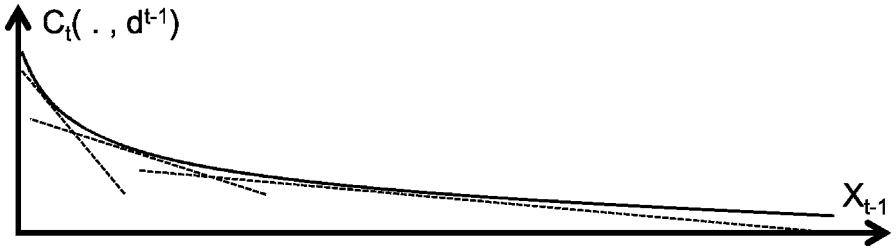

Fig. 7.9 In SDDP, for $t = 2,\ldots,T$ and for all possible d_{t-1}, the future cost functions $C_t(.,d_{t-1})$ are approximated from below by cutting planes (*dashed lines*) in the space of state variables x_{t-1} (e.g., storage levels)

of each $C_t(.,d_{t-1})$ for all x_{t-1}, to approximate these functions successively via collections of hyperplanes (*cutting planes*, or cuts) in $\mathbb{R}^n$ which are all located *below* but at some point close to $C_t(.,d_{t-1})$; cf. Fig. 7.9.

The solution procedure of SDDP is a loop over

1. A *forward pass*:
   (a) Sample a number of paths $d_1,\ldots,d_T$ from the Markov model/scenario tree;
   (b) For each sample path calculate $x_t(d_t)$ for $t = 1,2,\ldots$ as the solution of the problem $\min_{x_t}\{c_t \cdot x_t + C_{t+1}(x_t,d_t) | x_t \in X_t, A_{t,0}x_t + A_{t,1}x_{t-1} = h_t\}$ on the basis of the *currently available cutting plane approximation* of $C_{t+1}(.,d_t)$;
   (c) Store the pairs $d_t, x_t(d_t)$;
   (d) For each sample path sum up the values $c_t \cdot x_t(d_t)$ and calculate the average over all sample path $\rightsquigarrow$ *upper bound* (since the decisions $x_t(d_t)$ are not perfectly optimal); and
2. A *backward pass*:
   (a) Go through the sample paths from the forward pass starting at $t = T$;
   (b) For all the pairs (d_{t-1},x_{t-1}) stored within the forward pass calculate $C_t(d_{t-1},x_{t-1})$ according to (7.3) and obtain the *dual multipliers* (shadow prices) for the dynamic constraints $A_{t,0}x_t + A_{t,1}x_{t-1} = h_t$; together with the sampling point x_{t-1}, the associated dual multipliers define a cutting plane, i.e., a hyperplane below but close to $C_t(.,d_{t-1})$;
   (c) Add the new cutting planes to the respective collection for $C_t(.,d_{t-1})$;
   (d) Sum up the optimal values for each sample path and store the average $\rightsquigarrow$ *lower bound* (since the cutting plane approximation is always below $C_t(.,d_{t-1})$).

The procedure stops when lower and upper bounds are close enough. It can be shown that it converges in a finite number of time steps [27, 35]. In practice, satisfactory convergence speed can be achieved for much higher dimensional problems compared to SDP, because in SDDP $C_t(.,d_{t-1})$ does not need to be calculated *for all*

computation time by calculating cutting planes jointly for both arguments x_{t-1} and h_{t-1}; cf. [24]. We here restrict the presentation to the more general case.

x_{t-1} as in SDP. Furthermore, the algorithm is such that cutting planes are added only at points x_{t-1} which are on the way to the optimal solution; thus, the approximation of $C_t(.,d_{t-1})$ is only accurate where it is needed.

Since in each iteration of SDDP new cutting planes are added to the collection, significant additional speedup can be achieved by intelligently eliminating old cutting planes which are redundant or no longer interesting [21]. For efficiently determining whether a cut is redundant, heuristic methods are helpful.

7.4 Conclusion

Methods from Mathematical Finance [12, 18] often cannot be applied for power generation planning due to complex technical restrictions of power generating assets. Conventional deterministic optimization models, widely used in practice, can handle these technical aspects but cannot take into account stochastic decision structures induced by the electricity markets. In theory, stochastic optimization can take into account both stochastic decision structures and technical restrictions of assets. In practice, there are computational limits, simplifications may become necessary at some point. Stochastic optimization provides the flexibility with regard to which details should be included and which ones can be simplified. A power producer has to determine precisely to what purpose he/she wants to run this model; typically it makes sense to have different models for different purposes. In particular, there is the choice between classical multistage Stochastic Programming based on a non-recombining scenario tree (cf. Fig. 7.1b) and SDP/SDDP based on a recombining scenario tree (cf. Fig. 7.1c, d as well as Figs. 7.7 and 7.8).

For the short-term, a classical multistage stochastic program seems to be the natural choice since it can very well model the decision structure of day-ahead and intraday spot as well as the reserve market (cf. Fig. 7.5); moreover, time-spanning constraints such as storage balance equations as well as minimum up and downtimes can be modeled without limitation; see, e.g., [10, 13]. In case there are medium-term resources such as gas or seasonal hydro storages there has to be a coupling to a medium-term model as illustrated in Fig. 7.2.

For the medium-term, the answer is much more difficult; an SDP/SDDP model seems to be more attractive since obviously the representation of uncertainty can be much richer and short-term decision structures can be anticipated. However, this approach is limited with respect to the following issues:

- The stochastic input process must be Markov (cf. end of Sect. 7.3.3);
- The number of time-spanning, or dynamic, constraints is limited: even in SDDP it becomes difficult to handle more than, say, 20 state variables in practice (however, for some dynamic constraints such as minimum up and downtimes it may be acceptable to consider them only within the states and neglect them at the state transition time steps);
- The faster method SDDP cannot handle integer decision variables;
- The incorporation of a risk functional might be difficult in practice (though at least in theory it is possible [15, 36]).

On the other hand, a classical non-recombining scenario tree (cf. Fig 7.1b) can only cope with few branching points only. On the way from the root node $t = 1$ to the leave, a scenario of tree with, say, 1,000 scenarios, encounters only approx. 10 branching points on average; this seems to be not much for a medium-term time horizon of several years. Integer decision variables or risk functionals will make this issue even more serious. However, it can make sense to work with such models if only medium-term price trends and medium-term resources (such as seasonal hydro storages) are under consideration as, e.g., in [11].

References

1. APG: Austrian Power Grid (2013) Web page: www.apg.at
2. Birge JR, Louveaux F (1997) Introduction to stochastic programming. In: Springer series in operations research. Springer, New York
3. Boogert A, Dupont D (2008) When supply meets demand: the case of hourly spot electricity prices. IEEE Trans Power Syst 23:389–398
4. Burger M, Klar B, Müller A, Schindlmayr G (2004) A spot market model for pricing derivatives in electricity markets. Quant Financ 4:109–122
5. Burger M, Graeber B, Schindlmayr G (2007) Managing energy risk (Wiley Finance Series). Wiley, Chichester
6. EEX: European Energy Exchange (2013) Web page: www.eex.com
7. Eichhorn A, Römisch W (2005) Polyhedral risk measures in stochastic programming. SIAM J Optimiz 16:69–95
8. Eichhorn A, Heitsch H, Römisch W (2009) Scenario tree approximation and risk aversion strategies for stochastic optimization of electricity production and trading. In: Kallrath J et al (eds) Optimization in the energy industry, chap 14. Springer, Berlin, pp 321–346
9. EPEX Spot: European Power Exchange (2013) Web page: www.epexspot.com
10. Faria E, Fleten SE (2011) Day-ahead market bidding for a Nordic hydropower producer: taking the Elbas market into account. Comput Manage Sci 8:75–101
11. Fleten SE, Wallace SW (2009) Delta-hedging a hydropower plant using stochastic programming. In: Kallrath J et al (eds) Optimization in the energy industry, chap 22. Springer, Berlin, pp 507–524
12. Föllmer H, Schied A (2004) Stochastic finance: an introduction in discrete time. In: De Gruyter studies in mathematics, vol 27, 2nd edn. Walter de Gruyter, Berlin
13. Garcés L, Conejo A (2010) Weekly self-scheduling, forward contracting, and offering strategy for a producer. IEEE Trans Power Syst 25:657–666
14. GME: Gestore Mercati Energetici (2013) Web page: www.mercatoelettrico.org
15. Guigues V, Römisch W (2012) SDDP for multistage stochastic linear programs based on spectral risk measures. Oper Res Lett 40:313–318

16. Heitsch H, Römisch W (2009) Scenario tree modeling for multistage stochastic programs. Math Program 118:371–406
17. Heitsch H, Römisch W, Strugarek C (2006) Stability of multistage stochastic programs. SIAM J Optimiz 17:511–525
18. Hull J (2011) Options, futures and other derivatives, 8th edn. Pearson, Harlow
19. Kovacevic RM, Pflug GC (2013) Electricity swing option pricing by stochastic bilevel optimization: a survey and new approaches. www.speps.org (preprint)
20. Kovacevic RM, Wozabal D (2013) A semiparametric model for EEX spot prices. IIE Transactions. doi:10.1080/0740817X.2013.803640
21. Löhndorf N, Wozabal D, Minner S (2013) Optimizing trading decisions for hydro storage systems using approximate dual dynamic programming. Operations Research 61:810–823
22. Markowitz H (1952) Portfolio selection. J Financ 7:77–91
23. NordPool: Nord Pool Spot (2013) Web page: www.nordpoolspot.com
24. Pereira MVF, Pinto LMVG (1991) Multi-stage stochastic optimization applied to energy planning. Math Program 52:359–375
25. Pflug GC, Pichler A (2011) Approximations for probability distributions and stochastic optimization problems. In: Bertocchi M et al (eds) International series in operations research and management science, vol 163, chap 15. Springer, New York, pp 343–387
26. Pflug GC, Römisch W (2007) Modeling, measuring, and managing risk. World Scientific, Singapore
27. Philpott AB, Guan Z (2008) On the convergence of stochastic dual dynamic programming and related methods. Oper Res Lett 36:450–455
28. PJM: Pjm Interconnection (2013) Web page: www.pjm.com
29. Regelleistung.net: Internetplattform zur Vergabe von Regelleistung (2013) Web page: www.regelleistung.net
30. Rockafellar RT, Uryasev S (2002) Conditional value-at-risk for general loss distributions. J Bank Financ 26:1443–1471
31. RTE: Réseau de Transport d'Electricité (2013) Web page: www.rte-france.com
32. Ruszczyński A (2003) Decomposition methods. In: Ruszczyński A, Shapiro A (eds) Handbooks in operations research and management science, vol 10, chap 3, 1st edn. Elsevier, Amsterdam, pp 141–211
33. Ruszczyński A, Shapiro A (eds) (2003) Stochastic programming. In: Handbooks in operations research and management science, vol 10, 1st edn. Elsevier, Amsterdam
34. Ruszczyński A, Shapiro A (2006) Conditional risk mappings. Math Oper Res 31:544–561
35. Shapiro A (2011) Analysis of stochastic dual dynamic programming method. Eur J Oper Res 209:63–72
36. Shapiro A, Tekaya W, Da Costa J, Soares M (2013) Risk neutral and risk averse stochastic dual dynamic programming method. Eur J Oper Res 224:375–391
37. Weron R, Bierbrauer M, Trück S (2004) Modeling electricity prices: jump diffusion and regime switching. Physica A 336:39–48

Chapter 8
Risk Measures in Multi-Horizon Scenario Trees

Adrian S. Werner, Alois Pichler, Kjetil T. Midthun, Lars Hellemo,
and Asgeir Tomasgard

Abstract Production assurance requirements are used to ensure that the operation of natural gas transportation networks is robust with respect to flow and production disruptions. They also affect strategies for optimal infrastructure investments. Motivated by a combined investment and operational optimization model for natural gas transport, we describe how to address such requirements through risk measure formulations such as Average Value-at-Risk. The large number of operational scenarios required for a meaningful analysis of the risk measures creates a computational challenge. A new scenario tree structure, multi-horizon scenario trees, can improve computational tractability. We investigate properties of the risk measures such as time consistency for such scenario trees and illustrate this discussion with a stylized example.

8.1 Introduction

Industries with large capital investments are exposed to both long-term uncertainty and short-term or operational variations and uncertainty. Long-term uncertainty includes trends in demand or price developments, costs for infrastructure elements,

A.S. Werner (✉)
Department of Applied Economics and Operations Research,
SINTEF Technology and Society, Trondheim, Norway
e-mail: AdrianTobias.Werner@sintef.no

A. Pichler
Department of Industrial Economics and Technology Management, Norwegian University of Science and Technology (NTNU), Trondheim, Norway

K.T. Midthun • L. Hellemo
SINTEF, Trondheim, Norway

A. Tomasgard
Norwegian University of Science and Technology (NTNU), Trondheim, Norway

R.M. Kovacevic et al. (eds.), *Handbook of Risk Management in Energy Production and Trading*, International Series in Operations Research & Management Science 199,
DOI 10.1007/978-1-4614-9035-7_8,

and new resource discoveries. Short-term uncertainty comprises daily variations in demand or prices and unplanned events such as production stoppages and outages.

In optimization models, this uncertainty can be handled by applying stochastic programming methodology, where the uncertain parameters are represented by discrete values in scenario trees for possible future realizations of the parameters. Computational tractability is usually of great concern in such models, because variables and constraints are duplicated for each scenario in the scenario tree. If all daily variations are to be included, the scenario tree and, hence, the optimization model will become intractable. For this reason, operational details are often aggregated for strategic models. By applying a multi-horizon scenario tree structure as described in Sect. 8.2, relevant operational detail can be included into stochastic programming models without sacrificing computational tractability.

We discuss the application of risk measures on stochastic programming models with multi-horizon scenario trees and how risk measures can be applied not only to monetary performance. Our motivation stems from investment problems for the natural gas sector, in particular the project "Regularity and uncertainty analysis and management for the Norwegian gas processing and transportation system" (Ramona, funded by the Norwegian Research Council); see Sect. 8.3. This project developed new methodology and tools for optimizing production assurance and capacity utilization in natural gas production, processing, and transportation systems. During the project period, an optimization model was developed to find infrastructure solutions for processing and transporting natural gas from fields (reservoirs) to the markets. Such an infrastructure must be robust and flexible, allowing reliable and profitable operations under various, also adverse, situations. The main driving force is profitability, and the objective is to maximize net present value.

In addition to the obvious financial benefit of being able to fulfill contract obligations through delivery to the marketplaces, the producers value high production assurance, i.e., the capability of a system with respect to production performance or to meet the demand for deliveries. Hence, production assurance requirements can be imposed at both the production side (fields) and the consumption side (markets) of the network. The ability to deliver with high certainty is seen as a strategic goal, and the producers believe that a good reputation in this respect will translate to higher prices, increased sales, and better contract terms. These benefits are not straightforward to include directly into the model objective. Instead, we have chosen to add requirements on the deliveries using risk measures.

Risk measures (Sect. 8.4) are often applied to monetary losses where undesirable deviations are limited or penalized in the objective. We apply risk measures to the physical flow to ensure that the producer is able to deliver the contracted volumes with a high degree of certainty. The target production assurance is valid both in the short term and in the long term, and we apply the risk measures on the operational scale. We show that our approach implies time consistency of the risk measures for multi-horizon trees.

Moreover, problems found in the literature often consider risk aversion which means that one seeks to *minimize* risk by including risk measures in the objective function. In contrast, our application example *limits* the risk through constraints.

That is, we consider risk acceptable as long as it is below a given threshold—on the other hand, risk minimization (in the objective function) allows one to end up with an optimal level of risk above such a threshold value. Moreover, the objective function in our application example is a function of several decision variables and the risk measure is another function of a subset of these. Most applications consider a much more direct relation between the objective and risk functions, e.g., profits and losses.

Finally, a stylized example in Sect. 8.5 illustrates the application of production assurance requirements by way of Average Value-at-Risk constraints on a model with a multi-horizon tree structure and shows how different modeling choices can affect the optimal solution.

8.2 Treating Uncertainty: Multi-horizon Scenario Trees

Decisions on a long-term or strategic level often define a framework for short-term or operational aspects. Hence, when finding strategic decisions such as investments into equipment, it is important to assess their impact on operations and vice versa, and an optimization model should take into account both decision horizons. This becomes even more relevant when these decisions must be found under uncertainty. In many situations, one can distinguish between uncertainty on a longer-term perspective (trends in the development of consumer prices or demand, volumes available for production in newly discovered natural gas reservoirs, technology development, climate change, etc.) and uncertainty on a shorter time scale (daily price or demand variations, weather variations). Obviously, this suggests that a stochastic model combining both time scales in a common scenario tree should distinguish between scenario tree nodes dedicated to strategic uncertainty and decisions and nodes dealing with the operational uncertainty and decision process.

A straightforward combination of the two kinds of nodes in a common scenario tree structure leads to the tree size quickly growing out of hand: To represent the operational conditions during a strategic time period adequately, one should include quite a number of realizations of the uncertain operational parameters and, hence, branchings at the strategic nodes in the considered period. These nodes representing short-term uncertainty must then be combined with the nodes representing long-term uncertainty in the next strategic time period. In particular, if short-term and long-term uncertainties are independent, the resulting tree may contain many duplicate values.

Figure 8.1 shows an example of combining long- and short-term uncertainty in a traditional scenario structure where ◯ represents strategic and □ operational nodes. The tree spans just three strategic time periods with one operational period each. Strategic uncertainty is represented by three branchings at the first strategic stage and two at the second while there are two possible realizations of the uncertain operational parameters at each operational stage. This small example yields 48 scenarios and, in total, 93 tree nodes (62 operational and 31 strategic).

However, strategic decisions often depend on the *overall* operational performance in the time since the previous strategic decision rather than directly on a specific short-term scenario (and the corresponding operational decisions). For example, decisions about investments in natural gas transport infrastructure rarely depend on the infrastructure's performance on a specific day but rather on how the infrastructure is *expected* to perform under varying daily conditions. In this case, a multi-horizon scenario tree structure is well suited. With such a structure, it is sufficient to branch at the strategic nodes while the operational nodes are embedded as subtrees associated with the respective strategic node. Hence, the operational feasibility and profitability of the decisions made in the strategic nodes can be tested on the corresponding subtrees. Moreover, testing infrastructure reliability typically requires many operational scenarios to ensure robustness under a vast variety of situations. This indicates that a multi-horizon tree structure is particularly well suited for such purposes; see also the discussion in Sect. 8.3.3.

Kaut et al. [9] discuss this approach in more detail and draw comparisons to traditional scenario tree structures, also with respect to the growth in tree sizes.

An example of a multi-horizon scenario tree is given in Fig. 8.1b. The tree has the same number of stages and of realizations of the uncertain strategic parameters as the traditional scenario tree in Fig. 8.1a but twice as many realizations of the uncertain operational parameters (i.e., a finer presentation of the short-term uncertainty). Here, we have 40 operational and 6 strategic scenarios while the tree contains just 50 nodes (whereof 40 operational and 10 strategic).

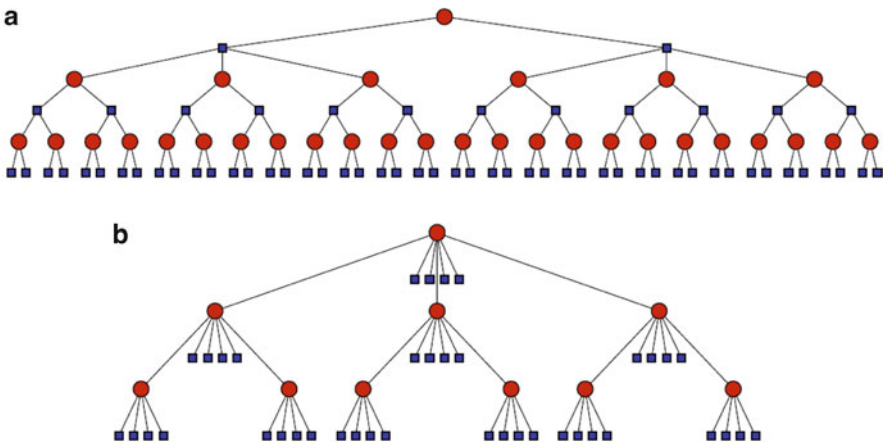

Fig. 8.1 Examples of traditional and multi-horizon scenario trees. (**a**) Traditional scenario tree structure with a combination of strategic and operational uncertainty. (**b**) Multi-horizon tree with the same number of strategic branchings but double number of operational branchings as the traditional tree

The multi-horizon scenario tree structure can be interpreted as a contingent scenario analysis of the operational problem for each strategic node. In general, it is a relaxation of the information structure represented by a traditional scenario tree.

If the strategic uncertainty is independent of the operational uncertainty and the strategic decisions do not depend on particular operational decisions at the previous strategic stage, the information structure is exactly the same as with the traditional approach. This condition is often satisfied quite easily. Moreover, in a multi-horizon tree structure, there is no connection between operational scenarios of two consecutive strategic nodes. Hence, the first operational decision associated with a strategic node should not depend on the last operational decision or state from the previous strategic period. This condition may require a careful definition of the strategic time periods; see also the discussion in Kaut et al. [9].

In the remainder of this chapter, the notation concerning uncertainty is chosen with a multi-horizon scenario tree structure in mind. For a strategic node $i \in \mathscr{N}^{Strat}$, we denote its time period by $\tau(i) \in \mathscr{T}$, relative probability by $\mathbb{P}_i^{Strat}$, and all operational nodes in the associated subtree by $j \in \mathscr{N}_i^{Op}$. The relative probability of an operational node $j \in \mathscr{N}_i^{Op}$ is denoted by $\mathbb{P}_j^{Op}$. Observe that a subtree representing the operational uncertainty is always associated with a certain strategic node. Hence, this strategic node can be interpreted as the root node of the considered subtree and the probabilities of the operational scenarios are considered only within the context of the respective subtree.

For the ease of notation we assume throughout this chapter that there are no bindings between consecutive operational time periods (e.g., due to storage modelling). That is, each operational node can be considered independent of other operational nodes. Hence, each operational node in a given subtree represents a single scenario. Consequently, we have just one operational time period in a subtree and we will, therefore, ignore the operational time index in the subsequent discussion. Moreover, the strategic node i is the parent node Pa_j of all operational nodes $j \in \mathscr{N}_i^{Op}$ in the associated subtree.

8.3 Application Example: The Ramona Optimization Model

The stochastic optimization model developed in the Ramona project combines both infrastructure and operational decisions under uncertainty in a common framework. It reflects, hence, both technological properties of the natural gas transport network and economic and business requirements.

Pressures and flows in one part of the network may influence transportation capacity in other infrastructure parts [14]. Such system effects must be taken into account when deciding about new investment in order to avoid a negative impact on the existing or future infrastructure. Hence, a portfolio perspective is advisable rather than an evaluation of single investment options in isolation and independently of the total system. For example, being able to address gas quality problems from new fields through blending in already existing facilities rather than investing in extra processing capacity can save investment costs. This means that, in addition to economics, operational aspects (physical processes and daily gas routing decisions) must be taken into account.

An increased focus on production assurance and security of supply makes it paramount to evaluate how the infrastructure will perform during daily operations and what the financial effects (costs and revenue) will be. For example, would a new pipeline allow to better satisfy delivery contracts in critical times or to route gas not bound in contracts to the most profitable markets? How would it affect gas flows in other pipelines? How does it affect operational costs and cash flow? Also the timing of investments is important for satisfying production obligations and developing new fields in a good way, e.g., allowing to reuse infrastructure.

Obviously, both investment and routing decisions are subject to various kinds of uncertainty. Discoveries of new reservoirs, gas composition and volumes in undeveloped reservoirs, or long-term changes (trends) in price and demand levels are examples of uncertainty on the strategic level. In contrast, uncertain parameters on the operational level may concern daily nominations in long-term delivery contracts or prices and demands at the markets. Also unplanned events such as network outages represent short-term uncertainty, reducing capacity in the affected infrastructure parts drastically reduced for some, often short, time and, hence, affecting production assurance. This combination of diverse kinds of long- and short-term uncertainty suggests the utilization of multi-horizon scenario trees in order to tackle realistically sized problem instances.

In order to further reduce the scenario tree size, one may consider only a representative selection of operational scenarios. For example, to estimate the profitability of the strategic decisions, it may be sufficient to study a few typical days in a year (spring, summer, winter, and a few variations). To test network flexibility and robustness, some extreme or "critical" scenarios are included. For example, this allows to assess the network's robustness in terms of production assurance, taking into account also unplanned events. Assigning (near-)zero probability to these "critical" scenarios, the feasibility of the strategic decisions also for these scenarios can be tested, but they do not affect profitability evaluations unduly.

The complete model constitutes a unified framework to analyze investment decisions under uncertainty in a portfolio perspective, taking into account physical properties of the network and the dynamics of short-term planning as well as long- and short-term uncertainty. Here, we present only the most important aspects of this multistage stochastic mixed-integer programming problem. Hellemo et al. [7] give an overview of a deterministic version of the model including a discussion of system effects and quality aspects while Hellemo et al. [8] present a comprehensive description of the full stochastic optimization model.

8.3.1 Model Overview

In the strategic nodes $i \in \mathcal{N}^{Strat}$, the (binary) decision variables $X^{Strat} = \{x_i^{Strat}, i \in \mathcal{N}^{Strat}\}$ refer to decisions about investments into network infrastructure elements (production facilities, pipelines, processing facilities, and markets). They establish

the framework for the decisions $X^{Op} = \{x_j^{Op}, j \in \mathscr{N}_i^{Op}\}$ in the associated operational nodes $j \in \mathscr{N}_i^{Op}$ which concern the flow f_{nj} through each network infrastructure element n and the pressures at its in- and outlets. These decisions, in turn, determine the cash flow and production assurance achievable with the found network design.

The objective of the model is to maximize the expected net present value (NPV) of investments and operations which is determined by the costs C_i^{Strat} and revenue R_i^{Strat} from decisions at all strategic nodes $i \in \mathscr{N}^{Strat}$ and costs C_j^{Op} and revenue R_j^{Op} from operations at the associated operational nodes $j \in \mathscr{N}_i^{Op}$:

$$NPV(X^{Strat}, X^{Op}) \tag{8.1}$$
$$= \sum_{i \in \mathscr{N}^{Strat}} \mathbb{P}_i^{Strat} \delta_{\tau(i)} \left(R_i^{Strat} - C_i^{Strat} + \sum_{j \in \mathscr{N}_i^{Op}} \mathbb{P}_j^{Op} \gamma_j \left(R_j^{Op} - C_j^{Op} \right) \right).$$

The costs C_i^{Strat} from investment decisions x_i^{Strat} at the node $i \in \mathscr{N}^{Strat}$ are mainly costs for installing new and decommissioning network infrastructure elements. In principle, they may depend on the strategic time period $\tau(i)$ or on the age of the element. The mathematical model uses two kinds of cost profiles to reflect this. However, we do not go into the details of these profiles here. Decommissioned network elements may have a positive salvage value contributing to revenue R_i^{Strat} directly arising from strategic decisions. The factor $\delta_{\tau(i)}$ denotes the discount factor at the time period of the strategic node i.

Operational costs C_j^{Op} are related to operating and maintaining the infrastructure. For each network element n, they are composed of operational expenditure (depending on the element's age), fixed (depending on calendar time), and variable costs (depending on the flow f_{nj}). Again, time-dependent profiles are used to express the different dependencies. As mentioned above, the operational nodes often represent only a selection of scenarios such that a scaling factor γ_j is applied to the values R_j^{Op} and C_j^{Op} which represents the weight of the considered operational node. For example, if all operational time periods have a length of one day and the strategic time period is a year, then the weights of all operational nodes associated with a strategic node must sum up to 365.

For a given market, the daily natural gas sales price may be stochastic, and the revenue R_j^{Op} achieved in an operational node $j \in \mathscr{N}_i^{Op}$ is the sum over sales at all markets.

Investment decisions are subject to constraints on the network elements' start-up (within a time window), shutdown, availability for production, and capacity. Operational decisions concern routing gas through the network and, obviously, the constraints on these decisions primarily model the physics of the network. The most important constraints express flow–pressure relationships and ensure mass balances as well as limits on the flow and pressure in each network element. The latter constraints take also care of capacity variations throughout the lifetime of an infrastructure element. Moreover, they can be used to model unforeseen events affecting

network capacity. Reservoir constraints ensure that the amount of gas produced at the single production facilities complies with yearly production plans and limits on the totally available gas in the corresponding reservoir.

8.3.2 Multi-Horizon Scenario Trees in the Ramona Model

A typical realistic model instance comprises about 200 network elements. While investment analysis may span a time horizon of between twenty and fifty years, operational decisions are found with a daily time resolution. Consequently, a three-stage stochastic model with 12 strategic periods, 10 branches per strategic node at each stage, and 10 operational profiles over 365 days will have about 100 million decision variables when using a multi-horizon scenario tree structure. With a traditional scenario tree structure, the model would be practically unsolvable—it would contain about 9 billion variables.

For this application, multi-horizon scenario trees represent an approximation of the information structure represented in a traditional scenario tree (cf. discussion in Sect. 8.2): Obviously, while production plans made at the strategic level give guidelines for the produced volumes to optimally deplete the reservoirs, the actual production depends on the operational scenarios. Hence, specific operational scenarios do indeed affect the decision space for the production plans in subsequent strategic periods to some degree. However, also the total volume in a reservoir available for production is not perfectly known, and the dependency of this volume on operational production decisions may be considered negligible.

8.3.3 Production Assurance Requirements

At the production side, production assurance requirements refer to the flow into the network relative to the production plan. On the other hand, at the consumption side, they consider the deviation of actually delivered volumes from the company's delivery obligations agreed upon in the contract with its customers. Moreover, production assurance may be measured at several levels: separately for each market or field, at a cluster of or at all markets fields, or even across the whole network. For our subsequent discussion, it is not important what exactly the notion refers to as we focus on a discussion of this concept in the context of the time and information structure provided by multi-horizon scenario trees. Therefore, we refrain from a network topology index for the concerned variables and parameters.

As an example, production assurance at a market can be found by analyzing *daily deliverability* over a given strategic time period. Typically, deliverability measures the deviation of the gas volume actually delivered in an operational node $j \in \mathscr{N}_i^{Op}$ (e.g., a day) against the demand, i.e., the nominated or contracted volume. Here, the nomination or contracted flow f_j^{Con} represents a random parameter while the delivered flow f_j^{Del} is a decision variable. Depending on the definition of

the production assurance requirement, f_j^{Del} may refer to the flow f_{nj} in a specific market node n but also to the aggregated flow in a cluster of (or all) market nodes. Obviously, a similar understanding holds for the nomination f_j^{Con}.

Often, there is no proper incentive to perform better than required (e.g., gas volume exceeding nominations may rather be sold in a more lucrative market). Hence, we define deliverability with a focus on cases where the delivered volume is insufficient compared to the nomination:

$$Del\left(f_j^{Del}, f_j^{Con}\right) = \min\left\{1, \frac{f_j^{Del}}{f_j^{Con}}\right\}, \quad \forall\, j \in \mathscr{N}_i^{Op},\ i \in \mathscr{N}^{Strat},$$

and ignore the (theoretically possible) option of delivering more than specified.

For the ease of notation, we suppress the reference to the volumes f_j^{Con} and f_j^{Del} when mentioning the deliverability in an operational scenario in the following: $Del_j = Del\,(f_j^{Del}, f_j^{Con})$.

Observe that production assurance requirements apply to underdeliveries compared to nominated gas volumes (or underproduction compared to planned gas volumes) while the profitability evaluations included in the objective function are based on the actually delivered (or produced) volumes and the gas price at the considered operational nodes. In other words, production assurance requirements do not consider the *value* of the gas. Rather, production assurance requirements must be satisfied no matter what the current gas price is. If the risk measure took into account gas prices, thus focusing on the risk of lost profit, a violation of delivery obligations would matter more when the gas price is high and less when it is low.

There is no clear-cut and unified way to specify production assurance requirements, and we discuss some ways to specify such requirements in the following.

One may state a threshold value PA_t below which the daily deliverability should not fall in all strategic nodes i in a time period t (i.e., for all nodes $i \in \mathscr{N}^{Strat}$ with $\tau(i) = t$). For example, one may require a deliverability of at least 99%. Due to the uncertainty about operational parameters—which also affects the daily deliverability—it may not be wise to formulate this requirement as a constraint to be satisfied under all circumstances:

$$Del_j \geq PA_{\tau(i)}, \quad \forall j \in \mathscr{N}_i^{Op},\ i \in \mathscr{N}^{Strat}.$$

This would result in a network infrastructure with a high degree of redundancy and, consequently, unduly high investment costs to ensure that this constraint is satisfied at any time. Instead one may allow a violation of the requirement, but encourage solutions ensuring a high degree of production assurance. For example, a penalty M may be imposed for each day $j \in \mathscr{N}_i^{Op}$ with insufficient deliverability:

$$M \max\{0, PA_{\tau(i)} - Del_j\},$$

which is then summed up in the objective function over the whole optimization horizon, taking into account the probabilities of the strategic scenarios and of the operational nodes within each subtree:

$$\sum_{i \in \mathcal{N}^{Strat}} \mathbb{P}_i^{Strat} \sum_{j \in \mathcal{N}_i^{Op}} \mathbb{P}_j^{Op} M \max\{0, PA_{\tau(i)} - Del_j\}. \tag{8.2}$$

This corresponds to penalizing the expected insufficient deliverability $\max\{0, PA_{\tau(i)} - Del_j\}$ over the whole horizon. Hence, the variability of this random variable cannot be taken into account properly and such a penalty term cannot control how the target values are satisfied over the operational nodes $j \in \mathcal{N}_i^{Op}$ associated with a strategic node $i \in \mathcal{N}^{Strat}$. In other words, a very low deliverability on one day and good performance else are considered comparable to a constant slight underperformance. Moreover, the penalty factor M must be chosen very carefully to achieve the desired results, weighting production assurance against expected net present value of investments and operations. As it is difficult to quantify such quality-oriented aspects, this is a rather daunting task.

The former challenge can be addressed by the following formulation: Given the threshold value PA_t for acceptable deliverability, set a limit on the percentage (or number) of days in the strategic node $i \in \mathcal{N}^{Strat}$ (with $\tau(i) = t$) where this threshold is not reached. This can be formulated by way of *chance constraints*. The threshold $\alpha_{\tau(i)}$ specifies the minimum percentage of operational scenarios (days) in strategic period i with sufficient deliverability:

$$\mathbb{P}\left\{Del_j \geq PA_{\tau(i)}, j \in \mathcal{N}_i^{Op}\right\} \geq \alpha_{\tau(i)}, \quad \forall i \in \mathcal{N}^{Strat}. \tag{8.3}$$

Such a formulation avoids the difficulty of quantifying the company's deliverability record. Since the operational scenarios represent a discretization of the distribution of the uncertain operational parameters, this constraint can be expressed by an LP formulation by introducing auxiliary binary and continuous variables [22, 27].

If, for example, it is required that "deliverability at a market node shall be over 0.99 on, at least, 97% of all days in a year" in all years during the optimization horizon (and, consequently, in all strategic tree nodes), constraint (8.3) would read

$$\mathbb{P}\left\{Del_j \geq 0.99, j \in \mathcal{N}_i^{Op}\right\} \geq 97\%, \quad \forall i \in \mathcal{N}^{Strat}.$$

In other words, this formulation ensures that the percentage of days with insufficient deliverability is not too high—but it also allows to underperform quite drastically in all those days.

The thread of excessive underperformance may be taken care of by setting a *lower limit on the average performance in the worst outcomes*. This way, some really low deliverability is still allowed, but only occasionally. However, it is difficult to define the "worst outcomes": If all days with a deliverability below PA_t are considered unacceptable, the requirement just leads to a lower average deliverability as there is no limit on the number of these days. Consequently, there is no incentive to perform better than PA_t.

Alternatively, one may sort all days $j \in \mathscr{N}_i^{Op}$ associated with a strategic node $i \in \mathscr{N}^{Strat}$ according to their deliverability and consider a given percentage $\tilde{\alpha}_{\tau(i)}$ of these days as the worst outcomes, no matter how "bad" they actually are. Then, the average deliverability on these days may be limited from below by a limit $\widetilde{Del}_{\tau(i)}$:

$$\mathbb{E}\{Del_j | Del_j \leq VaR_{\tilde{\alpha}_{\tau(i)}}(Del), j \in \mathscr{N}_i^{Op}\} \geq \widetilde{Del}_{\tau(i)}, \quad \forall i \in \mathscr{N}^{Strat}. \tag{8.4}$$

For example, one may require that the average deliverability on the 15% days with lowest deliverability in a strategic node i still shall be above 0.9:

$$\mathbb{E}\{Del_j | Del_j \leq VaR_{15\%}(Del), j \in \mathscr{N}_i^{Op}\} \geq 90\%, \quad \forall i \in \mathscr{N}^{Strat}.$$

Also this requirement can be reformulated to a set of linear constraints by introducing auxiliary continuous variables [22, 27].

Clearly, the deliverability targets PA_t (and threshold probabilities α_t) should be the same for all strategic nodes in a given time period t: The production assurance requirement should not depend on the gas network configuration or realizations of the uncertain strategic parameters. Moreover, they may also be the same for several or all strategic time periods t. The same holds for the lower limits $\widetilde{Del}_t$ and for the percentages $\tilde{\alpha}_t$ of the days with lowest deliverability within the time period t.

However, in general, there is no direct relationship between the target value PA_t and the limit $\widetilde{Del}_t$ or between the threshold α_t and the percentage $\tilde{\alpha}_t$ (although it might be more natural to assume a relationship between the latter than between the first). Intuitively, one may set $\tilde{\alpha}_t$ somewhat higher than α_t and / or $\widetilde{Del}_t$ somewhat lower than PA_t.

Evidently, production assurance is an operational concept and confined to a certain (strategic) time period, e.g., year or month. It is determined for a given infrastructure configuration (existing network and potential investment decisions) and requires, in order to give meaningful results, many operational scenarios associated with each strategic decision point. Hence, to compare or decide between several investment options under similar operational conditions, a traditional scenario tree structure would require a large degree of duplicate values. Consequently, one can solve only relatively small examples rather than realistic-sized cases.

Observe, however, that formulations (8.3) and (8.4) of production assurance requirements do not span several strategic time periods—they are considered independently for each strategic period. Hence, this aspect is well suited a model with a multi-horizon scenario tree structure. This tree structure allows many more operational scenarios for each strategic decision point than a traditional structure interspersing operational and strategic tree nodes in each scenario.

In the following section, we briefly introduce static and dynamic risk measures before we relate them to the multi-horizon scenario tree structure and the production assurance concepts discussed in Sect. 8.3.3. In particular, we turn our attention to the question of time consistency in a multistage setting.

8.4 Risk Measures

Section 8.3.3 outlined several approaches to model risk aversion when making decisions under uncertainty. On a more general level, risk measures as functionals on random variables have been studied intensely over the past decade and have become popular in particular in finance. The seminal paper by Artzner et al. [1] addresses axioms that are considered natural when quantifying risk by assigning a single number to the random variable representing potential outcomes. Krokhmal et al. [13] provide an overview over risk-modeling concepts in a static (single-period) setting.

The expectation operator $\mathbb{E}(\cdot)$ employed in formulation (8.2) represents the simplest form of a risk measure; it assigns the single number $\mathbb{E}Y$ to the possible outcomes represented by a random variable Y. For the example of production assurance described, the random parameter may be the daily nominations f_j^{Con} (and other uncertainties not discussed closer here), while the flow f_j^{Del} is the considered decision variable such that the random variable Y corresponds to the deliverability Del_j. The respective probabilities are modeled through a probability measure involved when calculating the expectation.

As mentioned above, this measure does not take into account properties of the considered random variable such as its variability, i.e., the distribution of the single values over all outcomes. Risk measures exploiting more properties of the random variable are, for example, chance constraints [exemplified by (8.3)] or the *Average Value-at-Risk* ($\mathbb{A}$V@R) as formulated in (8.4).

Chance constraints (8.3) ensure that the deliverability targets PA_t are satisfied with a certain probability as specified in the contract with the company's customers. However, this formulation treats any underdeliveries equally, no matter how large the shortfall is.

The $\mathbb{A}$V@R illustrated in (8.4) reflects another important risk measure which is also known as expected shortfall or Conditional Value-at-Risk (CVaR). This measure does not only take into account at which probability the demand is satisfied but also the level of demand satisfaction. This constraint is sometimes more conservative than a chance constraint. On the other hand, it is a convex constraint and certainly easier to handle computationally than, say, chance constraints. Our subsequent discussion will focus on this risk measure.

With our application in mind, the $\mathbb{A}$V@R of a random variable Y at the confidence level α can be defined generally as the expectation of all outcomes in the lower α-quantile of the probability distribution of Y:

$$\mathbb{A}\mathrm{V@R}_\alpha(Y) = \mathbb{E}\{Y | Y \leq V@R_\alpha(Y)\}. \tag{8.5}$$

However, in the case that the probability space contains atoms, this formulation has a drawback. In this case, the event $\{Y|Y \leq V@R_\alpha(Y)\}$ may have a probability other than α despite $V@R_\alpha(Y)$ being involved in the definition of this event. Exactly this situation occurs in the Ramona model involving finitely many (operational) scenarios of deliverability values as each has a (strictly) positive probability and, hence, is an atom.

Using the Fenchel–Moreau Theorem (cf. [26]), the $\mathbb{A}$V@R can be expressed by its dual formula:

$$\mathbb{A}\mathrm{V@R}_\alpha(Y) = \inf\left\{\mathbb{E}YZ | 0 \le Z \le \frac{1}{\alpha}, \mathbb{E}Z = 1\right\}. \tag{8.6}$$

This representation outlines that the mapping $Y \mapsto \mathbb{A}\mathrm{V@R}_\alpha(Y)$ is concave.

Alternatively, the $\mathbb{A}$V@R can be expressed as

$$\mathbb{A}\mathrm{V@R}_\alpha(Y) = \max_{q\in\mathbb{R}} \; q - \frac{1}{\alpha}\mathbb{E}(q-Y)_+,$$

where x_+ is the positive part, $x_+ := \max\{0, x\}$. This expression has been introduced by Rockafellar and Uryasev [19] while the general formulation is stated in Pflug [15]. It replaces the infimum in (8.6) by a maximum and has become popular in stochastic optimization (cf., e.g., [4]) and appears to be tailor-made for the model discussed here.

A risk constraint similar to the production assurance requirements presented in Sect. 8.3.3 may require that the Average Value-at-Risk of the random variable Y at the level α is above a given threshold $\tilde{q}$:

$$\mathbb{A}\mathrm{V@R}_\alpha(Y) \ge \tilde{q}.$$

8.4.1 Risk Measures in Multistage Optimization Problems

Intuitively, the static concept may be extended easily to a dynamic or multistage situation. However, due to relations between the decisions and parameters at the different stages affecting the properties of the risk measures, this is not quite straightforward and has spawned increased research interest in the recent years. For example, Kozmík and Morton [12] consider a stochastic programming problem structure with multiple recourse stages. They consider risk aversion, i.e., minimize risk and study stage-wise independent uncertain parameters and a risk measure as a function of the recourse value at each stage.

Risk measures may be applied separately at each stage of the underlying scenario tree or as a nested measure spanning several or all stages. We focus here on the former, conceptually simpler approach. More formally, we consider $\mathbb{R}$-valued random variables dependent on some previous decisions $\mathbf{x} \in \mathbb{X}$ and a random parameter $\xi \in \Xi$, that is, $Y = Y(\mathbf{x}, \xi) \in \mathbb{R}$ for the objective and (possibly different) random variables $Y_t^c = Y_t^c(\mathbf{x}, \xi) \in \mathbb{R}$ for the constraints, which are observed at the times $t \in \{0, \ldots, T\}$. The vector $\mathbf{x} = (x_0, \ldots, x_T)$ collects all decisions made at $T+1$ subsequent instants of time $t \in \{0, \ldots, T\}$. We use the notation $Y(\mathbf{x})$ also for the random variables $Y(\mathbf{x}) : \xi \mapsto Y(\mathbf{x}, \xi)$ (and $Y_t^c(\mathbf{x})$ for $Y_t^c(\mathbf{x}) : \xi \mapsto Y_t(\mathbf{x}, \xi)$, respectively). Importantly, the decisions up to time t are $x_0, \ldots, x_t$, and a random variable Y_t^c observed at that time t is determined by $x_0, \ldots, x_t$. Expressed in mathematical terms, it holds that

$$Y_t^c(\mathbf{x}) = Y_t^c(\mathbf{x}') \tag{8.7}$$

whenever $(x_0,\ldots,x_t) = (x'_0,\ldots,x'_t)$, where $\mathbf{x} = (x_0,\ldots,x_T)$ and $\mathbf{x}' = (x'_0,\ldots,x'_T)$.

Similar to the static formulation (8.5), an $\mathbb{A}$V@R measure with a (potentially different) level α_t can be applied to the random variable Y_t at each stage $t \in \{0,\ldots,T\}$ of the underlying scenario tree. Constraints may require that any $\mathbb{A}$V@R at a given stage t shall exceed a given threshold value q_t (recall that $\mathbb{A}$V@R is concave). Observe that, if this constraint is required to hold separately at each tree node at this stage, the $\mathbb{A}$V@R is considered conditionally on realizations of Y up to this stage. (Obviously, if there is only one $\mathbb{A}$V@R constraint involving all Y_t for given t, this measure does not depend on previous realizations.) This is made evident through the filtration $\mathscr{F}_t$ of the tree: the increasing sigma algebras $\mathscr{F}_t \subset \mathscr{F}_{t+1}$ represent the information available at time t [17].

Hence, a multistage stochastic optimization problem with risk constraints at each time period t can be formulated as

$$\begin{aligned}
\text{maximize} \quad & \mathbb{E}Y(\mathbf{x}) && && (8.8a)\\
\text{subject to} \quad & \mathbb{A}\text{V@R}_{\alpha_t}\left(Y_t^c(\mathbf{x})|\mathscr{F}_t\right) \geq q_t, && \forall t \in \{0,\ldots,T\}, && (8.8b)\\
& \mathbf{x} \in \mathbb{X}_0 \times \cdots \times \mathbb{X}_T, && && (8.8c)
\end{aligned}$$

where all $x_t, t \in \{0,\ldots,T\}$ are measurable with respect to the sigma algebra $\mathscr{F}_t$. The latter condition expresses the *nonanticipativity* constraints on the decisions x_t.

Observe that the risk measure applies to the random variables Y_t^c while the random variable Y employed in the objective function may be a different function of the decisions $\mathbf{x}$. For example, the Ramona model comprises risk constraints involving natural gas volumes delivered to the markets while maximizing the expected profit from these deliveries. Only the latter involves current market prices.

A sufficiently large sample size is necessary to get acceptable approximations, particularly as a non-biased estimator for the Average Value-at-Risk does not exist ([10]) in general. More specific, the number of considered realizations of Y should be of order $\frac{1}{\alpha} \approx \frac{1}{P(Y \leq VaR_\alpha(Y))}$. As a consequence, the size of multistage stochastic programming problems relying on a traditional scenario tree structure quickly grows out of hand if risk measures are applied not only to the leaf nodes at the final stage. Also Kozmík and Morton [12] point out the need for many scenarios at each stage and the resulting computational challenges when considering risk in a multistage setting: As only a small number of the realizations of the random parameter at each stage contribute to calculating the risk, a large number of nodes would be required. They suggest SDDP, i.e., sampling during the solution process, relying on stage-independent scenario trees. Note that this requirement of stage-wise independent random parameters excludes time-series models.

Alternatively, the scenario tree size may be reduced drastically without sacrificing model quality by utilizing the properties of the model at hand. For the Ramona model, we can distinguish clearly between operational and strategic decisions, and the risk measures apply only to all operational outcomes associated with a given strategic node. This indicates that the subtrees associated with each strategic node

at any stage can be of the required size and, consequently, risk constraints can be applied at "any" strategic stage throughout the optimization horizon.

With a multi-horizon scenario tree structure, one can calculate a risk measure at each strategic node i, spanning all operational outcomes (i.e., all operational nodes j representing days) associated with this node i. For all strategic nodes in a given strategic period t, that is, $\{i \in \mathcal{N}^{Strat} : \tau(i) = t\}$, the requirements are the same, i.e., they are characterized by the same parameters α_t and q_t. Consequently, decisions in any strategic node should be found such that (a) the risk requirement covering all operational nodes associated with this strategic node is satisfied, and (b) they allow to make decisions in all subsequent strategic nodes such that the corresponding risk requirements in these nodes are satisfied. In general, the operational scenarios associated with a strategic node are considered to be independent of the operational scenarios associated with other strategic nodes. However, due to b), operational scenarios associated with later strategic nodes in the same strategic scenario may affect earlier strategic decisions—in particular, if the strategic decision space in these later nodes is quite confined. We will resume these considerations in Sect. 8.4.2 discussing time consistency of dynamic risk measures.

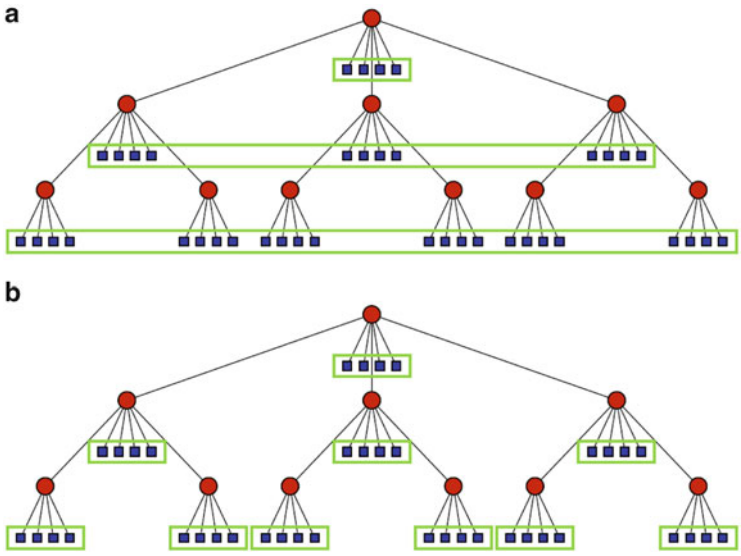

Fig. 8.2 Different scopes of risk measures. (**a**) Risk measure on a traditional scenario tree spanning all nodes at a stage. (**b**) Risk measures on a multi-horizon scenario tree spanning all operational nodes associated with a strategic node

8.4.2 Time Consistency

The principle of time consistency for multistage optimization problems derives from dynamic optimization [5] and has been analyzed from many perspectives in the lit-

erature; see, e.g., [2, 11, 24]. Although its meaning is intuitively evident, a common and widely accepted definition does, apparently, not exist yet. A common consensus, however, is that it means that decisions which are made at a certain stage of time should not be withdrawn at a later stage of time. That is, contradictions that may occur during the decision process should be excluded.

Carpentier et al. [3] informally formulate this principle as follows:

The principle of time (or dynamic) consistency. *The sequence of optimization problems is said to be dynamically consistent if the optimal strategies obtained when solving the original problem at time t_0 remain optimal for all subsequent problems. In other words, dynamic consistency means that strategies obtained by solving the problem at the very first stage do not have to be questioned later on.*

When discussing time consistency, one often distinguishes between optimization problems which involve a risk functional in the objective and which are solved at multiple, subsequent stages in time and problems which involve a risk functional in the constraints.

We will follow this distinction here and discuss time consistency for risk measures in the objective and the constraints separately. Then, we address time consistency on multi-horizon scenario trees.

8.4.2.1 Risk Measures in the Objective

Typically, multistage stochastic optimization problems with a risk measure in the objective employ a composition of risk measures at subsequent stages. Several results are known about such compositions of risk measures which often generalize initial properties of one-period risk measures ([6, 20, 21]). Schachermayer and Kupper [23] show that the only risk measure that is closed under time-consistent compositions is the functional

$$Y \mapsto u^{-1}\left(\mathbb{E}\left(u(Y)|\mathscr{F}_t\right)\right).$$

Moreover, Shapiro [25] states that the composition of risk functionals is not necessarily a law-invariant risk measure anymore, except for the expectation

$$\mathbb{AV@R}_1(\cdot) = \mathbb{E}(\cdot) \tag{8.9}$$

and the min-risk functional

$$\mathbb{AV@R}_0(\cdot) = \lim_{\alpha \searrow 0} \mathbb{AV@R}_\alpha(\cdot) = \operatorname{essinf}(\cdot). \tag{8.10}$$

Moreover, the composition lacks a natural interpretation: it is not clear what the Average Value-at-Risk of an Average Value-at-Risk could be. However, a composition of risk measures can be easily applied and it is convenient in computations. Often it is considered a—rather conservative—alternative to a single-period $\mathbb{AV@R}$ measure in the objective.

A possible way to overcome these challenges is by changing the level of the Average Value-at-Risk according to the representation

$$\mathbb{AV@R}_\alpha(Y) = \inf \mathbb{E}\, Z_t \cdot \mathbb{AV@R}_{\alpha \cdot Z_t}(Y|\mathscr{F}_t), \tag{8.11}$$

where the infimum is taken over all random variables Z_t, measurable with respect to $\mathscr{F}_t$, satisfying $\mathbb{E}\, Z_t = 1$ and $0 \leq Z_t \leq \frac{1}{\alpha}$. The essential difference to (8.6) is that the dual variable Z_t is measurable with respect to $\mathscr{F}_t$ and the level $\alpha \cdot Z_t$ in (8.11) is random itself [16].

This demonstrates that the level of the Average Value-at-Risk has to be changed in order to allow a combination of conditional Average Value-at-Risk measures. Equation (8.11) does not represent a composition but a change of measure instead (change of numéraire, cf. [18]).

Notably, the level of the Average Value-at-Risk represents the risk which the decision maker should accept in order to handle the optimization problem. Hence, one may conclude that the perception of risk may vary in different situations which, indeed, reflects a natural situation: Having observed a comfortable past which makes the initial objective more likely to achieve, a decision maker may be more relaxed in the future. Conversely, having observed a difficult past making the initial goal unlikely to be achieved, a decision maker may impose tougher conditions to ensure that the initial goal can still be achieved.

This is especially important in the case of rolling-horizon solution approaches.

8.4.2.2 Risk Measures in Constraints

The problem formulation (8.8) considers risk measures in the constraints at different levels. Studying time consistency of such risk measures, it appears natural to ask if $\mathbb{AV@R}_\alpha(Y|\mathscr{F}_t) \geq q$ means that also $\mathbb{AV@R}_\alpha(Y) \geq q$. More generally, if a random variable Y_1 is preferred over a variable Y_2 at a stage t, can it then be concluded that this random variable is preferable at an earlier stage as well; that is,

$$\mathbb{AV@R}_\alpha(Y_1|\mathscr{F}_t) \geq \mathbb{AV@R}_\alpha(Y_2|\mathscr{F}_t) \Longrightarrow \mathbb{AV@R}_\alpha(Y_1) \geq \mathbb{AV@R}_\alpha(Y_2).$$

Similar to the case of risk measures in the objective, this holds obviously for $\alpha = 1$ (the expectation) and $\alpha = 0$ (the min-risk functional).

Figure 8.3 illustrates that time consistency of the Average Value-at-Risk measure cannot be guaranteed for values of α other than 0 and 1: Assuming an Average Value-at-Risk at the level $\alpha = \frac{2}{3}$ at both stages, the example demonstrates that Y is acceptable when employing the criterion $\mathbb{AV@R}_{\frac{2}{3}} > 13$ at every subtree. However, applying the same criterion $\mathbb{AV@R}_{\frac{2}{3}} > 13$ to the complete problem, the variable Y is not acceptable. The simple Average Value-at-Risk is, therefore, not a time-consistent risk functional in this specified sense whenever $\alpha \in (0,1)$.

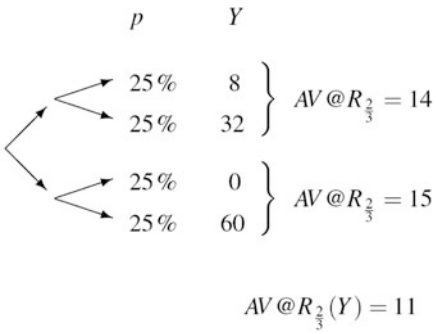

Fig. 8.3 This random variable Y satisfies $\mathbb{A}\mathrm{V@R}_{2/3} > 13$ for every partial observation in the subtree (specified by $\mathscr{F}_1$). However, the combined observation does not satisfy $\mathbb{A}\mathrm{V@R}_{2/3} > 13$. It is, however, correct, that $\mathbb{A}\mathrm{V@R}_\alpha \leq \min \mathbb{A}\mathrm{V@R}_\alpha(\cdot|\mathscr{F}_1)$

8.4.3 Consistency of Risk Measures in a Multi-horizon Tree Formulation

This section addresses time consistency of the multi-horizon problem (8.8). Note that its objective is an expectation and, moreover, the constraints in this problem are not compositions of risk measures as discussed in the previous section. The way the problem is formulated ensures its time consistency according to the principle formulated on page 192.

Proposition (Time consistency of the multi-horizon problem). *Let*

$$\mathbf{x}^* := (x_0^*, \ldots, x_T^*)$$

be an optimal solution of the multi-horizon problem (8.8)*. Then* $\mathbf{x}^*$ *solves also the problem with respect to the conditional probability measure* P^i*, where* $i \in \mathscr{N}_t^{Strat}$ *is an arbitrary (strategic) node at stage* t *and* $P^i(\cdot) = P(\cdot|i)$ *is the conditional probability satisfying* $P^i(i) = 1$.

Hence, the problem is time consistent in the sense of the principle given on page 192.

Remark. Incorporating the measure P^i—that is, conditioning on the node i—ensures that the tree process will move through the node i with probability one. The latter preposition ensures, therefore, that the problem can be reconsidered at a certain stage, and the initial solution will remain optimal even for the new subproblem which is reconsidered at a later time t ($i \in \mathscr{N}_t^{Strat}$). Hence, the problem is time consistent in the described sense and the initial solution does not have to be changed retrospectively.

Proof. Without loss of generality, we assume $P(i) > 0$.

Let $\mathbf{x}^*$ denote the optimal (minimal) solution, and assume that $\mathbf{x}^*$ is not optimal for the subproblem conditional on P^i. Denote the optimal solution of the subproblem with respect to P^i by $(x_0^*, \ldots, x_{t-1}^*, \tilde{x}_t^*, \ldots, \tilde{x}_T^*)$, and define the new decision as

$$\tilde{\mathbf{x}} := \begin{cases} (x_0^*, \ldots, x_{t-1}^*, \tilde{x}_t^*, \ldots, \tilde{x}_T^*) & \text{if } i \text{ is contained in the path,} \\ (x_0^*, \ldots, x_{t-1}^*, x_t^*, \ldots, x_T^*) & \text{else,} \end{cases}$$

which we apply to the initial problem.

The new strategy $\tilde{\mathbf{x}}$ is a potential solution of the initial problem as $\mathbb{X} = \mathbb{X}_0 \times \mathbb{X}_1 \times \ldots \mathbb{X}_T$. Moreover, $\tilde{\mathbf{x}}$ is feasible for the initial problem: Indeed, if $t' \leq t$, then, from (8.7), $Y_{t'}(\mathbf{x}) = Y_{t'}(\tilde{\mathbf{x}})$ as $x_{t'} = \tilde{x}_{t'}$ for all $t' \leq t$. Further, if $t' > t$, then

$$\mathbb{A}\mathrm{V@R}(Y_t(\mathbf{x})|\mathscr{F}_t) \geq q_t$$

in both cases, that is, no matter whether i is in the path or not.

Finally, the objective $\mathbb{E}Y(\tilde{\mathbf{x}})$ of the new strategy $\mathbf{x}$ is superior as

$$\begin{aligned} \mathbb{E}Y(\tilde{\mathbf{x}}) &= \mathbb{E}\mathbb{E}(Y(\tilde{\mathbf{x}})|\mathscr{F}_t) \\ &< \mathbb{E}\mathbb{E}(Y(\mathbf{x})|\mathscr{F}_t) = \mathbb{E}Y(\mathbf{x}) \end{aligned}$$

due to the assumption $P^i(i) > 0$ and since $\tilde{\mathbf{x}}$ is better than $\mathbf{x}$ on the node i.

Summarizing, $\tilde{\mathbf{x}}^*$ is a better strategy than $\mathbf{x}^*$ on the entire tree. This, however, contradicts the assumption that $\mathbf{x}^*$ is optimal. Hence, the strategy $\mathbf{x}^*$ is also optimal for the subproblem conditioned on P^i. This proves the assertion. □

Observe that the argument is valid also for operational nodes $j \in \mathscr{N}_i^{Op}$ associated with any strategic node $i \in \mathscr{N}^{Strat}$.

8.5 Illustrative Example

To illustrate the implementation of $\mathbb{A}\mathrm{V@R}$ on multi-horizon trees we use a stylized example. We show in the example that different ways of modeling risk aversion can change the optimal decisions in the optimization model. As a case study, we consider a network that consists of a field connected to a market through a single pipeline. This is illustrated in Fig. 8.4. The production capacity in the field node is assumed to not constrain our solution, but the pipeline that connects the two nodes has a capacity limit of 100 units. Furthermore, we assume that the market price is fixed at 1 million per unit, while the demand is stochastic. The scenario tree that we use in our example consists of two strategic periods and three strategic nodes (i.e., $i \in \{1,2,3\}$; see Fig. 8.5), each with an associated operational subtree.

We represent the demand uncertainty by 100 equiprobable scenarios in each of the operational subtrees. The demand uncertainty in the subtrees in strategic nodes 1 and 2 is identical and uniformly distributed between 50.5 and 100 units (plot to

Fig. 8.4 The simple network used in our example, consisting of a single field that supplies a market through a pipeline

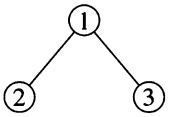

Fig. 8.5 Strategic nodes in the tree

the left in Fig. 8.6), while the demand uncertainty in the subtrees in strategic node 3 is uniformly distributed between 52.5 and 102 units (plot to the right in Fig. 8.6). The probability of strategic nodes 2 and 3 is equal (0.5).

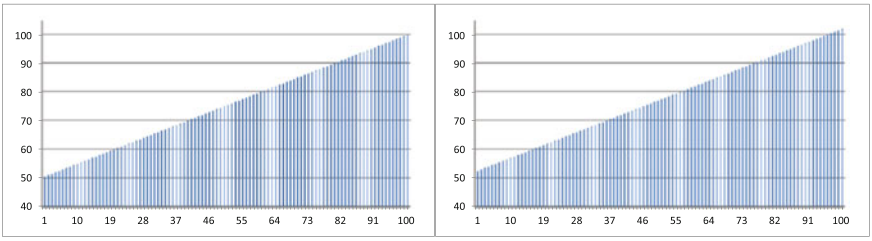

Fig. 8.6 Demand scenarios. The *left plot* shows the demand realizations in the operational subtree linked to the strategic nodes 1 and 2 (uniformly distributed between 50.5 and 100), while the *right plot* shows the slightly higher demand realizations for the subtree in strategic node 3

The company has an investment opportunity that will increase the pipeline capacity from 100 units to 110 units. The cost of this capacity increase is 0.1 million. To simplify our model we assume that the network is operated only in the two strategic periods considered in our example. This means that there are no end-of-horizon effects in our model. We also disregard discount rates and, as there is only a single field, pipeline, and market, we ignore network element indices. This simplified investment model can then be formulated as

$$\max_{\substack{\lambda_i, f_j \\ i \in \mathcal{N}^{Strat}, j \in \mathcal{N}_i^{Op}}} \sum_{i \in \mathcal{N}^{Strat}} \mathbb{P}_i^{Strat} \sum_{j \in \mathcal{N}_i^{Op}} \mathbb{P}_j^{Op} p f_j - \sum_{i \in \mathcal{N}^{Strat}} \mathbb{P}_i^{Strat} \lambda_i I, \tag{8.12}$$

where p is the price in the market, f_j is the volume sold in the market in operational scenario j, λ_i is the (binary) investment decision, and I is the investment cost. The production, flow in the pipeline, and sale in the market are constrained by the pipeline capacity K and the market demand F_j. The company can invest in additional capacity L. The set $\mathcal{N}_{A(i)}^{Strat}$ contains all ancestor nodes for node i as well as the node

i itself (i.e., all the nodes on the path from the root node to node i):

$$f_j \leq K + \sum_{i' \in \mathcal{N}^{Strat}_{A(i)}} \lambda_{i'} L, \quad j \in \mathcal{N}_i^{Op}, i \in \mathcal{N}^{Strat}, \tag{8.13a}$$

$$f_j \leq F_j, \quad j \in \mathcal{N}_i^{Op}, i \in \mathcal{N}^{Strat}. \tag{8.13b}$$

The company can make the investment only once in each scenario:

$$\sum_{i' \in \mathcal{N}^{Strat}_{A(i)}} \lambda_{i'} \leq 1, \quad i \in \mathcal{N}^{Strat}. \tag{8.14}$$

We can then solve the profit-maximizing model (8.12)–(8.14) to find the optimal investment decision. The solution to this problem is trivial since the only node where the capacity extension would influence the revenue is strategic node 3 (the demand associated with strategic nodes 1 and 2 is already covered by the capacity of the pipeline without the investment). The additional expected revenues from having a capacity of 110 units in node 3 are 0.05 million (additional sales in the four scenarios where demand exceeds 100). Since these revenues are smaller than the investment cost, the investment will not be made in any of the strategic nodes.

Let us now consider how risk measures may influence this solution. We assume that the company that operates the field has an obligation to deliver according to the demand level in the market node. The performance is regulated with an $\mathbb{A}$V@R constraint enforcing that the expected delivery rate in the worst 5% of the scenarios should be at least 0.995 (meaning that the expectation of the actually delivered volumes divided by the demand in the 5% worst scenarios should be at least 0.995). Figures 8.7 and 8.8 show two different ways of implementing this $\mathbb{A}$V@R constraint. Figure 8.7 illustrates the traditional approach where the $\mathbb{A}$V@R constraint is based on all operational observations within a given *time period* while Fig. 8.8 illustrates the approach used in the Ramona model. In this case, a separate $\mathbb{A}$V@R constraint refers to all operational subtrees associated with a strategic *node*. In the following, we show that these different $\mathbb{A}$V@R calculations can indeed influence the investment decision in the model.

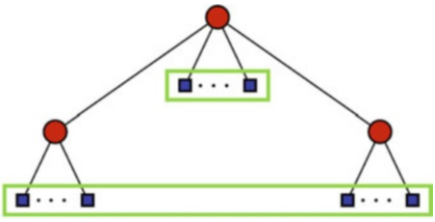

Fig. 8.7 Stage-wise constraints on $\mathbb{A}$V@R, involving all operational nodes in a strategic period. With this implementation of $\mathbb{A}$V@R, the optimal investment decision in our model is to not invest in capacity extension in any of the strategic nodes

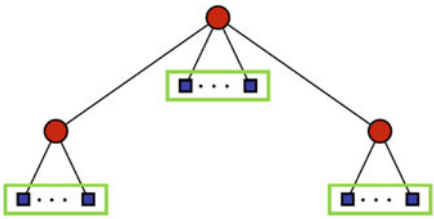

Fig. 8.8 Node-wise constraints, involving all operational nodes associated with a strategic node. With this implementation of the $\mathbb{A}$V@R constraints, the optimal decision in our example is to invest in additional pipeline capacity in strategic node 3

The mathematical formulation of the $\mathbb{A}$V@R constraints that include all operational observations within a given time period can be given as

$$\mathbb{E}\{Del_j | Del_j \leq VaR_{\tilde{\alpha}_t}(Del), j \in \widetilde{\mathcal{N}_t^{Op}}\} \geq \widetilde{Del}_t, \quad \forall t \in \mathcal{T}, \tag{8.15}$$

where the set $\widetilde{\mathcal{N}_t^{Op}}$ includes all operational nodes linked to a strategic time period t. The mathematical formulation used in the Ramona model can be given as

$$\mathbb{E}\{Del_j | Del_j \leq VaR_{\tilde{\alpha}_{\tau(i)}}(Del), j \in \mathcal{N}_i^{Op}\} \geq \widetilde{Del}_{\tau(i)}, \quad \forall i \in \mathcal{N}^{Strat}. \tag{8.16}$$

For more explanation on this $\mathbb{A}$V@R formulation, see the discussion linked to Equation 8.4 on page 187.

In our example, $\widetilde{Del}_t$ and $\widetilde{Del}_{\tau(i)}$ are equal to 0.995 while $\tilde{\alpha}_t$ and $\tilde{\alpha}_{\tau(i)}$ are equal to 5%.

First, let us consider the $\mathbb{A}$V@R constraints that are based on all observations in a time period. We already know that the investment project is not profitable, so to solve the model we only need to check if the $\mathbb{A}$V@R constraint holds. For the subtree linked to the first strategic node, this constraint is clearly satisfied, since the demand is met in all scenarios (the expected delivery rate in the 5% worst scenarios is 1). Considering the operational subtrees linked to the second strategic period, we find that the 5% worst scenarios have an expected delivery rate of 0.995 without any investments (calculated as the expected delivery rate in the 10 scenarios with highest demand). This means that including the $\mathbb{A}$V@R constraint on the model will not alter the optimal decisions.

Now, let us study how $\mathbb{A}$V@R constraints on all operational subtrees linked to a strategic node will influence the optimal solution from the model. Again, we know that the investment option alone is not profitable and we only need to check the $\mathbb{A}$V@R constraints. Obviously, in the operational subtrees linked to strategic nodes 1 and 2, these constraints are satisfied even without the investment (the demand will not exceed the original capacity of the pipeline). In the operational subtree linked to strategic node 3, however, the expected delivery rate of the 5% worst scenarios (the 5 scenarios with highest demand) is 0.990 without the investment. If the pipeline capacity is increased to 110, the $\mathbb{A}$V@R constraint is also satisfied in these

operational scenarios (the new expected delivery rate is 1). This means that the inclusion of $\mathbb{A}$V@R constraints in every subtree will force an investment in additional pipeline capacity in strategic node 3.

Obviously, this example is simplified and rather far removed from real investment decisions. The influence of the modeling choice of $\mathbb{A}$V@R constraints on the decision space and, hence, the optimal decisions found by the model is, however, a general result. We chose a simple example to transparently illustrate this effect. We can also note that while the first approach to modeling $\mathbb{A}$V@R is not time consistent, the second approach is. It can be easily seen that the optimal decisions on a rolling horizon will change when using the first approach: it will be necessary to change the original decision of not investing in strategic node 3. The second approach to modeling $\mathbb{A}$V@R is, however, time consistent, and the decisions will not change if we consider a rolling-horizon solution approach.

Acknowledgements The presented work builds on research performed in the Ramona project (The Research Council of Norway, project number 175967) on regularity and security of supply for natural gas transport. We thank our industrial partners Statoil and Gassco for many inspiring discussions and support.
We would also like to thank Michal Kaut at SINTEF for his valuable help while preparing this manuscript.

References

1. Artzner P, Delbaen F, Eber JM, Heath D (1999) Coherent measures of risk. Mathe Financ 9:203–228
2. Artzner P, Delbaen F, Eber JM, Heath D, Ku H (2007) Coherent multiperiod risk adjusted values and Bellman's principle. Ann Oper Res 152:5–22. doi: 10.1007/s10479-006-0132-6
3. Carpentier P, Chancelier JP, Cohen G, de Lara M, Girardeau P (2012) Dynamic consistency for stochastic optimal control problems. Ann Oper Res 200(1):247–263. doi:10.1007/s10479-011-1027-8
4. Fábián CI (2008) Handling CVaR objectives and constraints in two-stage stochastic models. Eur J Oper Res 191:888–911
5. Fleming WH, Soner HM (2006) Controlled markov processes and viscosity solutions. Springer, New York
6. Gutjahr WJ, Pichler A (2013) Stochastic multi-objective optimization: a survey on non-scalarizing methods. Ann Oper Res. doi:10.1007/s10479-013-1369-5
7. Hellemo L, Midthun K, Tomasgard A, Werner A (2012) Natural gas infrastructure design with an operational perspective. Energ Procedia 26:67–73. Proceedings of the 2nd Trondheim Gas Technology Conference
8. Hellemo L, Midthun K, Tomasgard A, Werner A (2013) Multi-stage stochastic programming for natural gas infrastructure design with a production perspective. In: Gassmann HI, Wallace SW, Ziemba WT (eds) Stochastic programming: applications in finance, energy, planning and logistics, world scientific series in finance. World Scientific, Singapore

9. Kaut M, Midthun KT, Werner AS, Tomasgard A, Hellemo L, Fodstad M (2012) Dual-level scenario trees – scenario generation and applications in energy planning. Optimization Online. URL http://www.optimization-online.org/DB$_$HTML/2012/08/3551.html. Report 2012/08/3551
10. Ko B, Russo RP, Shyamalkumar ND (2009) A note on the nonparametric estimation of the CTE. ASTIN Bull 39(2):717–734
11. Kovacevic R, Pflug GCh (2009) Time consistency and information monotonicity of multiperiod acceptability functionals. In: Albrecher H, Runggaldier WJ, Schachermayer W (eds) Advanced financial modelling. In Radon Series on Computational and Applied Mathematics, vol 8. de Gruyter, Berlin, pp 347–369
12. Kozmìk V, Morton DP (2013) Risk-averse stochastic dual dynamic programming. Optimization Online. URL http://www.optimization-online.org/DB$_$HTML/2013/02/3794.html
13. Krokhmal P, Zabarankin M, Uryasev S (2011) Modeling and optimization of risk. Surv Oper Res Managem Sci 16:49–66
14. Midthun KT, Bjørndal M, Tomasgard A (2009) Modeling optimal economic dispatch and system effects in natural gas networks. Energ J 30:155–180
15. Pflug GCh (2000) Some remarks on the value-at-risk and the conditional value-at-risk. In: Uryasev S (ed) Probabilistic constrained optimization: methodology and applications. Kluwer Academic Publishers, Dordrecht
16. Pflug GCh, Pichler A (2011) Multistage optimization. Stochastic Programming E-Print Series. URL http://edoc.hu-berlin.de/series/speps/2011-3/PDF/3.pdf
17. Pflug GCh, Römisch W (2007) Modeling, measuring and managing risk. World Scientific, River Edge
18. Pichler A (2013) Evaluations of risk measures for different probability measures. SIAM J Optim 23(1):530–551. doi:http://dx.doi.org/10.1137/110857088
19. Rockafellar RT, Uryasev S (2000) Optimization of conditional value-at-risk. J Risk 2:21–41
20. Ruszczyński A, Shapiro A (2006) Conditional risk mappings. Math Oper Res 31(3):544–561
21. Ruszczyński A, Shapiro A (2006) Optimization of convex risk functions. Mathe Oper Res 31(3):433–452
22. Sarykalin S, Serraino G, Uryasev S (2008) Value-at-risk vs. conditional value-at-risk in risk management and optimization. Tutorials in Operations Research, INFORMS
23. Schachermayer W, Kupper M (2009) Representation results for law invariant time consistent functions. Math Financ Econ 2(3):189–210. doi:10.1007/s11579-009-0019-9
24. Shapiro A (2009) On a time consistency concept in risk averse multistage stochastic programming. Oper Res Lett 37(37):143–147
25. Shapiro A (2012) Time consistency of dynamic risk measures. Oper Res Lett 40(6):436–439. doi:10.1016/j.orl.2012.08.007

26. Shapiro A, Dentcheva D, Ruszczyński A (2009) Lectures on stochastic programming. In MQS-SIAM Series on Optimization, vol 9, SIAM, Philadelphia
27. Zenios SA (2007) Practical financial optimization. Decision making for Engineers. Blackwell-Wiley, New York

Chapter 9
Controlled Islanding as Robust Operator Response Under Uncertainty

A. Grothey, W. Bukhsh, K.I.M. McKinnon, and P.A. Trodden

Abstract In the past decade there have been multiple high-profile cases of cascading blackouts, often resulting in the disconnection of tens of millions of consumers in large areas. It appears that in hindsight many of these disturbances could have been prevented by timely interventive action. In the actual cases, however, lack of complete knowledge about the state of the system undergoing a blackout event has prevented such action. This chapter reviews approaches to the problem of finding optimal interventions for a power system in the early stages of a cascading blackout. Conceptually the problem is one of optimization under uncertainty or robust optimization: the goal is to find a set of corrective actions that will guarantee power supply to as many customers as possible, in all, or at least most, of the possible states that the system may be in. To tackle the problem directly as a stochastic or robust optimization problem is intractable due to the complexities involved, foremost the number of possible states that would have to be considered. We argue, guided by example, that a robust response is to disconnect lines in such a manner as to create an island containing the affected part of the network. We give an overview of such approaches, notably those involving mixed-integer programming to directly design islands that admit a stable steady-state operating point.

A. Grothey (✉)
School of Mathematics, University of Edinburgh, James Clerk Maxwell Building, King's Buildings, Mayfield Road, Edinburgh EH9 3JZ, UK
e-mail: A.Grothey@ed.ac.uk

W. Bukhsh • K.I.M. McKinnon
University of Edinburgh, Edinburgh, UK

P.A. Trodden
Department of Automatic Control & Systems Engineering, University of Sheffield, Mappin Street, Sheffield, S1 3JD

R.M. Kovacevic et al. (eds.), *Handbook of Risk Management in Energy Production and Trading*, International Series in Operations Research & Management Science 199,
DOI 10.1007/978-1-4614-9035-7_9, © Springer Science+Business Media New York 2013

9.1 Introduction

The last decade has seen a number of notable cases of wide-area blackouts as a consequence of severe disturbances and cascading failures. In 2003, separate blackouts in Italy [13], Sweden/Denmark [19] and USA/Canada [31] affected millions of customers. The August 2003 blackout in Northeastern America alone disconnected 50 million people and 62 GW to an area spanning eight states and two provinces. The wide-area disturbance in 2006 to the European system caused the system to split in an uncontrollable way [14], forming three islands.

While the exact causes of wide-area blackouts differ from case to case, some common driving factors emerge. Modern power systems are being operated closer to limits: liberalization of the markets, and the subsequent increased commercial pressures and change in expenditure priorities, has led to a reduction in security margins [3, 7, 9]. A more recently occurring factor is increased penetration of variable distributed generation, notably from wind power, which brings significant challenges to secure system operation [23]. In addition power systems in developing countries are typically operating close to their operating limits, contributing to their vulnerability to cascading blackouts [2].

Although automated preventive and corrective systems exist to ameliorate the effects of system faults, aversion of a cascading blackout in severe disturbance events often requires active intervention of the system operator to manually adjust generation output, open or close lines or shed some load on the system. The correct response, however, is dependent on the exact state of the system and being able to pinpoint the exact nature and extent of the disturbance. Studies of historical blackout events however show that in many cases control room staff were unaware of the exact state of the system and the nature of the disturbance. Causes include failure of monitoring and protection equipment, software failures, lack of adherence to protocol, human error and even unawareness that there is a problem with the security systems at all. If, as in these cases, the system state is unknown, it becomes increasingly difficult to take controlled action to avert the impeding cascading blackout.

Our aim in this chapter is to investigate possible ways to find a robust operator response to avert an impeding cascading blackout given that there is significant uncertainty about the current state of the system. Conceptually the problem is one of optimization under uncertainty: Given a set of scenarios that describe the possible states the system could be in, the operator seeks a set of actions that guarantee that the system survives in each of these states or, since that is likely not possible, minimizes the expected loss (in terms of load shed or any other socioeconomic measure) of the system disturbance. Such an approach would naturally lead to a robust or stochastic optimization model.

Indeed we present such a conceptual model in Sect. 9.5. However, there are a number of problems associated, starting from the difficulty of obtaining input data: what are the possible system scenarios and their respective probabilities, to the sheer complexity of such a problem (which would be mixed integer, nonlinear and stochastic, potentially with a huge number of scenarios). We therefore concentrate on alternative, computationally tractable ways of optimizing the operator response.

Typically the initial system disturbance is limited to a relatively small geographic area. A possible robust solution is to electrically isolate the affected area from the remaining system, thus forming an *at-risk* island. We argue that such an islanding solution is often the optimal solution to the problem of finding a robust response to the system emergency. Indeed for several large disturbance events, e.g. [31], studies have shown that a wide-area blackout could have been prevented by intentionally splitting the system into islands [36]. Thus we tackle the uncertainty by splitting the at-risk area from the remainder of the network instead of dealing with every possible system state separately, decreasing the complexity of the model.

In the second half of this chapter we will focus on several concrete models that have been suggested to deal with the conceptual and computational complexity of islanding for blackout prevention.

The models primary target is to design islands such that one of the islands will isolate the system disturbance; it will contain all the lines and generators which are either known to be faulty or whose state is uncertain. This is the *at-risk* section. The remaining islands should contain only those parts of the network whose state is certain. Islands are required not only to be load–generation balanced but also to admit a feasible steady-state transmission solution that respects physical network operation constraints. Typically this requires the shedding of loads in some or all of the islands; the objective is to find such an islanding scheme that minimizes a combination of load assigned to the risky island and load shed in the remaining islands.

Due to the nonlinear nature of the physical network constraints the resulting model is a mixed-integer nonlinear programming (MINLP) problem. To make it computational tractable, two schemes are suggested. In the first, a standard linear approximation to the physical power transmission equations (the "DC" model) is used, resulting in a mixed-integer linear programming (MILP) formulation. A disadvantage of the DC power flow model is that reactive power flows and bus voltages are not considered. As a result it may not be possible to operate islands produced by this model, even if some corrective post-islanding action, such as load shedding, is taken. In the second approach a piecewise linear (PWL) approximation to the full AC power flow equations is developed. Again this model leads to an MILP formulation, although now with a larger number of binary variables. Results obtained for test networks demonstrate the advantage of considering voltage and reactive power when deciding how to island, since instances of AC-infeasible islands are eliminated.

The organization of the paper is as follows. In the following section we review previous work on the islanding of power systems. In Sect. 9.3 we review some necessary background of power system operation and their reaction to contingencies and review the standard power flow models. An analysis of the events leading to some of the prominent blackouts from the last decade is given in Sect. 9.4 and we discuss lessons that can be learnt in order to actively prevent cascading blackouts. A conceptual nonlinear mixed-integer stochastic programming model is derived in Sect. 9.5 and a case study to shed light on the possible solution space of such a model is presented in Sect. 9.6. In Sect. 9.7 we present an islanding model that is still nonlinear and mixed integer, but has now removed the stochastic aspect. Section 9.8 presents two linear approximations of this model: first a "DC"-based approximation

that includes line losses and then a piecewise linear SOS-set approximation. We give a brief summary of the computational performance of the models in Sect. 9.9 and discuss possible extensions in Sect. 9.10.

9.2 Islanding of Power Systems

Controlled islanding or *network splitting* of power system is not a new concept. In the traditional sense islanding is used to protect small self-contained load-generator pairs from system disturbances [26] or to isolate subcomponents of the network that are prone to known problems in order to create a safe fall back state from which it is easy to reconnect the network [4]. While systems of this kind are usually small scale, islanding along pre-determined break lines is used in Japan to prevent the spread of system disturbances into the Tokyo metropolitan area and has been successful in preventing at least one potential blackout event [1].

In cases where islanding is done to protect fundamentally self-sufficient parts of the network, the splitting is done along pre-determined breaks, and the system is specifically designed to survive these network splitting events. Dynamic network splitting of a relatively tightly connected network has been initially studied as a tool to ease power flow computations ([22] and references therein). Islanding used as a dynamic protection mechanism to limit the spread of system disturbances is more recent. The challenge is that, if an island is to be feasible, it must satisfy both static constraints—load–generation balance, network constraints, system limits—and dynamic constraints, i.e. for electromechanical and voltage stability. Moreover, the act of islanding must not cause a loss of synchronism or voltage collapse. It is not computationally practical to tackle all these aspects of the problem simultaneously within a single optimization, and approaches in the literature differ according to which aspect is treated as the primary objective. Additionally different search methods have been proposed for identifying the island boundaries.

Most approaches concentrate on identifying sections of the network that are in some sense weakly linked and then employ a graph partitioning or clustering algorithm to create the islands. Wang et al. [33] and Peiravi and Ildarabadi [24] use weighted multilevel k-means where the weights are obtained by the power flow along the lines in the pre-disturbance steady-state solution. In [22, 25] the network is clustered according to electrical distances (measure by line susceptances) using centroid sorting. Buses that are electrically close to each other are placed into the same cluster. Wang et al. [34] employed a power flow tracing algorithm to first determine the domain of each generator, i.e. the set of load buses that "belong" to each generator. Subsequently, the network is coarsely split along domain intersections before refinement of boundaries to minimize imbalances.

Spectral methods [16, 20, 36] guide the partitioning by using the eigenvector corresponding to the smallest nonzero eigenvalue of the weighted graph Laplacian (the *Fiedler vector*), where edge weights are chosen to represent the strength of the link.

In other approaches the primary objective is to split the network into electromechanically stable islands, commonly by splitting so that generators with coherent oscillatory modes are grouped. If the system can be split along boundaries

of coherent generator groups while not causing excessive imbalance between load and generation, then the system is less likely to lose stability. Determining the required cutset of lines involves, as a secondary objective, considerations of load–generation balance and other constraints; algorithms include exhaustive search [37], minimal-flow minimal-cutset determination using breadth-/depth-first search [32], and graph simplification and partitioning [35]. The authors of [18] propose a framework that, iteratively, identifies the controlling group of machines and the contingencies that most severely impact system stability and uses a heuristic method to search for a splitting strategy that maintains a desired stability margin. A review of different (graph theoretic) approaches to use islanding to prevent blackouts is given in [21].

In most approaches the creation of islands that allow a feasible power flow, or are even just load–generation balanced, is not a primary objective of the islanding algorithm. In some cases power flow analysis is performed on the considered islands to exclude strategies that violate operating constraints, e.g. line limits [28].

An alternative to graph partitioning methods are methods based on mixed-integer programming. Integer programming has many applications in power systems, but its use in network splitting and blackout prevention is limited. Bienstock and Mattia [10] propose an IP-based approach to the problem of designing networks that are robust to sets of cascading failures and thus avoid blackouts; the binary decision is whether to upgrade a line's capacity. The authors of [15] propose a method for optimal transmission switching for the problem of minimizing the cost of generation dispatch by selecting a network topology to suit a particular load. In common with the formulation presented here, binary variables represent switches that open or close each line and the DC power flow model is used, resulting in an MILP problem. Finally Fan et al. [12] propose model for controlled islanding where binary variables indicate islanding decisions and a linear "DC" network model is used.

9.3 Power System Operation

Consider a network $\mathscr{N}$ that comprises a set of buses $\mathscr{B} = \{1, 2, \ldots, n^{\mathrm{B}}\}$ and a set of lines $\mathscr{L}$. The two vectors F and T describe the connection topology of the network: a line $l \in \mathscr{L}$ connects bus F_l to bus T_l. There exists a set of generators $\mathscr{G}$ and a set of loads $\mathscr{D}$. A subset $\mathscr{G}_b$ of generators is attached to bus $b \in \mathscr{B}$; similarly, $\mathscr{D}_b$ contains the subset of loads present at bus $b \in \mathscr{B}$.

Power is distributed through the network by alternating current. Unlike other utility networks such as gas, water or telecommunication, for power networks the operator has no direct influence over the routing of the power; rather the network will arrange itself into a steady state that is governed by Kirchhoff's laws. The state of the network can be described by the voltage level v_b and phase angle δ_b at each bus $b \in \mathscr{B}$ in the network. Voltages are measure in per units (p.u.) against the nominal bus voltage whereas phase angles are measured against some arbitrary reference bus. Current is driven through lines by differences of voltages and phase angles at each end of the line. It is convenient to decompose an AC power flow into its real (i.e. the net energy flow over a cycle) and reactive components. The characteristics

of a line l are described by its conductance g_l (the reciprocal of its resistance) and its susceptance b_l.

The real and reactive power injections into a line l from bus b to b' at bus b are given by the power injection equations

$$p_{bb'} = g_l v_b^2 - v_b v_{b'} \left(g_l \cos\theta_l + b_l \sin\theta_l \right), \tag{9.1a}$$

$$q_{bb'} = -b_l v_b^2 - v_b v_{b'} \left(g_l \sin\theta_l - b_l \cos\theta_l \right), \tag{9.1b}$$

where $\theta_l = \delta_b - \delta_{b'}$ is the phase angle difference across line l. Note that these are not symmetric: the power injection at bus b is not the negative of the power injections at the opposite bus b', representing the effect of line losses. Here and in the following we interchangeably use the index l or bb' for a line; the latter is used where we want to make the index of its start and end bus explicit. We also use the notation v to refer to the vector of all bus voltages or $v_{\mathscr{G}}$ for the voltages at generator buses only. Real and reactive power balances are imposed at each bus $b \in \mathscr{B}$ giving

$$\sum_{g \in \mathscr{G}_b} p_g^{\mathrm{G}} - \sum_{l \in \mathscr{L}: F_l = b} p_{bb'} - \sum_{l \in \mathscr{L}: T_l = b} p_{bb'} - g_b^{\mathrm{B}} v_b^2, = \sum_{d \in \mathscr{D}_b} p_d^{\mathrm{D}}, \tag{9.2a}$$

$$\sum_{g \in \mathscr{G}_b} q_g^{\mathrm{G}} - \sum_{l \in \mathscr{L}: F_l = b} q_{bb'} - \sum_{l \in \mathscr{L}: T_l = b} q_{bb'} + b_b^{\mathrm{B}} v_b^2, = \sum_{d \in \mathscr{D}_b} q_d^{\mathrm{D}}. \tag{9.2b}$$

Here g_b^{B} and b_b^{B} are the shunt conductance and susceptance at bus b, representing the fact that equipment installed at the bus itself may absorb real and reactive power. Further p_g^{G} and q_g^{G} are the real and reactive generation at generator g, while p_d^{D} and q_d^{D} are the real and reactive power demand of load d.

In operating the network real and reactive demands p_d^{D} and q_d^{D} at each bus are given and the operator decides on real generation p_g^{G} and voltage v_g at all generator buses. The real and reactive power flow through the lines is thus fully determined by the network topology, demand and generation and voltage at the generators.

For later discussion it will be convenient to partition the variables describing the network into controls $(p_{\mathscr{G}}^{\mathrm{G}}, v_{\mathscr{G}})$ and state variables $x = (q_{\mathscr{G}}^{\mathrm{G}}, v_{\mathscr{B}\setminus\mathscr{G}}, \delta, p_{bb'}, q_{bb'})$ and summarize the power flow equations (9.1) and (9.2) with a slight abuse of notation as

$$PF_{\mathscr{N}}(p_{\mathscr{G}}^{\mathrm{G}}, v_{\mathscr{G}}; x) = (p^{\mathrm{D}}, q^{\mathrm{D}})^{\top}, \tag{9.3}$$

where the subscript $\mathscr{N}$ indicates that the form of the constraints depends on the network topology.

Not all levels of generation are feasible. A typical network will impose operational constraints on the state variables. Voltages are typically required to be within a specified range of the nominal voltage

$$V^{\min} \le v_b \le V^{\max}, \qquad \forall b \in \mathscr{B}, \tag{9.4}$$

where $V^{\min} = 0.94\,\mathrm{p.u.}, V^{\min} = 1.06\,\mathrm{p.u.}$ are commonly used, while heating limits impose a bound on line flows

$$p_{bb'}^2 + q_{bb'}^2 \leq (S_l^{\max})^2, \qquad \forall l = (b, b') \in \mathscr{L}, \tag{9.5}$$

further generator reactive power output must be within the operating limits

$$Q_g^{\mathrm{G,min}} \leq q_g^{\mathrm{G}} \leq Q_g^{\mathrm{G,max}}, \qquad \forall g \in \mathscr{G}. \tag{9.6}$$

Again it will be convenient to express the state constraints (9.4)–(9.6) short as

$$x \in \mathscr{H}.$$

Since the power flow equations are nonlinear, the state of the power system is not necessarily unique and there may be several feasible solutions to (9.3). Indeed, so called "spurious" solutions to the power flow equations are well documented [17]. It is unclear what their prominence in real-life power systems is.

9.3.1 The DC Model

Under normal system operating conditions it is reasonable to make the simplifying assumptions that the reactance of each line is much larger than its resistance ($b_l \gg g_l \approx 0$), all voltages are close to the nominal system voltage 1 p.u. and the phase angle difference θ_l across all lines is small. The standard "DC" approximation to AC power flow linearizes the power flow equations around this operating point: if we substitute $v_b = 1$ and $g_l = 0$ into the power injection equation, we obtain

$$p_{bb'} = -b_l \sin\theta_l, \qquad q_{bb'} = -b_l + b_l \cos\theta_l.$$

If we in addition use the small-angle approximations

$$\cos\theta \approx 1, \qquad \sin\theta \approx \theta,$$

we get

$$p_{bb'} = -b_l \theta_l, \qquad q_{bb'} = 0, \tag{9.7}$$

thus eliminating all reactive power flows and voltage magnitudes from the model. If we further neglect the shunt term in (9.2a), it becomes

$$\sum_{g \in \mathscr{G}_b} p_g^{\mathrm{G}} = \sum_{d \in \mathscr{D}_b} p_d^{\mathrm{D}} + \sum_{l \in \mathscr{L}: F_l = bb'} p_l^L + \sum_{l \in \mathscr{L}: T_l = bb'} p_l^L. \tag{9.8}$$

The resulting model which only uses the real power flow balance (9.8) and the approximation (9.7) to the power injection equation is known as the *DC model*. Its advantage is that it is a linear model. It is a good approximation under normal operating conditions but has a number of disadvantages, principally that the model does not include either reactive power or bus voltages and, because there are no resistances, it does not model line losses.

9.3.2 Contingency Response

We are interested in the response of the network to an unforeseen contingency, such as equipment failure. In the case of a line or bus failure the power flows through the network will rearrange themselves instantaneously according to the power flow equations in the remaining network. In this short-term response it is usually assumed that the real and reactive loads imposed on the network as well as real power generation at the generating buses stay constant. Generators are equipped with voltage regulators that will automatically adjust the generators reactive power output (within the generators design limits) in order to maintain constant voltage. Generator buses are thus assumed to have constant real power and voltage post-contingency (PV-buses), while all other buses are assumed to have constant real and reactive load (PQ-buses).

Naturally, a contingency may isolate a load or generator from the remainder of the network, or it may directly impact a generator bus, leading to over- or undersupply of power in the network (and thus to an infeasible post-contingency load-flow problem). In a real network such a situation is handled through automatic frequency-driven generation and load shedding response: an AC power system has a system frequency (50 or 60 Hz) which is the same throughout the system. In a generation-starved system, the system frequency will drop (since some of the energy stored as rotational energy will be converted to electrical energy). This drop in system frequency is automatically detected by regulators who will attempt to increase the generator outputs. This can be successfully done for small disturbances (such as changes in load arising in standard network operation); however, it may be insufficient to deal with large disturbances such as the loss of a generator due to a contingency. In that case the next automated response would be a pre-determined load shedding program that starts to intentionally disconnect loads from the network driven solely by a severe drop in system frequency. Power systems are usually operating in an $(n-1)$ secure state, that is after the failure of any one line or bus the system is guaranteed to still be in a safe operating state.

All these responses are automatic and do not involve active intervention by the system operator. It is important to note that each response is triggered by information available in the locality without awareness of the full state of the system. In the type of major system disturbances under discussion in this chapter such full knowledge of the system state may not be available.

9.3.2.1 Dynamic System Stability

The solution given by the power flow equations describes the *steady state* of the system, that is the state that the system will settle to once any oscillations have died down due to damping. The discussion of the system response to a contingency in the previous section has assumed that the system operating point will move instantaneously from the old steady state to the new one. In reality power systems show dynamic behaviour: a power system includes a significant amount of mechanical

inertia (e.g. rotors in generators); thus as a result of a contingency the system state will start oscillating about the new steady-state point. While there is usually enough damping for those oscillations to die down rapidly, occasionally they may be violent enough to trigger protection mechanisms that switch off some of the equipment. Thus although there is a feasible steady state in the contingency the system may collapse in the attempt to reach it.

The dynamic behaviour is governed by differential equations, so is difficult to include in an optimization model. Instead surrogates such as slow-coherency analysis have been suggested. We come back to this point in Sect. 9.10.

9.4 Causes of Recent Blackouts

A review of recent power system blackouts and a discussion of their possible causes can be found in [8]. We give here a brief summary of the sequence of events that has led to two of the major blackouts in the last decade.

The August 2003 blackout in the Northeastern USA and Canada started as a localized fault in a small region of northern Ohio, but spread quickly to blackout the entire American Northeast. The disturbance started shortly after 3 pm on August 14 with the tripping of three local transmission lines due to sagging into trees. All three lines were well within their rated capacity (48%, 88% and 93%, respectively). The line trips were mainly due to the environmental conditions (hot, still day), thus reducing ambient cooling of the line. The trips were largely independent, although the failure of the first line did increase the loading and hence the sagging of the other two lines. After the first trip the system was outside its $(n-1)$ operating limits; therefore, intervention by the system operator would have been necessary. However due to a software fault, the control centre was unaware of the line faults. As a result of these faults between 3.39 pm and 4.05 pm further 16 local (138 kV) lines were disconnected automatically due to exceeding their safe operating load. At this point local load shedding would likely have still been able to avert further consequences. However at 4.05 pm a crucial regional (345 kV) line tripped due to being overloaded. At this point the cascade was unavoidable. As the power flows tried to circumvent the downed lines to keep supplying northern Ohio, more and more lines were first overloaded and then automatically tripped, escalating the problem. By 4.13 pm, within 8 min of the start of the cascade the fault had spread to the entire US Northeast and Canada, and 50 million customers were without power. Figure 9.1 shows the progress of the blackout over those 8 min. While the initial cause of the fault was inadequate tree maintenance, the reason for the cascade was lack of situational awareness on behalf of the control centre.

A second example is the Italian blackout [13, 27] in the early morning of 28 September 2003 that isolated Italy from the rest of the European network and, due to severe generation shortage in Italy, resulted in power loss in the majority of Italy. Fifty-six million people were left without power. At the time of the blackout Italy was importing 6.7 GW (or 25% of its total load) from other European countries.

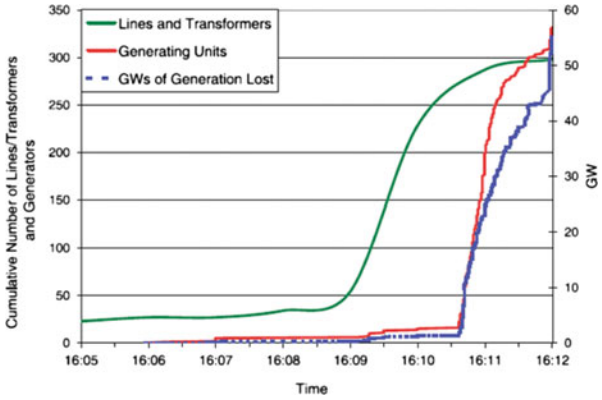

Fig. 9.1 Progress of the 2003 US blackout: cumulative number of lines and generators lost over time (plot taken from [31])

This was 300 MW over the agreed limit. As a result several lines in Switzerland were close to their operational limit. At 3.11 am a major transmission line from Switzerland to Italy failed due to tree contact. The system was then operating outside the planned contingency state and required manual intervention. While some load was reduced, again due to lack of situational awareness, too little was done too late and by 3.25 two more lines from Switzerland to Italy tripped. The resulting instability of the network caused all remaining transalpine connections to trip within seconds, isolating Italy from the rest of Europe. Due to the severe generation shortage in Italy, frequency dropped and automatic load shedding started. However, the frequency drop was so severe that automatic protection mechanism of several generators took them off the grid, escalating the problem. Three minutes after the separation of Italy all power generators in Italy were switched off.

In both cases the power system has been driven by external events to a state outside its normal operating regime in which operator intervention would have to be taken. However, due to insufficient or wrong assumptions about the state of the network on behalf of the operators, they were unable to take the appropriate action to prevent the cascading of the network fault.

The occurrence of these situations is our motivation to use optimization to find a robust response that can be used by the operator to return the system to a safe operating state, even if there is uncertainty about the current state of the network.

9.5 Stochastic Network Defence Model

A common observation of recent wide-area blackouts has been the lack of complete information about the actual state of the system. Commonly the fault was more widespread that initially assumed with more lines and or generators having

been lost. Although operators did take deliberate or automatic action to prevent a spreading of the fault, these actions were based on incomplete knowledge of the system state and hence failed in resolving the problem.

In fact the problem faced by the system operator in such a situation is one of optimization under uncertainty: one would like to take the best corrective action given uncertainties about the actual state of the system. Typical actions available to the operator are the (partial) disconnection of loads, the adjustment of generator voltages and real power outputs (within short-term ramping limits) and the rearrangement of the network topology—in the extreme case isolating the fault from the remainder of the network. While the first two can (and are) employed in a largely automatic fashion on detection of a fault, the rearrangement of the network topology is a more drastic action whose effects are more difficult to predict in a simple manner. We will therefore concentrate on this action.

Our conceptual setup is as follows: the system operator is faced with a disturbance that affects a certain region of the network. While some equipment (lines, buses, generators) are known to work or not to work, for others there may be an uncertainty about their current state (given by a probability). The operator has a number of possible remedial actions that he can take; once the action has been taken, the system should be stable, independent of which of the uncertain states the network is actually in. This situation can be modelled in a stochastic programming setup: *First-stage decisions* (the active decisions taken by the system operator in response to the system emergency) are adjustments to the real power generation Δp_g^{G}, load shedding for individual demands $\alpha_d \in [0,1]$ and line disconnection decisions y_l. We assume that loads can be shed individually and that real and reactive load of a given demand d will be scaled uniformly by the factor α_d.

The *scenarios* are the possible states of the network topology. Let $\mathscr{B}^0 \subseteq \mathscr{B}$ and $\mathscr{L}^0 \subseteq \mathscr{L}$ be the set of buses and lines, respectively, whose state is uncertain. Let $\mathscr{E}^0 = \mathscr{B}^0 \cup \mathscr{L}^0$ and for every piece of equipment $e \in \mathscr{E}^0$. Then the set of possible network topologies $\mathscr{N}_s$ is given by the power set $\mathbb{P}(\mathscr{E}^0)$ of $\mathscr{E}^0$. We denote by $\mathscr{S}$ the scenario set and for every $s \in \mathscr{S}$ let $\mathscr{N}_s$ be the network topology corresponding to scenario s where scenario probabilities π_s could be estimated from historical data of equipment failures. In this setup the number of scenarios would be exponential in the number of uncertain pieces of equipment. The *recourse problem* would ensure that the system has a stable operating state x_s for every contingency scenario s, i.e. there is a feasible point for

$$PF_{\mathscr{N}_s}(p^{\mathrm{G}}, v_{\mathscr{G}}, x_s) = (p^{\mathrm{D}}, q^{\mathrm{D}})^{\top}, \qquad x_s \in \mathscr{H}.$$

The recourse problem would further model the automatic system response to the contingency. Since we assume that the system operator is never aware of the actual state of the system, no active decisions are possible. Rather we assume that automatic generation adjustment and load shedding are taking place in the following manner.

For every generator there is a frequency response curve $\overline{\alpha}_g(\sigma)$ that, for a given *contingency severeness* σ, would adjust the generator output Δp_g^{G} within the

generators short-term ramping limits $|\Delta p_g^{\mathrm{G}}| \le r_g$. In the real system σ would be measured by the system frequency; the standard AC power flow model presented above, however, does not model system frequency. Nevertheless automated network response driven by system frequency can be modelled by an automatic response driven by real power load–generation mismatch. Note that σ is a single system wide parameter, modelling a uniform system wide response. This reflects the fact that any response is automatic with no opportunity for individual generator decisions by the operator. Once all generators are at the short-term adjustment limit any further system response will be by automated load shedding. This again is modelled by a prescribed load shedding curve $\overline{\alpha}_d(\sigma)$ for every demand d which determines the proportion of the demand $(p_d^{\mathrm{D}}, q_d^{\mathrm{D}})$ that is to be shed. There is thus a unique σ for each contingency case that will ensure real power load–generation balance in the contingent system. In our conceptual model this σ will be set by the recourse problem. Note that in the discussion so far we have assume that since the system frequency is used as a feedback mechanism, there is a single parameter σ to control the system response. This is correct as long as the post-contingency system remains connected. If this is not the case there will be a separate frequency (and hence the need for a separate σ) in each connected component. In order not to overburden the notation we use a single σ in the presented model.

The system operator would thus be faced with the following stochastic programming problem (variables are $\Delta p^{\mathrm{G}}, v_{\mathscr{G}}, \alpha, y, \hat{p}^{\mathrm{D}}_{\mathscr{D}}, \hat{q}^{\mathrm{D}}_{\mathscr{D}}$):

$$\begin{aligned} \max_{\Delta p^{\mathrm{G}}, v_{\mathscr{G}}, \alpha, y} \ & \sum_{s\in\mathscr{S}} \pi_s Q_s(\mathscr{N}(y), v_{\mathscr{G}}, \hat{p}^{\mathrm{G}}_{\mathscr{G}}, \hat{p}^{\mathrm{D}}_{\mathscr{D}}, \hat{q}^{\mathrm{D}}_{\mathscr{D}}) \\ \text{s.t.} \ & |\Delta p_g^{\mathrm{G}}| \le r_g, && \forall g \in \mathscr{G} \\ & (\hat{p}_d^{\mathrm{D}}, \hat{q}_d^{\mathrm{D}}) = \alpha_d(p_d^{\mathrm{D}}, q_d^{\mathrm{D}}), && \forall d \in \mathscr{D} \\ & \hat{p}^{\mathrm{G}}_{\mathscr{G}} = p^{\mathrm{G}}_{\mathscr{G}} + \Delta p^{\mathrm{G}}_{\mathscr{G}} \end{aligned} \tag{9.9}$$

where $Q_s(\hat{\mathscr{N}}, v_{\mathscr{G}}, \hat{p}^{\mathrm{G}}_{\mathscr{G}}, \hat{p}^{\mathrm{D}}_{\mathscr{D}}, \hat{q}^{\mathrm{D}}_{\mathscr{D}})$ is the amount of load that can be supplied after possible automatic load shedding in scenario s, given that as a result of the first-stage decisions the network topology is $\hat{\mathscr{N}}$ and the *intended loads* are given by $(\hat{p}^{\mathrm{D}}_{\mathscr{D}}, \hat{q}^{\mathrm{D}}_{\mathscr{D}})$:

$$\begin{aligned} Q_s(\hat{\mathscr{N}}, v_{\mathscr{G}}, \hat{p}^{\mathrm{G}}_{\mathscr{G}}, \hat{p}^{\mathrm{D}}_{\mathscr{D}}, \hat{q}^{\mathrm{D}}_{\mathscr{D}}) = \max_{\sigma_s, x_s, v_{\mathscr{G},s}} \ & \sum_{d\in\mathscr{D}} \overline{p}^{\mathrm{D}}_{d,s} \\ \text{s.t.} \ & PF_{\hat{\mathscr{N}}\cap\mathscr{N}_s}(\overline{p}^{\mathrm{G}}_s, v_{\mathscr{G},s}, x_s) = (\overline{p}^{\mathrm{D}}_{\mathscr{D},s}, \overline{q}^{\mathrm{D}}_{\mathscr{D},s})^\top \\ & \overline{p}^{\mathrm{G}}_{g,s} = \overline{\alpha}_g(\sigma_s)\hat{p}^{\mathrm{G}}_g, \quad \forall g \in \mathscr{G} \\ & (\overline{p}^{\mathrm{D}}_{d,s}, \overline{q}^{\mathrm{D}}_{d,s}) = \overline{\alpha}_d(\sigma_s)(\hat{p}^{\mathrm{D}}_d, \hat{q}^{\mathrm{D}}_d), \quad \forall d \in \mathscr{D} \\ & x_s \in \mathscr{H}. \end{aligned} \tag{9.10}$$

Note that there are no degrees of freedom in this formulation: per the discussion above, the recourse problem does not model any active intervention of the system operator but solely the automatic response which is driven by the single parameter σ. We use the convention that the maximum over an empty set evaluates to $-\infty$ to capture the event that there is no feasible operating state (for any amount of automatic load shed) in any one of the scenarios.

The main objections to this model are its large size if all possible network situations should be considered as individual scenarios and the impossibility to obtain much of the necessary input data (e.g. equipment outage probabilities π_e).

It has been observed (e.g., in [36]) that a wide-area blackout could have been prevented by intentionally splitting the system into islands in such a way that the fault is contained in one of the islands and the remaining islands are viable [36]. We use this observation to derive a tractable approximation to the stochastic network defence problem presented above: Rather than leaving it to the optimization model to find a network topology that will survive in all considered disturbance scenarios, we make the assumption that such a topology will have to isolate the fault, by creating firebreaks that limit the spread of the fault into the remainder of the network. In other words we seek to break the network into electrically isolated islands, one of which will contain the faults. By this assumption we avoid having to explicitly model every possible combination of affected equipment, such are able to reduce the model complexity significantly.

9.6 A Case Study

In order to investigate the spread of cascading blackouts and the effects of possible interventions we have set up the following experiments: at the heart is a simulator that simulates the spread of a cascading blackout. It is designed to simulate the effects of normal control actions (such as frequency-driven load shed and generation adjustment), but not active intervention by system operators. Following the failure of some of the operating equipment (lines, buses, generators), the simulation performs the following steps:

1. Evaluate the load–generation mismatch in each connected component of the network as a proxy for system frequency and uniformly shed load and (within the ramping constraints) adjust generator output to achieve load–generation balance.
2. Perform a load-flow calculation to identify overloaded lines and voltage problems.
   (a) Any line that is overloaded at more than 115% of its emergency rating, is deemed to trip at this iteration and is removed from the network.
   (b) If no voltage feasible solution to a connected component exists, the whole connected component is deemed to have failed due to voltage collapse.
3. The whole analysis is repeated until no more lines are tripped or the complete network has collapsed.

We have investigated the following scenario: in the 118-bus system (see Fig. 9.2 for a part of this network) we start from an $(n-1)$ secure operating point, that is, the system is designed to survive any one line failure. We now consider a scenario where a disturbance event impacts the neighbourhood of bus 61. In fact lines 59–61, 60–61 and 61–62 are all down, but we are also unsure about the state of bus 60 and the line

61–64. We assume that the system operator only knows that the state of these four lines and the bus is uncertain, but not which ones are operating. Pre-disturbance the situation of the network is such that there is a large load of 277 MW at bus 59 and some smaller loads of 63, 84 and 78 MW at buses 55, 56 and 60, respectively. The generators at buses 55 and 56 are not operating; instead much of the load is supplied from the generators at buses 59 and 61 which are supplying 123.4 and 121.3 MW, respectively.

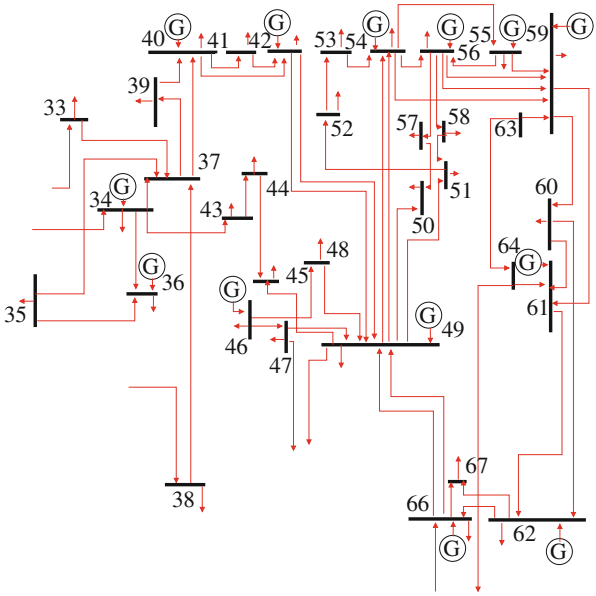

Fig. 9.2 A part of the 118 bus test network

According to the simulation the following sequence of events unfolds: The power generated at bus 61 is now routed via bus 64. The line from 61 to 64 can cope with this increased flow, but the lines 63–64 and 63–59 are overloaded and trip in the first stage of the simulation. Power is now routed through bus 56 and first the lines 58–51 and then 54–56 and 56–57 fail. There are still connections that supply bus 59 through bus 54 and now lines supplying bus 54 (namely 49–54 and 52–53) fail. At this point several buses in the remaining network are not able to maintain voltage within operating bounds and the system collapses.

This case is instructive since the sequence of events is similar to that of the North American blackout of August 2003 in that a small localized fault spreads to affect most of the network by sequentially overloading and tripping lines as power flows take increasingly complex routes to supply loads.

The correct operator response in this case is not obvious, partly because of the uncertainty of the state of the lines around bus 60. From the sequence of events that unfolds it would seem advisable to shed all or part of the load at bus 59.

In fact the problem is caused by the fact that the generator at bus 61 is still operational and connected to the network through line 61–64. While this line itself has enough capacity, there is insufficient transmission capacity away from bus 64, leading to the system collapse by successive overloading of lines. The optimal operator response is to isolate the problem area by islanding buses 60 and 61. The generation lost from bus 61 can be compensated by increasing generation at other buses throughout the network without overloading lines. Islanding of the two buses is thus a safe solution for all possible combinations of failures at buses 60 and 61 and the lines adjacent to them.

9.7 MINLP Islanding Formulation

9.7.1 Motivation

The analysis above suggests a robust approach in which the optimal solution should physically separate the faulty area from the remainder of the network. We thus limit the possible impact of the uncertainty about the network state.

The model presented here aims to physically separate all lines and buses into two sections: a "healthy" Section 1 that only includes equipment that is certain to work and an "unhealthy" Section 0 that includes all doubtful equipment. This is done subject to the constraint that the remaining network in the healthy section must have an AC feasible steady state, i.e. feasible power flows that respect the power flow equations and operational constraints must exist. In particular this means that the islands have to be load–generation balanced. As in the stochastic network defence model, it may be necessary to shed load or adjust generator output to achieve this target. The objective of the model is to minimize the necessary load shed.

Figure 9.3a illustrates the situation: uncertain lines and buses are indicated by a "?". Figure 9.3b shows a possible islanding solution for this network: all uncertain buses have been placed in Section 0 and all lines connecting Sections 0 and 1 as well as all uncertain lines with at least one end in Section 1 are disconnected. Thus the fault is electrically isolated from the remainder of the network. Note that neither section is required to be connected so may contain more than one island: in Fig. 9.3b, Section 1 comprises islands 1, 3 and 4, and Section 0 is a single island.

Loads may be placed in the unhealthy Section 0. If so, there is a certain probability that the load cannot be supplied, namely if the unhealthy section collapses. The optimization will minimize the amount of load placed in Section 0 as well as the planned load shed.

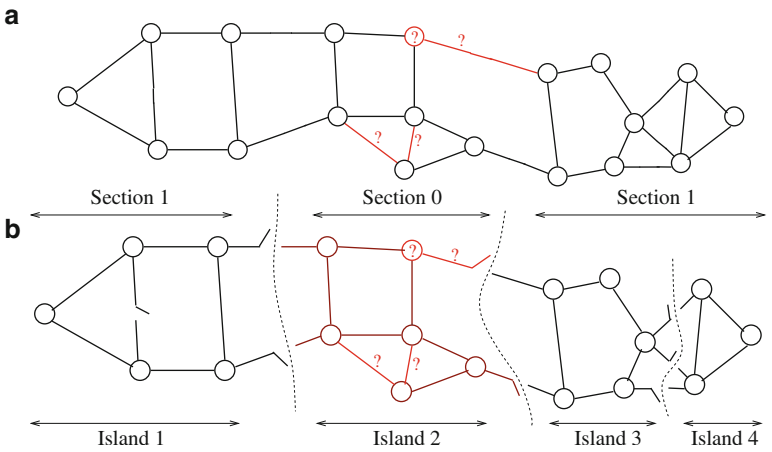

Fig. 9.3 (**a**) *Above*: illustration of a network with uncertain buses and lines. (**b**) *Below*: the islanding of that network by disconnecting lines

9.7.2 The Model

9.7.2.1 Sectioning Constraints

As before let $\mathscr{B}^0 \subseteq \mathscr{B}$ and $\mathscr{L}^0 \subseteq \mathscr{L}$ be the subset of buses and lines that are suspected to be faulty or at risk. To place buses into sections a binary variable γ_b is defined for each bus $b \in \mathscr{B}$; γ_b is set equal to 0 if b is placed in Section 0 and $\gamma_b = 1$ otherwise:

$$\gamma_b = 0, \ \forall b \in \mathscr{B}^0. \tag{9.11a}$$

A binary variable ρ_l is defined for each $l \in \mathscr{L}$; $\rho_l = 0$ if line l is disconnected and $\rho_l = 1$ otherwise. Lines are not placed into sections: rather any line that connects buses in different sections and uncertain lines whose two buses are in the healthy section are required to be disconnected. In our model this is enforced by constraints (9.11b) and (9.11c) below. The first set of constraints (9.11b) applies to lines in $\mathscr{L} \setminus \mathscr{L}^0$ and expresses that a line has to be cut if its two end buses are in different sections (i.e. $|\gamma_{F_l} = |\gamma_{Fl} - \gamma_{T1}| = 1$):

$$\rho_l \leq 1 + \gamma_{F_l} - \gamma_{T_l}, \quad \rho_l \leq 1 - \gamma_{F_l} + \gamma_{T_l}, \qquad \forall l \in \mathscr{L} \setminus \mathscr{L}^0. \tag{9.11b}$$

Constraints (9.11c) apply to lines assigned to $\mathscr{L}^0$ and disconnect any *at-risk* line if at least one of the ends is in Sect. 9.1. Thus, an uncertain line either (a) shall be disconnected if entirely in Section 1, (b) shall be disconnected if between Sections 0 and 1 or (c) may remain connected if entirely in Section 0:

$$\rho_l \leq 1 - \gamma_{F_l}, \quad \rho_l \leq 1 - \gamma_{T_l}, \qquad \forall l \in \mathscr{L}^0. \tag{9.11c}$$

Given some assignments to $\mathscr{B}^0$ and $\mathscr{L}^0$, the optimization will disconnect lines and place buses, hence partitioning the network into Sections 0 and 1. What else is placed in Section 0, what other lines are cut and which loads and generators are adjusted are degrees of freedom for the optimization and will depend on the objective function.

9.7.2.2 The Power Flow Equations

Power transmission within each island has to satisfy the power flow equations. The power balance equation can be imposed in its original form (9.2), while the power injection equation needs to be modified so that for a cut line:

- The real and reactive flow injected into the line at both ends is zero,
- Bus voltages and phase angles at either end are free to move independently.

This is achieved by using the following modification of the power injection equation:

$$p_{bb'} = g_l \hat{v}_b^2 - \hat{v}_{Ib}\hat{v}_{Ib'}\left(g_l \cos\theta_l + b_l \sin\theta_l\right), \tag{9.12a}$$

$$q_{bb'} = -b_l \hat{v}_b^2 - \hat{v}_{lb}\hat{v}_{lb'}\left(g_l \sin\theta_l - b_l \cos\theta_l\right). \tag{9.12b}$$

Here $\hat{v}_{Ib}$ are auxiliary variables that are required to equal the bus voltages when the line is connected and are forced to $V^{\min}$ otherwise:[1]

$$-(V^{\max} - V^{\min})(1-\rho_l) \le \hat{v}_{Ib} - v_{Ib} \le (V^{\max} - V^{\min})(1-\rho_l), \tag{9.13a}$$

$$-(V^{\max} - V^{\min})\rho_l \le \hat{v}_{lb} - V^{\min} \le (V^{\max} - V^{\min})\rho_l. \tag{9.13b}$$

In addition

$$-\Theta\rho_l \le \theta_l \le \Theta\rho_l, \tag{9.14a}$$

$$-M(1-\rho_l) \le \theta_l - \delta_{F_l} + \delta_{T_l} \le M(1-\rho_l). \tag{9.14b}$$

Here, Θ is the maximal permissible angle difference across a connected line (usually $\Theta = \pi/2$), while M is a "big-M" constant. If line l is connected, (9.14b) will impose θ_l to be equal to the angle difference across a line, while for a disconnected line (9.14a) will force $\theta_l = 0$ without imposing equality of the corresponding bus phase angles. For a disconnected line l, $\theta_l = 0$ and $\hat{v}_{Ib} = \hat{v}_{Ib'} = V^{\min}$ and thus by (9.12) we have $p_{bb'} = q_{bb'} = 0$.

[1] The value of $V^{\min}$ is somewhat arbitrary, and any other value (notably $\hat{v} = 1$) could be used for the auxiliary variables $\hat{v}_{Ib}$ for disconnected lines. It is important that both ends of the disconnected line are set to the same value, so that the expressions in (9.12) cancel for $\theta_l = 0$. Using $V^{\min}$ allows (9.13a) to be modelled with a single set of inequalities each.

9.7.2.3 Generation Constraints

In the short time available when islanding in response to a contingency it is not possible to start up generators. Generators that are operating can have either their input power disconnected, in which case their real output power drops to zero, or their output can be changed to a value within a small interval, $\left[P_g^{\mathrm{G,min}}, P_g^{\mathrm{G,max}}\right]$ say for generator g, around their pre-islanded value. The limits will depend on the ramp and output limits of the generator, and the amount of immediate or short-term reserve capacity available to the generator. This alternative operating regime is modelled by the constraint

$$\zeta_g P_g^{\mathrm{G,min}} \leq p_g^{\mathrm{G}} \leq \zeta_g P_g^{\mathrm{G,max}}, \tag{9.15}$$

where ζ_g is a binary variable. If $\zeta_g = 0$, then generator g is switched off and $p_g^{\mathrm{G}} = 0$; otherwise $\zeta_g = 1$ and the generator output is $p_g^{\mathrm{G}} \in \left[P_g^{\mathrm{G,min}}, P_g^{\mathrm{G,max}}\right]$.

9.7.2.4 Load Shedding

Because of the limits on generator power outputs and network constraints it may not be possible to fully supply all loads after islanding. It is therefore necessary to permit some shedding of loads. Note that this is *intentional* load shedding, not automatic shedding as a result of low frequency. As in Sect. 9.5 we assume that load can be shed individually at each bus, in which case real and reactive demand at that bus are scaled by the same factor α_d:

$$\hat{p}_d^{\mathrm{D}} = \alpha_d p_d^{\mathrm{D}}, \qquad \hat{q}_d^{\mathrm{D}} = \alpha_d q_d^{\mathrm{D}}. \tag{9.16}$$

9.7.2.5 Objective Function

The overall goal in islanding is to split the network and leave it in a secure steady state while maximizing the expected value of the load supplied. However, if a certain load is placed in Section 0, then because this section is vulnerable, it is assumed there is a risk of not being able to supply power to that load. Accordingly, a load loss penalty $0 \leq \beta < 1$ is defined, which may be interpreted as the probability of being able to supply a load if placed in Section 0. With this interpretation the expected value of the load supplied is J:

$$J = \sum_{d \in \mathscr{D}} p_d^D \left(\beta \alpha_{0d} + \alpha_{1d}\right),$$

where

$$\alpha_d = \alpha_{0d} + \alpha_{1d}, \quad \forall d \in \mathscr{D}, \tag{9.17a}$$

$$0 \leq \alpha_{0d} \leq 1, \quad \forall d \in \mathscr{D}, \tag{9.17b}$$

$$0 \leq \alpha_{1d} \leq \gamma_b, \quad \forall b \in \mathscr{B}, d \in \mathscr{D}_b. \tag{9.17c}$$

Here a new variable α_{sd} is introduced for the load d delivered in section $s \in \{0,1\}$. If $\gamma_b = 0$ (and so the load at bus b is in Section 0), then $\alpha_{1d} = 0, \alpha_{0d} = \alpha_d$ and the expected load delivered is $\beta p_d^D \alpha_d$. On the other hand, if $\gamma_b = 1$, then $\alpha_{1d} = \alpha_d$ and $\alpha_{0d} = 0$, giving a larger expected load delivered of $p_d^D \alpha_d$. Thus maximizing J gives a preference for $\gamma_b = 1$ and a smaller Section 0.

9.7.2.6 The Complete Model

The complete MINLP islanding model is thus to maximize the expected load supplied

$$\max J = \sum_{d \in \mathscr{D}} p_d^{\mathrm{D}} (\beta \alpha_{0d} + \alpha_{1d}) \tag{9.18}$$

subject to sectioning constraints (9.11), modified power flow equations (9.2), (9.12) together with the line cutting constraint (9.13) and (9.14), load shedding (9.16), (9.17) and operating limits on line heating (9.5) and generator real (9.15) and reactive (9.6) output.

The presented conceptual islanding model is a MINLP problem. In addition the nonlinear constraints are nonconvex and even for a fixed setting of binary variables there is a possibility of local solutions to the resulting optimal power flow problem. Such a model is very difficult to solve. Indeed attempts to solve it with a general MINLP solver such as Couenne [6] exceed the time and memory limits even for small problems with 9 and 14 buses.

If the (debatable) assumption is made that the continuous NLP subproblems that result from relaxing the integrality requirements can be solved to global optimality by a local NLP solver, then a convex MINLP solver such as Bonmin [11] can be used. However, Bonmin still requires an excessive amount of computational time, frequently more than 10 h for a problem with 14 buses.

9.8 MILP Islanding Formulation

To overcome the computational complexity of the conceptual MINLP formulation of the previous chapter we present two simplifications that were developed in [29, 30] to provide tractable approximations to the MINLP formulation presented in the previous section.

9.8.1 DC Power Flow Model with Line Losses

An appealing idea is to use the DC OPF model instead of the AC OPF model within the islanding model (as is done in [12]), since this would lead to a linear problem. The derivation of the DC model assumes that the system is close to normal

operation conditions, something that is unlikely to be the case in the situation we are interested in. Two shortcomings of the DC model in particular are relevant to our application: the neglectance of line losses and reactive power. If line losses are neglected the model may choose islands that appear to have sufficient power generation, but in fact have not. This effect can be somewhat mitigated by including an approximation of line losses in a DC-based islanding model.

Such an islanding model is suggested in [29]. It employs a power flow model that is a variant of the "DC" model with the only difference from the standard DC load-flow model being that line losses are approximated by constant terms. Indeed the power balance equation (9.8) is replaced by

$$\sum_{g\in\mathscr{G}_b} p_g^{\mathrm{G}} = \sum_{d\in\mathscr{D}_b} p_d^{\mathrm{D}} + \sum_{l\in\mathscr{L}:F_l=b} p_l^{\mathrm{L}} - \sum_{l\in\mathscr{L}:T_l=b} (p_l^{\mathrm{L}} - \bar{h}_l^{\mathrm{L}}), \tag{9.19}$$

where $\bar{h}_l^{\mathrm{L}}$ is a (constant) estimate of the real power line loss along the line. We can use

$$\bar{h}_l^{\mathrm{L}} = \rho_l h_l^{\mathrm{L0}}, \tag{9.20}$$

where h_l^{L0} is the loss immediately before islanding. The inclusion of ρ_l ensures the modelled loss is zero if the line is cut. The rest of the model is the same as the MINLP islanding model of Sect. 9.7, resulting in a mixed-integer linear (MILP) formulation.

The issues related to reactive power however are more subtle: lack or surplus of reactive power may result in the system not being able to support bus voltages within their operating limits. Note that reactive power needs to be available locally, so it is not sufficient to ensure that each island has sufficient reactive power generation to supply all loads. Indeed Trodden et al. [29] report on an example where a shunt reactor used to compensate reactive power on an underground cable depresses voltages after that cable has been cut. The "DC" model is blind to such issues.

9.8.2 Piecewise Linear AC Power Flow Approximation

Motivated by this, in [30], this model is extended to include a PWL approximation to AC power flow, in which voltage and reactive power are modelled. This is done by using a PWL model of the cosine term in the power injection equation that would be the dominant source of error in a standard linear formulation. Depending on the number of pieces used in the PWL approximation, this model of the cosine can be of arbitrary precision.

Its starting point is to linearize the power flow equations around unit voltages and zero-phase angle differences ($v = 1$ and $\theta = 0$) to obtain

$$\sum_{g\in\mathrm{G}_b} p_g^{\mathrm{G}} = \sum_{d\in\mathscr{D}_b} P_d^{\mathrm{D}} + \sum_{(b,b')\in\mathscr{L}} p_{bb'} + \sum_{(b',b)\in\mathscr{L}} p_{bb'} + g_b^{\mathrm{B}}(2v_b - 1), \tag{9.21a}$$

$$\sum_{g\in\mathscr{G}_b} q_g^{\mathrm{G}} = \sum_{d\in\mathscr{D}_b} Q_d^{\mathrm{D}} + \sum_{(b,b')\in\mathscr{L}} q_{bb'} + \sum_{(b',b)\in\mathscr{L}} q_{bb'} - b_b^{\mathrm{B}}(2v_b - 1), \tag{9.21b}$$

and

$$p_{bb'} = g_l(2v_b - 1) - g_l\big(v_b + v_{b'} + c_l - 2\big) - b_l s_l, \tag{9.22a}$$
$$q_{bb'} = b_l(1 - 2v_b) + b_l\big(v_b + v_{b'} + c_l - 2\big) - g_l s_l, \tag{9.22b}$$

where $c_l = \cos\theta_l$ and $s_l = \sin\theta_l$. As a next step it would be natural to use the small-angle approximations $c_l \approx 1, s_l \approx \theta_l$, which incur an error of order θ_l^2 and θ_l^3, respectively. Table 9.1 gives the maximum absolute errors for each of the constituent

Table 9.1 Approximation errors in line flow terms

Term	Approximation	Max abs error on $\mathscr{X}$
v_b^2	$2v_b - 1$	0.0025
$v_b v_{b'} c_l$	$v_b + v_{b'} + c_l - 2$	0.0253
$v_b v_{b'} s_l$	s_l	0.0659
c_l	1	0.2340
s_l	θ_l	0.0553

terms in the thus approximated flows, over a typical range of operating voltages and angles (i.e. the domain $\mathscr{X} = \big\{(v_b, v_{b'}, \theta_l) : 0.95 \le v_b, v_{b'} \le 1.05, |\theta_l| \le 40^\circ\big\}$). As can be seen the small-angle cosine approximation incurs the largest error. To overcome this instead of the small-angle approximation an N-piece piecewise linear PWL approximation to the cosine is used:

$$c_l = c_{l,i}\theta_l + d_{l,i}, \forall\theta_l \in [x_{l,i}, x_{l,i+1}], i \in \{0,\ldots,N-1\}, \tag{9.23}$$

where $c_{l,i}$ and $d_{l,i}$ are obtained by evaluating $\cos x$ at breakpoints $\{x_{l,0},\ldots,x_{l,N}\}$. Note that the modelling of (9.23) requires the use of integer variables. This can however be done by using special ordered sets (SOS) of type 2, which some solvers (e.g., CPLEX) can handle directly and efficiently.

9.8.3 Post-Islanding AC Optimal Load Shedding

Both the DC and PWL islanding model use an approximation to Kirchhoff's laws. Thus neither can guarantee that an islanding solution obtained from the models has a feasible steady-state operating point. In order to evaluate the quality of the islands returned by each of the above scheme a standard AC optimal flow problem has been solved for each of the islands generated. Load shedding and adjusting generator real outputs within ramping limits (9.15) are allowed; however, all binary variables are fixed to the solution from the islanding models.

9.9 Computational Results

This section presents computational results using the above islanding formulations. First the construction of the test problems is described, then the quality of the islanding solutions returned by the two MILP-based models is evaluated and finally we comment on the necessary solution times. We only give a summary of results here; a fuller set is available from [29, 30].

9.9.1 Islanding Test Cases

A set of islanding test cases was built based on IEEE systems with between 9 and 300 buses. For a network with n^{B} buses, n^{B} scenarios were generated by assigning in turn each single bus to $\mathscr{B}^0$.

Voltage limits were set to $0.94\,\text{p.u.} \leq v_b \leq 1.06\,\text{p.u.}$ for all buses. The lower and upper bounds on real power output for generator g, $P_g^{\mathrm{G,min}}$ and $P_g^{\mathrm{G,max}}$, were defined as $\pm 5\%$ of the pre-islanding output, $P_g^{\mathrm{G}*}$, obtained by solving an AC OPF on the intact network prior to islanding. Reactive power outputs were allowed to vary over their full range. In the objective function, a value of 0.75 is used for the load loss penalty β. For the PWL approximation for a line $l12$ pieces are used to cover a range of 10° either side of $\pm|\theta_l^*|$, where θ_l^* is the pre-islanding phase difference across the line.

All the test are performed on a 64-bit Linux machine with Intel i7-2600 processor and 8 GiB RAM, using CPLEX 12.4 as the solver.

9.9.2 AC Feasibility and Solution Accuracy

We first investigate the quality of the solutions obtained by the two MILP-based islanding models. Table 9.2 gives the percentage of test cases for various network sizes for which the post-islanding optimal load shedding problem (AC-OLS) could not find a feasible solution, i.e. the solution produced by the islanding model is not operationally feasible. On this measure the islands created by the DC model are infeasible in 10–20% of cases across all network sizes. In all these cases the infeasibilities were caused by bus voltages exceeding their limits: if voltage bounds are relaxed by a further 0.06 p.u., i.e. to $0.88\,\text{p.u.} \leq v_b \leq 1.12\,\text{p.u.}$, then the islanding solution is AC feasible in all cases. This may represent an acceptable solution in an emergency, but such voltages would leave the system in a weakened state which is undesirable.

The PWL-based islanding model on the other hand is able to produce islands that allow AC feasible solutions in all cases. This is due to the fact that, as intended, the PWL model is able to approximate reactive power and voltages. Figure 9.4

compares the voltages at all buses and reactive flow in all lines as predicted by the PWL approximation and the full AC model for the 24-bus test network. It demonstrates a good matching of the predicted with the actual values.

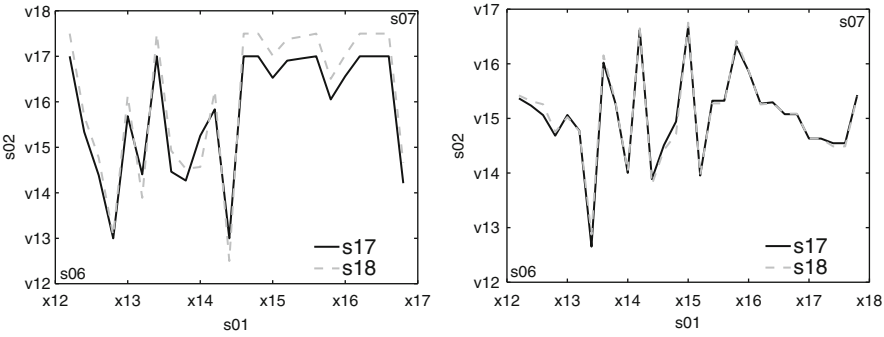

Fig. 9.4 Comparison of AC PWL and AC solutions for 24-bus network

Table 9.2 Percentage of AC-infeasible solutions

n^B	9	14	24	30	39	57	118	300
DC	0	13	21	23	17	10	12	18
PWL-AC	0	0	0	0	0	0	0	0

9.9.3 Computation Times and Optimality

Table 9.3 reports the mean time over all test cases to solve the models to 1% optimality gap. Results are encouraging in that all problems can be solved within 30 s on average with the PWL model taking slightly longer, but not prohibitively so. It can also be seen that islands in some systems are much harder to find than in others, irrespective of network size. However, a few remarks are in order:

- The solution of an MILP problem often makes fast progress in finding a solution and then takes a long time to prove that there are no possible better solutions. In practice, since islanding is a measure of last resort, proven optimality is not a priority, and being able to provide good, feasible islands quickly is more important. We therefore report on times to achieve a solution that is proven to be within 1% of the true optimal. The analysis of the AC feasibility of the islands discussed earlier in Table 9.2 was based on these solutions.
- We report average solution times over all test cases, when in fact the times are highly variable. Indeed the median for all cases is below 1 s. On the other hand there are a small number of cases where 1% optimality gap cannot be reached

within the time limit of 500 s. In all cases feasible islanding solutions are found very fast and the actual MIP gap of the best solutions found after the time limit is reached is always acceptable (see Table 9.4).

Nevertheless we wish to comment on how an islanding scheme based on the presented models could be used in practice. We assume that a decision if and how to island has to be taken very quickly, probably within 5 min. We would suggest to run the islanding model as soon as a fault is suspected. Any integer feasible solution found by the solver (even if not optimal) should be returned immediately

Table 9.3 Mean times (s) to solve models to 1% MIP gaps

n^B	9	14	24	30	39	57	118	300
DC	0.03	0.08	0.13	0.10	0.15	25.4	4.66	0.61
PWL	0.0	0.6	0.5	0.9	0.6	28.8	16.5	14.5

Table 9.4 MIP gaps (%) of solutions returned at 500 s

n^B	39	57	118	300
Number	2	4	4	3
Min	0.34	0.21	0.61	0.11
Mean	0.37	2.37	1.22	0.89
Max	0.41	5.12	1.69	2.37

and checked (on a different processor) for AC feasibility. In this way a record of the best found AC feasible islanding solution would be kept and our experiments suggest that the first implementable solution would be found very quickly. In order to speed the process up further and to cover against cases where the problem is very difficult the islanding model could be run as a matter of course under normal network operation periodically to have access to an islanding plan when the need arises. If typical patterns of uncertain regions can be identified, islanding solutions for these patterns can be computed off-line. Finally, to provide a backup, one of schemes based on graph partitioning could be run in parallel to our model—these would be expected to find a partitioning very fast, however without any guarantee that this partitioning leads to an AC feasible operating point of the islands.

9.10 Extensions: Dynamic Stability

We have concentrated on islanding schemes whose primary focus is to create islands that pose an AC feasible steady-state operating point at a minimum level of expected load lost. This disregards the issue of dynamic stability, namely whether

such an operating point will be reached in practice or whether the shock to the system created by the cutting of lines may cause uncontrollable oscillations.

The optimal islanding problem with objective J often has multiple feasible solutions with objectives close to the optimal value. This flexibility can be exploited by including penalty terms $\varepsilon \sum_{l \in \mathscr{L}} W_l \rho_l$, with appropriately chosen weights W_l, in the objective that for example discourages the disconnection of lines with large pre-islanding power flow. This will minimize the disruption caused to the system by the act of islanding and thus increase the likelihood of dynamic stability.

Indeed as reported in [29] we have tested the islanding solutions found by the DC model in a time-domain simulation and have found that 14 out of a total of 452 test runs led to dynamically unstable solutions. In all cases, however, resolving the model with penalties on cutting high-flow lines resulted in dynamically stable islands. In addition coherent groups of generators can be determined by slow-coherency analysis pre-islanding [5] and the placing of generators in the same coherent group into different islands can be discouraged by penalties in the islanding objective. This approach is described in more detail in [30].

9.11 Conclusions

In this chapter the problem of finding a robust operator response to prevent an imminent cascading blackout when the exact state of some parts of the system is uncertain has been studied.

We have cast the problem within a conceptual stochastic programming setup. However due to the number of possible system states that need to be considered such a model is intractable. Instead we argue that a possible (and quite likely the only) robust operator intervention is to island the system, i.e. to electrically isolate the area affected by the disturbance.

An optimization-based approach for intentional or controlled islanding has been presented. Starting from a (still very difficult) MINLP formulation that takes into account nonlinear network constraints, we have presented two possible mixed-integer linear models based on different approximations of the network model. The approach is flexible with respect to the aims and objectives of islanding and finds islands that are balanced and satisfy real and reactive power flow and operating constraints.

The simpler DC-based approach, while faster to solve, has been observed to occasional lead to islands with out-of-bound voltages. The more refined PWL model has been able to overcome these problems in all cases at the expense of a moderate increase in computation time. Both proposed models find good feasible islanding solutions very quickly in almost all cases. We have suggested how such a scheme can be used in practice.

While the dynamic response is not explicitly modelled in the optimizations, time-domain simulations of the islanding solutions have indicated that instability is avoided by appropriate choices of penalties on cuts to high-flow lines and disconnections of generating units, both of which discourage disruption to the network.

Acknowledgements

This work has been supported by the UK Engineering and Physical Sciences Research Council (EPSRC) under Grant EP/G060169/1. The authors would like to thank Andreas Wank for providing the cascading blackout simulation and Ian Wallace for providing solution times for the MINLP islanding model.

References

1. Agematsu S, Imai S, Tsukiu R, Watanabe H, Nakamura T, Matsushima T (2001) Islanding protection system with active and reactive power balancing control for Tokyo metropolitan power system and actual operational experience. In: Proceedings of the 7th IEE international conference developments in power systems protection 2001. IEEE Computer Society, London, pp 351–354
2. Ahmed SS, Sarker NC, Khairuddin AB, Ghani MRBA, Ahmed H (2003) A scheme for controlled islanding to prevent subsequent blackout. IEEE Trans Power Syst 18(18):136–143
3. Andersson G, Donalek P, Farmer R, Hatziargyriou N, Kamwa I, Kundur P, Martins N, Paserba J, Pourbeik P, Sanchez-Gasca J, Schulz R, Stankovic A, Taylor C, Vittal V (2005) Causes of the 2003 major grid blackouts in North America and Europe, and recommended means to improve system dynamic performance. IEEE Trans Power Syst 20(4):1922–1928
4. Archer BA, Davis JB (2002) System islanding considerations for improving power system restoration at Manitoba Hydro. In: Proceedings of the 2002 Canadian conference on electrical and computer engineering. IEEE Computer Society, London, pp 60–65
5. Avramovic B, Kokotovic PK, Winkelman JR, Chow JH (1980) Area decomposition for electromechanical models of power systems. Automatica 16:637–648
6. Belotti P, Lee J, Liberti L, Margot F, Wächter A (2009) Branching and bounds tightening techniques for non-convex MINLP. Optim Methods Software 24(4–5):597–634
7. Bialek JW (2003) Are blackouts contagious? IEE Power Eng 17(6):10
8. Bialek JW (2004) Recent blackouts in us and continental Europe: is liberalisation to blame? Technical report CWPE 0407, University of Cambridge, Department of Applied Economics
9. Bialek JW (2005) Blackouts in the US/Canada and continental Europe in 2003: is liberalisation to blame? In: IEEE powertech conference, Russia, June 2005
10. Bienstock D, Mattia S (2007) Using mixed-integer programming to solve power grid blackout problems. Discrete Optim 4:115–141
11. Bonami P, Biegler LT, Conn AR, Cornuejols G, Grossmann IE, Laird CD, Lee J, Lodi A, Margot F, Sawaya N, Wächter A (2005) An algorithmic framework for convex mixed integer nonlinear programs. Technical research report RC23771, IBM, Oct 2005

12. Fan N, Izraelevitz D, Pan F, Pardalos PM, Wang J (2012) A mixed integer programming approach for optimal power grid intentional islanding. Energy Syst 3:77–93
13. Final Report of the Investigation Committee on the 28 September 2003 Blackout in Italy (2004) Final report, Union for the Coordination of the Transmission of Electricity (UCTE), April 2004
14. Final Report System Disturbance on 4 November 2006 (2007) Final report, Union for the Coordination of the Transmission of Electricity (UCTE)
15. Fisher EB, O'Neill RP, Ferris MC (2008) Optimal transmission switching. IEEE Trans Power Syst 23(3):1346–1355
16. Hamada IA, Israelsa B, Rikvolda PA, Porosevab SV (2010) Spectral matrix methods for partitioning power grids: applications to the italian and floridian high-voltage networks. Phys Procedia 4:125–129
17. Hiskens IA, Davy RJ (2001) Exploring the power flow solution space boundary. IEEE Trans Power Syst 16:389–395
18. Jin M, Sidhu TS, Sun K (2007) A new system splitting scheme based on the unified stability control framework. IEEE Trans Power Syst 22(1):433–441
19. Larsson S, Danell A (2006) The black-out in southern Sweden and eastern Denmark, September 23, 2003. In: IEEE power systems conference and exposition, Atlanta
20. Li H, Rosenwald GW, Jung J, Liu C (2005) Strategic power infrastructure defence. IEEE Proc 93:918–933
21. Liu Y, Liu Y (2006) Aspects on power systems islanding for preventing widespread blackout. In: Proceedings of the 2006 IEEE international conference on networking, sensing and control, ICNSC'06. IEEE, Ft. Lauderdale, pp 1090–1095
22. Müller N, Quintana VH (1992) A sparse eigenvalue-based approach for partitioning power networks. IEEE Trans Power Syst 7:520–527
23. Newman DE, Carreras BA, Kirchner M, Dobson I (2011) The impact of distributed generation on power transmission grid dynamics. In: 44th Hawaii international conference on system science, Hawaii
24. Peiravi A, Ildarabadi R (2009) A fast algorithm for intentional islanding of power systems using the multilevel kernel k-means approach. J Appl Sci 9:2247–2255
25. Quintana VH, Müller N (1991) Partitioning of power networks and applications to security control. In: IEE Proc Generation, Transmission and Distribution, Part C 138:535–545
26. Rajamani K, Hambarde UK (1999) Islanding and load shedding schemes for captive power plants. IEEE Trans Power Delivery 14:805–809
27. Report on the blackout in Italy on 28 September 2003 (2003) Technical report, Swiss Federal Office of Energy (SFOE), November 2003
28. Sun K, Zheng D-Z, Lu Q (2003) Splitting strategies for islanding operation of large-scale power systems using OBDD-based methods. IEEE Trans Power Syst 18:912–923

29. Trodden PA, Bukhsh WA, Grothey A, McKinnon KIM (2013) MILP formulation for controlled islanding of power networks. Int J Electr Power Energy Syst 45:501–508
30. Trodden PA, Bukhsh WA, Grothey A, McKinnon KIM (2013) Optimization-based islanding of power networks using piecewise linear AC power flow. Technical report ERGO 13-011, Edinburgh Research Group in Optimization, School of Mathematics, University of Edinburgh, April 2013
31. US-Canada Power System Outage Task Force (2004) Final report on the August 14, 2003 blackout in the United States and Canada: causes and recommendations. Final report, April 2004
32. Wang X, Vittal V (2004) System islanding using minimal cutsets with minimum net flow. In: IEEE power systems conference and exposition, New York
33. Wang CG, Zhang BH, Li P, Shu J, Cheng LY, Hao ZG, Bo ZQ, Klimek A (2008) Power systems islanding based on multilevel reduced graph partitioning algorithm. In: Proceedings of the 43rd international universities power engineering conference, vol 25, pp 1–6
34. Wang CG, Zhang BH, Hao ZG, Shu J, Li P, Bo ZQ (2010) A novel real-time searching method for power system splitting boundary. IEEE Trans Power Syst 25(4):1902–1909
35. Xu G, Vittal, V (2010) Slow coherency based cutset determination algorithm for large power systems. IEEE Trans Power Syst 25(2):877–884
36. Yang B, Vittal V, Heydt GT (2006) Slow-coherency-based controlled islanding—a demonstration of the approach on the August 14, 2003 blackout scenario. IEEE Trans Power Syst 21:1840–1847
37. You H, Vittal V, Wang X (2004) Slow coherency-based islanding. IEEE Trans Power Syst 19(1):483–491

Chapter 10
Complementarity and Game-Theoretical Models for Equilibria in Energy Markets: Deterministic and Risk-Averse Formulations

Juan Pablo Luna, Claudia Sagastizábal*, and Mikhail Solodov

Abstract Electricity and natural gas transmission and distribution networks are subject to regulation in price, service quality, and emission limits. The interaction of competing agents in an energy market subject to various regulatory interventions is usually modeled through equilibrium problems that ensure profit maximization for all the agents. These types of models can be written in different manners, for example, by means of mixed complementarity problems, variational inequalities, and game-theoretical formulations. More generally, we consider energy markets both in deterministic and stochastic settings and explore theoretical relations between the various formulations found in the literature and in practice. Our analysis shows that the profit-maximization complementarity model is equivalent to a game with agents minimizing costs if the setting is deterministic or risk neutral. On the other hand, when the agents exhibit risk aversion which is natural in this type of markets, the equivalence no longer holds. This gives rise to an interesting economical interpretation. As a complement to our theoretical study, and for the European natural gas market with deterministic data, we present some numerical results showing the impact of market power on equilibrium prices.

10.1 Introduction

In spite of an undeniable worldwide trend of liberalization, industries dealing with energy networks (and to a lesser extent with water supply) continue to be subject to regulation in price, entry, and service quality of the network. Regarding electricity

*Visiting researcher at IMPA, Brazil. On leave from INRIA, France.

J.P. Luna • C. Sagastizábal (✉) • M. Solodov
IMPA - Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro, RJ 22460-320, Brazil
e-mail: jluna@impa.br; sagastiz@impa.br; solodov@impa.br

R.M. Kovacevic et al. (eds.), *Handbook of Risk Management in Energy Production and Trading*, International Series in Operations Research & Management Science 199,
DOI 10.1007/978-1-4614-9035-7_10,

and natural gas transmission and distribution, the specific mechanism chosen for regulation impacts significantly competition and affects the network prices, investment, and reliability.

In general, good performance of the regulatory framework results in lower operation and transmission costs, better service quality, and investment to expand the network and face future changes in demand and supply. Regulation plays an important role also with respect to environmental concerns, for example, encouraging carbon trading to reduce CO2 emissions.

In a market of energy that is subject to various regulatory interventions it is very important to fully understand the interaction of competing agents. Due to the presence of relatively few companies generating power in a given region, electricity markets are naturally set in an oligopolistic competition framework. A similar situation arises in the natural gas industry.

In a centralized environment the paradigm of cost minimization defines energy prices based on marginal costs or shadow prices obtained by optimization. In a liberalized setting, by contrast, prices are computed through equilibrium models aimed at ensuring profit maximization for all the agents. These types of models can be formulated in different ways, for example, by means of mixed complementarity problems (MCP), bi-level programming, and mathematical programs with equilibrium constraints. We mention [1, 2, 4, 5, 8, 17, 18, 19, 26, 34], without the claim of being exhaustive.

In this work we explore the relations between mixed complementarity, variational inequality, and game-theoretical formulations of energy markets both in deterministic and stochastic settings. Our analysis shows that the profit-maximization complementarity formulation is equivalent to a game with agents minimizing costs if the setting is deterministic or risk neutral. On the other hand, when the agents in the market exhibit risk aversion, which is natural in this type of markets, the equivalence no longer holds. More precisely, the risk-averse game becomes equivalent to a complementarity model where agents maximize the expected remuneration and hedge risk only in the cost.

In the development that follows, we consider a stylized energy market that is general enough to cover the generation capacity expansion model [9] as well as the European natural gas market model in [14]. For the latter market and in a deterministic setting, we also present some numerical results showing the impact of market power on equilibrium prices.

Some comments about our notation and terminology are in order. For x, y in any given space, we denote by $\langle x, y\rangle$ the usual (Euclidean) inner product, and we write $x \perp y$ to say that $\langle x, y\rangle = 0$. By $\mathcal{N}_D(x)$ we denote the normal cone to the convex set D at x, that is, $\mathcal{N}_D(x) = \{w : \langle w, y - x\rangle \leq 0, \text{ for all } y \in D\}$ if $x \in D$ and $\mathcal{N}_D(x) = \emptyset$ otherwise.

The variational inequality (VI) [11] associated to a mapping F and a convex set D consists in determining a point $\bar{x} \in D$ such that the following inequality holds, for every $y \in D$: $\langle F(\bar{x}), y - \bar{x}\rangle \geq 0$. In terms of the normal cone, this means that $0 \in F(\bar{x}) + \mathcal{N}_D(\bar{x})$. The latter inclusion is called a generalized equation (GE). The MCP is a VI (equivalently, GE) with the set D defined by box constraints (where some bounds can be infinite).

10.2 A Simple Network of Agents

Our initial market is composed of producers, traders, and one end-consumption sector like in Fig. 10.1. Producers generate some kind of good (electricity, natural gas) that is sold to traders in an amount $S_P^i q_p^i$ for the ith producer. The jth trader buys from the producers an amount $B_T^j q_T^j$ and sells to consumers the product, after transporting and possibly modifying it, in an amount $S_T^j q_T^j$.

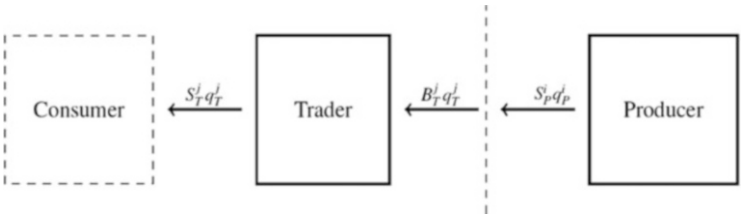

Fig. 10.1 A simple market

We shall see in Sect. 10.5.1 that the model can easily incorporate pipeline and storage operators, marketers, and other outsourcing agents like in [14]. For simplicity, and without loss of generality, in our presentation we analyze a network with only producers and traders that captures the main properties of the market model. Differently from [14], we consider a setup suitable for [9], in which decision variables are separated in two stages. For producers, for instance, some investment to increase capacity has to be decided at stage 0, in order to decide how much produce at stage 1. Another example is, in the presence of uncertainty, when the second-stage variables are a recourse to correct first-stage decisions, taken before knowing the realization of uncertainty [6].

In what follows, at equilibrium, all variables are denoted with a bar; for instance, $\bar{\pi}$ stands for an equilibrium price.

10.2.1 Producers, Traders, and Market Clearing

There are N_P producers, each one with decision variable (z_P^i, q_P^i). As mentioned, the variable z_P^i could refer to decisions concerning capacity or technological investments with a smooth convex cost $I_P^i(z_P^i)$. The variable q_P^i is related to operational activities involving a smooth convex cost $c_{P_i}(q_P^i)$. All the producer decision variables are taken in some set X_P^i which represents technological and resource constraints. After transformation of the raw materials, expressed by a matrix S_P^i of suitable dimensions, the producer has the quantity $S_P^i q_P^i$ for sale. In our model, we assume that producers are of the *price taker* type: there exists a market price that they cannot influence directly. So, for a given price π_P (exogenous to the players), each producer tries to maximize profit by solving the following problem:

$$\begin{cases} \max \left\langle S_P^i q_P^i, \pi_P \right\rangle - c_P^i(q_P^i) - I_P^i(z_P^i) \\ \text{s.t.} \quad (z_P^i, q_P^i) \in X_P^i . \end{cases} \tag{10.1}$$

The trader's model is similar; for $j = 1, \ldots, N_T$, the jth trader has decision variable (z_P^j, q_T^j). Given a transformation matrix B_T^j of suitable size, the trader buys $B_T^j q_T^j$ from the producers at price π_P. After modifying and/or transporting the product via a matrix S_T^j of suitable dimensions, the quantity $S_T^j q_T^j$ is sold to consumers at price π_T. The trader may have some additional (smooth convex) operational expenses $c_T^j(q_T^j)$ along the process and maximizes revenue by solving the following problem:

$$\begin{cases} \max \left\langle S_T^j q_T^j, \pi_T \right\rangle - \left\langle B_T^j q_T^j, \pi_P \right\rangle - c_T^j(q_T^j) - I_T^j(z_T^j) \\ \text{s.t.} \quad (z_T^j, q_T^j) \in X_T^j . \end{cases} \tag{10.2}$$

We shall see below that, as in [14], traders have a special role in the market and can exert market power by withholding supply from end costumers.

When the market is at equilibrium, there is no excess of generation and the producers' supply meets the traders' demand:

$$\sum_{i=1}^{N_P} S_P^i \bar{q}_P^i - \sum_{j=1}^{N_T} B_T^j \bar{q}_T^j = 0 \qquad \text{(mult. } \bar{\pi}_P\text{)}. \tag{10.3}$$

The rightmost notation means that the producers are remunerated at a price that clears the market: $\bar{\pi}_P$ is the multiplier corresponding to (10.3) at an equilibrium.

An environmentally responsible regulator can also impose a CO2 clearing condition, similar to (10.3), but involving different emission factors, depending on the technology employed to generate energy; see for instance [22, 31]. The essential feature of such constraints is that they *couple* the actions of different agents and in this sense (10.3) suffices for our development.

10.2.2 Consumer Modeling

The representation of the end-consumption sector can be done in different ways, depending on the manner price-taking producers operate in an imperfectly competitive market. Market imperfections can originate in regulatory measures such as price caps and emission limits and/or in traders exerting market power. We now review some alternatives that fit our general modeling.

10.2.2.1 Consumer Via Inverse-Demand Function

When a price-sensitive demand curve is available, the consumers' needs are represented implicitly by their inverse-demand function. Following [14], we model the demand curve by an affine function $P \cdot + d_0$, depending on given intercept d_0 and

matrix P. The dimension of d_0 is the same as of the traders' selling price [π_T in (10.2)]; the matrix P is of order $|\pi_T| \times |S_T^j q_T^j|$. At equilibrium the constraint

$$\sum_{j=1}^{N_T} P S_T^j \bar{q}_T^j + d_0 - \bar{\pi}_T = 0 \tag{10.4}$$

must be satisfied.

The inverse-demand function is useful to model the influence that the traders may exert on the market, a typical phenomenon in oligopolies. Instead of selling all the goods at price π_T (exogenous, hence not controllable), the trader sells a portion δ^j at price $\sum_{k=1}^{N_T} P S_T^k q_T^k + d_0$ (that depends on the amount of product the trader offers to the market). The factor $\delta^j \in [0,1]$ determines the strength of the influence the trader can have on the market. Accordingly, now the trader's problem (10.2) is

$$\begin{cases} \max \left\langle S_T^j q_T^j, \pi_T \right\rangle - \left\langle B_T^j q_T^j, \pi_P \right\rangle - c_T^j(q_T^j) - I_T^j(z_T^j) \\ \qquad\qquad + \delta^j \left\langle S_T^j q_T^j, \sum_{k=1}^{N_T} P S_T^k q_T^k + d_0 - \pi_T \right\rangle \\ \text{s.t.} \quad (z_T^j, q_T^j) \in X_T^j . \end{cases} \tag{10.2}_{\delta^j}$$

For future use, note that the initial problem (10.2) amounts to setting $\delta^j = 0$ for all the traders. Like for (10.2), both prices π_P and π_T are exogenous for the traders.

10.2.2.2 Consumer Via Explicit Demand Constraint

Sometimes instead of inverse-demand function there is a load duration curve segmented into blocks defining a vector D, which represents the consumers' demand. Accordingly, letting q^0 denote a nonnegative variable, at the equilibrium the constraint

$$\sum_{j=1}^{N_T} S_T^j \bar{q}_T^j + \bar{q}^0 - D = 0 \qquad \text{(mult. } \bar{\pi}_T\text{)} \tag{10.5}$$

should be satisfied. To prevent traders from exerting market power, and following [9], the deficit variable is related in a dual manner to a price cap imposed by the regulating agency:

$$\bar{\pi}_T \le PC \qquad \text{(mult. } \bar{q}^0\text{)}$$

with PC being the maximal allowed price. Note that, in view of their definitions, the variables q^0 and π_T have the same dimension.

In what follows, we refer to (10.1), $(10.2)_{\delta^j}$, (10.3), (10.4) as *implicit* model, while (10.1)–(10.3), (10.5) and the price-cap condition define the *explicit* model.

10.3 Equilibrium: Mixed Complementarity Formulation

For both consumer models, the equilibrium problem consists in computing prices $\bar{\pi}$ and decision variables $(\bar{z}, \bar{q})$ such that:

- For the ith producer, problem (10.1) written with price $\pi_P := \bar{\pi}_P$ is solved by $(\bar{z}_P^i, \bar{q}_P^i)$.
- For the jth trader, problem $(10.2)_{\delta^j}$ written with prices $(\pi_P, \pi_T) := (\bar{\pi}_P, \bar{\pi}_T)$ is solved by $(\bar{z}_T^i, \bar{q}_T^i)$, keeping in mind that if the explicit model is used, then $\delta^j = 0$ for all the traders.
- The market is cleared and (10.3) holds.
- Regarding the price at which the traders sell the final product:
  - If the implicit model is used, the relation (10.4) holds.
  - If the explicit model is used, both (10.5) and the price-cap conditions (cf. (10.8) below) hold.

For the sake of clarity we derive first the MCP when the consumers model is explicit, i.e., the trader's problem is (10.2) and both (10.5) and the price-cap condition hold.

10.3.1 MCP in the Presence of Explicit Demand Constraint

We start by writing down the Karush–Kuhn–Tucker (KKT) optimality conditions for the profit-maximization problems of the producers and traders. Typically, in (10.1) and (10.2), the feasible sets X_P^i and X_T^j are polyhedra, say, of the form

$$Z_P^i z_P^i + Q_P^i q_P^i \geq b_P^i \quad \text{and} \quad Z_T^j z_T^j + Q_T^j q_T^j \geq b_T^j,$$

respectively. Let μ_P^i and μ_T^j denote the corresponding Lagrange multipliers. The KKT conditions for the producers' problems (10.1), dropping the super-indices i to alleviate notation, are

$$\begin{aligned} 0 &= I_P'(z_P) - Z_P^\top \mu_P, \\ 0 &= c_P'(q_P) - Q_P^\top \mu_P - S_P^\top \pi_P, \\ 0 &\leq Z_P z_P + Q_P q_P - b_P \perp \mu_P \geq 0. \end{aligned} \tag{10.6}$$

Similarly for the traders, dropping the super-indices j, we write

$$\begin{aligned} 0 &= I_T'(z_T) - Z_T^\top \mu_T, \\ 0 &= c_T'(q_T) - Q_T^\top \mu_T + B_T^\top \pi_P - S_T^\top \pi_T, \\ 0 &\leq Z_T z_T + Q_T q_T - b_T \perp \mu_T \geq 0. \end{aligned} \tag{10.7}$$

The system is completed with (10.3), (10.5), and the price cap, written in the form:

$$0 \leq PC - \pi_T \perp q^0 \geq 0. \tag{10.8}$$

To write the associated GE in a compact form, we use the primal and dual variables defined by

$$\mathfrak{p} := \left((z_P^i)_{i=1}^{N_P}, (q_P^i)_{i=1}^{N_P}, (z_T^j)_{j=1}^{N_T}, (q_T^j)_{i=1}^{N_T}, q^0\right) \text{ and } \mathfrak{d} := \left((\mu_P^i)_{i=1}^{N_P}, (\mu_T^j)_{j=1}^{N_T}, \pi_P, \pi_T\right)$$

$$\text{over the sets} \quad \mathfrak{P} := \mathbb{R}^{\Sigma_{i=1}^{N_P}(|z_P^i|+|q_P^i|)+\Sigma_{j=1}^{N_T}(|z_T^j|+|q_T^j|)} \times \mathbb{R}^{m^0}_{\geq 0} \tag{10.9}$$

$$\text{and} \quad \mathfrak{D} := \mathbb{R}^{\Sigma_{i=1}^{N_P}|\mu_P^i|+\Sigma_{j=1}^{N_T}|\mu_T^j|}_{\geq 0} \times \mathbb{R}^{|\pi_P|+|\pi_T|}, \tag{10.10}$$

where $|q^0| = |\pi_T|$, by construction. For convenience, we introduce the operations $\texttt{diag}(\cdot)$, $\texttt{col}(\cdot)$, and $\texttt{row}(\cdot)$ for matrices $M^k, k = 1, \ldots, K$:

$$\texttt{diag}(M^k) := \begin{pmatrix} M^1 & & \\ & \ddots & \\ & & M^K \end{pmatrix}, \texttt{col}(M^k) := \begin{pmatrix} M^1 \\ \vdots \\ M^K \end{pmatrix}, \texttt{row}(M^k) := \begin{bmatrix} M^1 & \ldots & M^K \end{bmatrix}.$$

With this notation, the matrix below has $|\mathfrak{D}|$ rows and $|\mathfrak{P}|$ columns:

$$B := \begin{pmatrix} \texttt{diag}(Z_P^i) & \texttt{diag}(Q_P^i) & 0 & 0 & 0 \\ 0 & 0 & \texttt{diag}(Z_T^j) & \texttt{diag}(Q_T^j) & 0 \\ 0 & \texttt{row}(S_P^i) & 0 & -\texttt{row}(B_T^j) & 0 \\ 0 & 0 & 0 & \texttt{row}(S_T^j) & I \end{pmatrix}, \tag{10.11}$$

where I is an identity matrix of order $|\pi_T| = |q^0|$. Finally, we define the following operator acting on primal variables only and the following dual vector:

$$F(\mathfrak{p}) := \begin{pmatrix} (I_P^{i\prime}(z_P^i))_{i=1}^{N_P} \\ (c_P^{i\prime}(q_P^i))_{i=1}^{N_P} \\ (I_T^{j\prime}(z_T^j))_{j=1}^{N_T} \\ (c_T^{j\prime}(q_T^j))_{j=1}^{N_T} \\ PC \end{pmatrix} \quad \text{and} \quad b := \begin{pmatrix} (b_P^i)_{i=1}^{N_P} \\ (b_T^j)_{j=1}^{N_T} \\ 0 \\ D \end{pmatrix}. \tag{10.12}$$

The GE that results from putting together the relations in (10.6), (10.7), (10.3), (10.5), and (10.8) is

$$0 \in \begin{bmatrix} 0 & -B^\top \\ B & 0 \end{bmatrix} \begin{pmatrix} \mathfrak{p} \\ \mathfrak{d} \end{pmatrix} + \begin{pmatrix} F(\mathfrak{p}) \\ -b \end{pmatrix} + \mathscr{N}_{\mathfrak{P}\times\mathfrak{D}}(\mathfrak{p}, \mathfrak{d}). \tag{10.13}$$

10.3.2 MCP in the Presence of Inverse-Demand Function

As the traders' conditions are more involved when there is market power, we shall keep the super-indices j (as otherwise there might be some confusion); the optimality system for the traders then reads as follows:

$$
\begin{aligned}
0 &= I_T^{j\prime}(z_T^j) - Z_T^{j\top}\mu_T^j,\\
0 &= c_T^{j\prime}(q_T^j) - Q_T^{j\top}\mu_T^j + B_T^{j\top}\pi_P - (1-\delta^j)S_T^{j\top}\pi_T\\
&\qquad -\delta^j S_T^{j\top}\Big(\sum_{k=1}^{N_T} PS_T^k q_T^k + d_0\Big) - \delta^j S_T^{j\top} P^\top S_T^j q_T^j,\\
0 &\le Z_T^j z_T^j + Q_T^j q_T^j - b_T^j \perp \mu_T^j \ge 0\,.
\end{aligned}
\tag{10.14}
$$

As before, the KKT conditions (10.6) and (10.14), together with the market clearing condition (10.3) and the implicit representation of consumers via (10.4), give a GE on both primal and dual variables. There are a few differences with (10.13), though:

- There is no deficit q^0, so the primal variables and primal feasible set are now

  $$\tilde{\mathfrak{p}} := \Big((z_P^i)_{i=1}^{N_P}, (q_P^i)_{i=1}^{N_P}, (z_T^j)_{j=1}^{N_T}, (q_T^j)_{j=1}^{N_T}\Big) \text{ and } \tilde{\mathfrak{P}} := \mathbb{R}^{\sum_{i=1}^{N_P}(|z_P^i|+|q_P^i|)+\sum_{j=1}^{N_T}(|z_T^j|+|q_T^j|)}\,.$$

  Accordingly, instead of the matrix B from (10.11), we consider the sub-matrix $\tilde{B}$ obtained by eliminating from B the last row and column. Dual variables remain unchanged, so the GE uses $\tilde{B}$ and an additional row to represent (10.4).
- The market-power terms in the third line in (10.14) enter the primal operator

  $$\tilde{F}(\tilde{\mathfrak{p}}) := \begin{pmatrix} I_P^{i\prime}(z_P^i)\\ c_P^{i\prime}(q_P^i)\\ I_T^{j\prime}(z_T^j)\\ c_T^{j\prime}(q_T^j)\end{pmatrix} - \begin{pmatrix} 0\\ 0\\ 0\\ \delta^j S_T^{j\top}\Big(\sum_{k=1}^{N_T} PS_T^k q_T^k + d_0\Big) + \delta^j S_T^{j\top} P^\top S_T^j q_T^j\end{pmatrix}.$$

  To alleviate the writing we omitted the super-indices ranges: $i = 1,\ldots,N_P$ and $j = 1,\ldots,N_T$, which are clear from the context; see (10.12).
- Replacing (10.5) by (10.4) modifies the dual vector as follows: $\tilde{b} := \Big(b_P^i, b_T^j, 0, -d_0\Big)^\top$, where, once again, i and j run in their respective ranges, as in (10.12).

Finally, the GE with the implicit model is

$$0 \in \tilde{A}\begin{pmatrix}\tilde{\mathfrak{p}}\\ \mathfrak{d}\end{pmatrix} + \begin{pmatrix}\tilde{F}(\tilde{\mathfrak{p}})\\ -\tilde{b}\end{pmatrix} + \mathcal{N}_{\tilde{\mathfrak{P}}\times\mathfrak{D}}(\mathfrak{p},\mathfrak{d})\,, \tag{10.15}$$

for a matrix $\tilde{A}$ that, unlike the one in (10.13), is not skew-symmetric (and, moreover, has a last line relating primal and dual elements):

$$\tilde{A} := \begin{bmatrix} & 0 & & -\tilde{B}^\top \begin{pmatrix} 0\\ 0\\ 0\\ -\mathtt{col}((1-\delta^j)S_T^{j\top})\end{pmatrix}\\ \tilde{B} & & & 0\\ \begin{bmatrix}0 & 0 & \mathtt{row}(PS_T^j) & 0\end{bmatrix} & & 0 & -I\end{bmatrix}.$$

We shall see in Sect. 10.4 that GEs of the form (10.13) can be reduced to VI in smaller dimensions, which can in turn be interpreted in terms of a Nash game with shared constraints. By contrast, the GE (10.15) cannot be reformulated the same way directly. We next rewrite (10.15) in an equivalent form that does have the desired properties.

10.3.3 Inverse-Demand Function and an Extra Variable

Taking inspiration from the explicit model, we introduce a new primal variable p^0, gathering the portion of supply that the traders cannot influence by exerting market power. Thus, we require the relation

$$\sum_{j=1}^{N_T} (1-\delta^j) S_T^j \bar{q}_T^j - \bar{p}^0 = 0 \tag{10.16}$$

to be satisfied when the market is at an equilibrium point. In view of its definition, this new variable has the same dimension as the deficit variable q^0 from (10.5) in the explicit model (and, hence, $|p^0| = |\pi_T|$).

The GE gathering (10.6), (10.14), (10.3), (10.4), and (10.16) now employs the primal objects

$$\hat{\mathfrak{p}} := \left(\tilde{\mathfrak{p}}, p^0\right) \text{ and } \hat{\mathfrak{P}} := \mathbb{R}^{\sum_{i=1}^{N_P}(|z_P^i|+|q_P^i|)+\sum_{j=1}^{N_T}(|z_T^j|+|q_T^j|)+|p^0|}, \tag{10.17}$$

noting that the dual variables remain the same from the explicit model, given in (10.10). The primal sets in the implicit and explicit models, from (10.17) and (10.9), respectively, only differ in their last component (q^0 and p^0, respectively). While in the explicit model the deficit is nonnegative [as a multiplier of the price cap (10.8)], in the implicit model the new primal variable is unconstrained. So the normal cone to q^0 will be the null vector and we can require satisfaction of the inverse-demand relation (10.4) in the corresponding new component of the GE. This eliminates the primal–dual coupling in the last line of matrix $\tilde{A}$ in (10.15). Similarly for ensuring (10.16), recalling that the π_T-component of the dual set is the whole space.

The resulting GE is

$$0 \in \begin{bmatrix} 0 & -\hat{B}^\top \\ \hat{B} & 0 \end{bmatrix} \begin{pmatrix} \hat{\mathfrak{p}} \\ \mathfrak{d} \end{pmatrix} + \begin{pmatrix} \hat{F}(\hat{\mathfrak{p}}) \\ -\hat{b} \end{pmatrix} + \mathcal{N}_{\hat{\mathfrak{P}}\times\mathfrak{D}}(\mathfrak{p}, \mathfrak{d}), \tag{10.18}$$

where we defined a matrix $\hat{B}$ of order $|\hat{\mathfrak{P}}|$ and $|\mathfrak{D}|$:

$$\hat{B} := \begin{pmatrix} & \tilde{B} & & 0 \\ \begin{bmatrix} 0 & 0 & 0 & \texttt{row}((1-\delta^j)S_T^j) \end{bmatrix} & & & -\hat{I} \end{pmatrix}, \tag{10.19}$$

using an identity $\hat{I}$ of order $|\pi_T| = |p^0|$ and the primal operator and dual vector

$$\hat{F}(\hat{\mathfrak{p}}) := \begin{pmatrix} I_P^{i\prime}(z_P^i) \\ c_P^{i\prime}(q_P^i) \\ I_T^{j\prime}(z_T^j) \\ c_T^{j\prime}(q_T^j) - \delta^j S_T^{j\top}(\sum_{k=1}^{N_T} PS_T^k q_T^k + d_0) - \delta^j S_T^{j\top} P^\top S_T^j q_T^j \\ -\sum_{k=1}^{N_T} PS_T^k q_T^k - d_0 \end{pmatrix} \quad \text{and} \tag{10.20}$$

$$\hat{b} := \begin{pmatrix} b_P^i \\ b_T^j \\ 0 \\ 0 \end{pmatrix}.$$

Like for (10.12), in both vectors $i = 1,\ldots,N_P$ and $j = 1,\ldots,N_T$.

10.4 Equivalent Mixed Complementarity Formulations

Both GEs (10.13) and (10.18) are defined using very simple normal cones and have a very specific primal–dual structure. The size of both GEs is the same: the respective primal and dual sets only differ in their last primal component; $q^0 \geq 0$ in the explicit model and unconstrained p^0 in the implicit one.

To establish the relation of the MCP models with a game-theoretical formulation, we state a result from [27]; see also [15]. Our GEs are a particular case of the setting covered by the reduction method in [27], as the primal sets $\mathfrak{P}$ are cones in both models. Here, the relation with a game could actually be also shown directly, by comparing the KKT conditions of the MCP model with those for a game. We prefer to state the more general result, because it includes a nice characterization of dual variables as solutions to a certain linear programming problem, defined a posteriori, once the primal solution is available. As dual variables have economical meaning as prices, this is an interesting feature; see Remark 1.

Theorem 1. *The following statements are equivalent:*

Primal–dual GE: *the primal–dual pair* $(\bar{\mathfrak{p}}, \bar{\mathfrak{d}})$ *satisfies* (10.13).
Primal GE + dual LP: *the primal variable* $\bar{\mathfrak{p}}$ *solves the generalized equation*

$$0 \in F(\mathfrak{p}) + \mathcal{N}_{\mathfrak{P}^0}(\mathfrak{p}), \tag{10.21}$$

where $\mathfrak{P}^0 := \mathfrak{P} \cap \mathcal{S}$ *and*

$$\mathcal{S} := \left\{ \mathfrak{p} : b - B\mathfrak{p} \in \mathbb{R}^{\Sigma_{i=1}^{N_P} |\mu_P^i| + \Sigma_{j=1}^{N_T} |\mu_T^j|}_{\leq 0} \times \{0 \in \mathbb{R}^{|\pi_P| + |\pi_T|}\} \right\}.$$

As for the dual variable, $\bar{\mathfrak{d}}$ solves the linear programming problem

$$\begin{cases} \min \ \langle B\bar{\mathfrak{p}} - b, \mathfrak{d} \rangle \\ \text{s.t.} \ \ B^\top \mathfrak{d} - F(\bar{\mathfrak{p}}) \in \mathscr{N}_{\mathfrak{P}}(\bar{\mathfrak{p}}) \\ \qquad \mathfrak{d} \in \mathfrak{D}\,. \end{cases} \tag{10.22}$$

Proof. The statement is just a rewriting of Propositions 1 and 2 in [27] in our notation. Specifically, the respective correspondence for primal elements is $(p, d(p), P) = (\mathfrak{p}, -F(\mathfrak{p}), \mathfrak{P})$, for the dual ones $(y, Y) = (\mathfrak{d}, \mathfrak{D})$, and for the matrix and vector $(A, f) = -(B^\top, b)$. Our set $\mathscr{S}$ corresponds to Z in Proposition 2, using the fact that in our setting the polar cone therein, $Y^0 = \mathfrak{D}^0$, has a very simple expression. □

Existence of solutions to the GE (10.21) can be guaranteed under mild assumptions, such as continuity of F and convexity and compactness of $\mathfrak{P}^0$, [11, Corollary 2.2.5]. These conditions are natural in our context: components of F consist of derivatives of smooth convex functions and the feasible set $\mathfrak{P}^0$ represents limited resources. Furthermore, the existence of solutions of (10.21) implies the existence of solutions of the (bounded) linear program (10.22), whose optimal value is zero.

The interest of Theorem 1 is twofold. First, the GE (10.21) is in primal variables only, stated over a set that (for both of our models) is a simple polyhedron. It is therefore a VI with linear constraints. We shall see that in some cases the multipliers corresponding to the constraints provide the equilibrium prices. Once a primal solution is at hand, the dual component of the MCP solution can be found by solving an easy linear program. This feature is attractive to identify (undesirable) situations in which equilibrium prices are not unique, even if the primal part of the equilibrium points is unique (the linear program solution will not be unique in this case; see Remark 1 below). A second advantage of the equivalent formulation is that, in addition to providing a mechanism for ensuring existence of solutions of the game, the reformulation reveals the particular structure of the set $\mathfrak{P}^0$, amenable to decomposition. More precisely, without the coupling constraints (some components in $b - B\mathfrak{p}$, hence in $\mathscr{S}$), the feasible set is decomposable [like $\mathfrak{P}$ from (10.9), (10.17)]. This decomposable structure can be exploited by decomposition methods, like the Dantzig–Wolfe algorithms developed in [20]; see also [3, 13].

10.4.1 Game for the Explicit Model

For the market in Sect. 10.2, instead of viewing the agents as maximizing revenue like in the complementarity model, we consider a generalized Nash equilibrium problem (GNEP) [10] with players minimizing costs. The coupling constraints in the game are (10.3) and (10.5). In addition to the traders and producers, there is an additional player, indexed by number "0," in charge of capping prices. Specifically, the purpose of the game is to find $\tilde{\mathfrak{p}} = \left((\tilde{z}_P^i)_{i=1}^{N_P}, (\tilde{q}_P^i)_{i=1}^{N_P}, (\tilde{z}_T^j)_{j=1}^{N_T}, (\tilde{q}_T^j)_{i=1}^{N_T}, \tilde{q}^0 \right)$ such that the following minimization problems are solved by $\tilde{\mathfrak{p}}$:

$$\textbf{Producers} \quad \begin{cases} \min_{(z_P^i, q_P^i) \in X_P^i} \; I_P^i(z_P^i) + c_P^i(q_P^i) \\ \text{s.t.} \qquad\qquad S_P^i q_P^i + \sum_{i \neq k=1}^{N_P} S_P^k \tilde{q}_P^k - \sum_{j=1}^{N_T} B_T^j \tilde{q}_T^j = 0, \end{cases} \tag{10.23}$$

$$\textbf{Traders} \quad \begin{cases} \min_{(z_T^j, q_T^j) \in X_T^j} \; I_T^j(z_T^j) + c_T^j(q_T^j) \\ \text{s.t.} \qquad -B_T^j \tilde{q}_T^j + \sum_{i=1}^{N_P} S_P^i \tilde{q}_P^i - \sum_{j \neq k=1}^{N_T} B_T^k \tilde{q}_T^k = 0, \\ \qquad\qquad S_T^j q_T^j + \sum_{j \neq k=1}^{N_T} S_T^k \tilde{q}_T^k + \tilde{q}^0 - D = 0, \end{cases} \tag{10.24}$$

$$\textbf{Consumer representative} \quad \begin{cases} \min_{q^0 \geq 0} \; \langle PC, q^0 - D \rangle \\ \text{s.t.} \qquad \sum_{j=1}^{N_T} S_T^j \tilde{q}_T^j + q^0 - D = 0. \end{cases} \tag{10.25}$$

In the GNEP (10.23)–(10.25), the market between producers and traders is cleared, and demand is satisfied up to certain deficit, q^0. The deficit is minimized by the action of the additional player, who tries to reduce the impact of imposing a price cap. In Corollary 1 below it is shown that (the negative of) the multiplier of the coupling constraint (10.5) is precisely the traders' remuneration in (10.2). We shall also see that in the game formulation, the price cap is maintained in an indirect manner, via (10.25).

In the game, the solution of each individual problem depends on the decisions of the other agents in the market: for instance (10.24) is an optimization problem on the jth trader variables (say, $\mathfrak{p}_j$) which depends on actions of other traders (say, on $\mathfrak{p}_{-j}$). A primal point $\bar{\mathfrak{p}}$ is a Nash equilibrium for the game (10.23)–(10.25) when each player's optimal decision (say, $\bar{\mathfrak{p}}_j$) is obtained by solving the individual problem [say, (10.24)] after fixing the other players' decisions to the corresponding entries on $\bar{\mathfrak{p}}$ (say $\bar{\mathfrak{p}}_{-j}$).

As this notion is so general that it includes points contradicting the natural intuition of what an equilibrium must be, it is further specialized to the notion of *variational equilibrium*, as follows. Note that the value function for the producers

$$v_P^i(x) := \begin{cases} \min_{(z_P^i, q_P^i) \in X_P^i} \; I_P^i(z_P^i) + c_P^i(q_P^i) \\ \text{s.t.} \qquad\qquad S_P^i q_P^i + \sum_{i \neq k=1}^{N_P} S_P^k \tilde{q}_P^k - \sum_{j=1}^{N_T} B_T^j \tilde{q}_T^j = x, \end{cases}$$

is convex. Furthermore, because in (10.23) all constraints are linear and the objective function is differentiable, there exists a Lagrange multiplier $\check{\pi}_P^i$ associated to the equality constraint. This multiplier represents a marginal cost, since it satisfies the inclusion $-\check{\pi}_P^i \in \partial v_P^i(0)$ [16, Theorem VII.3.3.2]. The issue with a generic Nash equilibrium like $\bar{\mathfrak{p}}$ above is that it may have multipliers associated to coupling constraints of the players' problems that are different for different players. In

economical terms, this means that the equilibrium is "unfair," because it benefits some players more than others. To avoid this undesirable feature, we shall solve a VI derived from the game and find a variational equilibrium (VE) [10] of the GNEP (10.23)–(10.25), ensuring that the multipliers associated with the coupling constraints are the same.

By Theorem 1, the GE (10.13) is equivalent to solving the GE (10.21), written with the data from Sect. 10.2.2.2. Putting together (10.11), (10.9), and (10.12) yields for (10.21) the following:

$$0 \in F(\mathfrak{p}) + \mathcal{N}_{\mathfrak{P}^0}(\mathfrak{p}),$$

where the feasible set $\mathfrak{P}^0 := \prod_{i=1}^{N_P} X_P^i \times \prod_{j=1}^{N_T} X_T^j \times \mathbb{R}^{m^0}_{\geq 0} \cap \mathscr{S}$ depends on the coupling set $\mathscr{S} := \{\mathfrak{p} = (z_P^i, q_P^i, z_T^j, q_P^j, q^0) : (10.3) \text{ and } (10.5) \text{ hold}\}$.

The equivalence between the MCP formulation and the generalized Nash game results from Theorem 1.

Corollary 1 (Game Formulation for the Explicit Model). *The MCP in Sect. 10.2.2.2 and the game* (10.23)–(10.25) *are equivalent in the following sense. Suppose the game has a variational equilibrium*

$$\bar{\mathfrak{p}} := \left((\breve{z}_P^i)_{i=1}^{N_P}, (\breve{q}_P^i)_{i=1}^{N_P}, (\breve{z}_T^j)_{j=1}^{N_T}, (\breve{q}_T^j)_{i=1}^{N_T}, \breve{q}^0 \right),$$

with $(\breve{\mu}_P^i)_{i=1}^{N_P}, (\breve{\mu}_T^j)_{j=1}^{N_T}$ *being the corresponding multipliers for the constraints in* (10.23) *and* (10.24)*, and let* $\breve{\pi}_P$ *and* $\breve{\pi}_T$ *be the multipliers associated to the coupling constraints* (10.3) *and* (10.5)*.*

Then the primal–dual pair $(\bar{\mathfrak{p}}, \bar{\mathfrak{d}})$ *with* $\bar{\mathfrak{d}} := (\breve{\mu}_P, \breve{\mu}_T, -\breve{\pi}_P, -\breve{\pi}_T)$ *solves the MCP given by* (10.1)–(10.3)*,* (10.5)*, and* (10.8)*.*

Proof. By Theorem 1, for the result to hold, $\bar{\mathfrak{d}}$ needs to solve the linear program therein. For the objects in (10.13), and for the normal cone to the primal set $\mathfrak{P}$ from (10.9), this linear program is

$$\left\{ \begin{array}{ll} \min\limits_{\substack{\mu_P, \mu_T \geq 0 \\ \text{any } \pi_P, \pi_T}} & \sum\limits_{i=1}^{N_P} \left\langle Z_P^i \breve{z}_P^i + Q_P^i \breve{q}_P^i - b_P^i, \mu_P^i \right\rangle + \sum\limits_{j=1}^{N_T} \left\langle Z_T^j \breve{z}_T^j + Q_T^j \breve{q}_T^j - b_T^j, \mu_T^j \right\rangle \\ \text{s.t.} & Z_P^{i\top} \mu_P^i = I_P^{i\prime}(\breve{z}_P^i), \, Z_T^{j\top} \mu_T^j = I_T^{j\prime}(\breve{z}_T^j), \\ & Q_P^{i\top} \mu_P^i + S_P^{i\top} \pi_P = c_P^{i\prime}(\breve{q}_P^i), \\ & Q_T^{j\top} \mu_T^j - B_T^{j\top} \pi_P + S_T^{j\top} \pi_T = c_T^{j\prime}(\breve{q}_T^j), \\ & \pi_T \leq PC \text{ and } \pi_T^k = PC^k \text{ whenever } \breve{q}^{0k} > 0. \end{array} \right. \tag{10.26}$$

The optimality conditions for problems (10.23) and (10.24) amount to $\breve{\mu}_P, \breve{\mu}_T, -\breve{\pi}_P$, and $-\breve{\pi}_T$ to satisfy the first four equalities in the feasible set of (10.26). Note also that, by complementarity, the (nonnegative) objective function attains its minimum value at $\breve{\mu}_P, \breve{\mu}_T$. The last line in (10.26), written with $-\breve{\pi}_T$, is $\breve{q}^0 \perp PC + \breve{\pi}_T \geq 0$. As these relations result from the optimality condition of (10.25), the desired result follows. □

10.4.2 Game for the Implicit Model

We now apply Theorem 1 to the GE (10.18). Writing (10.21) with the data from Sect. 10.2.2.1, that is, using (10.19), (10.17) and (10.20), we have

$$0 \in \hat{F}(\hat{\mathfrak{p}}) + \mathscr{N}_{\mathfrak{P}^0}(\hat{\mathfrak{p}}) \quad \text{where } \mathfrak{P}^0 := \prod_{i=1}^{N_P} X_P^i \times \prod_{j=1}^{N_T} X_T^j \times \mathbb{R}^{|p^0|} \cap \mathscr{S},$$

for $\mathscr{S} := \{(z_P^i, q_P^i, z_T^j, q_P^j, p^0) : (10.3) \text{ and } (10.16) \text{ hold}\}$.

The MCP formulation of (10.1), $(10.2)_{\delta^j}$, (10.3), and (10.4) is now equivalent to finding a variational equilibrium of the following GNEP:
the point $\tilde{\mathfrak{p}} = \left((\tilde{z}_P^i)_{i=1}^{N_P}, (\tilde{q}_P^i)_{i=1}^{N_P}, (\tilde{z}_T^j)_{j=1}^{N_T}, (\tilde{q}_T^j)_{i=1}^{N_T}, \tilde{q}^0\right)$ solves the problems

Producers same as (10.23),

$$\textbf{Traders} \quad \begin{cases} \min\limits_{(z_T^j, q_T^j) \in X_T^j} & I_T^j(z_T^j) + c_T^j(q_T^j) \\ & \quad - \delta^j \left\langle \sum\limits_{k=1}^{N_T} PS_T^k q_T^k + d_0, S_T^j q_T^j \right\rangle \\ \text{s.t.} & \sum\limits_{i=1}^{N_P} S_P^i q_P^i - \sum\limits_{k=1}^{N_T} B_T^k q_T^k = 0, \\ & \sum\limits_{k=1}^{N_T} (1-\delta^k) S_T^k q_T^k - p^0 = 0, \end{cases} \tag{10.27}$$

$$\textbf{Consumer representative} \quad \begin{cases} \max\limits_{p^0} & \left\langle \sum\limits_{k=1}^{N_T} PS_T^k q_T^k + d_0, p^0 \right\rangle \\ \text{s.t.} & \sum\limits_{j=1}^{N_T} (1-\delta^j) S_T^j q_T^j - p^0 = 0. \end{cases} \tag{10.28}$$

The game (10.23) and (10.27)–(10.28) can be interpreted as follows. The additional player tries to maximize the traders' revenue that is market-power-free (given in terms of the inverse-demand function). The traders see their influence on the market as a way of reducing costs or of increasing their income [the negative δ^j term in the objective function from (10.27)]. Transactions between producers and traders are cleared, as before. Regarding the traders remuneration π_T [i.e., the multiplier of constraint (10.16)], we now show that the additional player controls it in a manner ensuring satisfaction of (10.4).

Corollary 2 (Game Formulation for the Implicit Model). *The MCPs in Sects. 10.2.2.1 and 10.3.3 and the game* (10.23) *and* (10.27)–(10.28) *are equivalent in the following sense. Suppose the game has a variational equilibrium*

$$\breve{\mathfrak{p}} := \left((\breve{z}_P^i)_{i=1}^{N_P}, (\breve{q}_P^i)_{i=1}^{N_P}, (\breve{z}_T^j)_{j=1}^{N_T}, (\breve{q}_T^j)_{i=1}^{N_T}, \breve{p}^0\right),$$

with $(\breve{\mu}_P^i)_{i=1}^{N_P}, (\breve{\mu}_T^j)_{j=1}^{N_T}$ *being the corresponding multipliers for the constraints in* (10.23) *and* (10.27)*, and let* $\breve{\pi}_P$ *and* $\breve{\pi}_T$ *be the multipliers associated to the coupling constraints* (10.3) *and* (10.16)*.*

Then the primal–dual pair $(\bar{\mathfrak{p}}, \bar{\mathfrak{d}})$ *with* $\bar{\mathfrak{d}} := (\breve{\mu}_P, \breve{\mu}_T, -\breve{\pi}_P, -\breve{\pi}_T)$ *solves the MCP* (10.18)*, which is equivalent to* (10.15)*.*

Proof. Like for Corollary 1, we only need to show that $\bar{\mathfrak{d}}$ solves the linear program in Theorem 1. In this case, the normal cone to the primal set $\hat{\mathfrak{P}}$ from (10.17) is just the null vector and, hence, the linear program is

$$\begin{cases} \min\limits_{\substack{\mu_P, \mu_T \geq 0 \\ \text{any } \pi_P, \pi_T}} \sum\limits_{i=1}^{N_P} \left\langle Z_P^i \breve{z}_P^i + Q_P^i \breve{q}_P^i - b_P^i, \mu_P^i \right\rangle + \sum\limits_{j=1}^{N_T} \left\langle Z_T^j \breve{z}_T^j + Q_T^j \breve{q}_T^j - b_T^j, \mu_T^j \right\rangle \\ \text{s.t.} \quad Z_P^{i\top} \mu_P^i = I_P^{i\prime}(\breve{z}_P^i), \quad Z_T^{j\top} \mu_T^j = I_T^{j\prime}(\breve{z}_T^j), \\ \qquad Q_P^{i\top} \mu_P^i + S_P^{i\top} \pi_P = c_P^{i\prime}(\breve{q}_P^i), \\ \qquad Q_T^{j\top} \mu_T^j - B_T^{j\top} \pi_P + (1-\delta^j) S_T^{j\top} \pi_T = c_T^{j\prime}(\breve{q}_T^j) \\ \qquad\qquad\qquad - \delta^j S_T^{j\top} \left(\sum\limits_{k=1}^{N_T} P S_T^k \breve{q}_T^k + d_0 \right) - \delta^j S_T^{j\top} P^\top S_T^j \breve{q}_T^j \\ \qquad \pi_T = \sum\limits_{k=1}^{N_T} P S_T^k \breve{q}_T^k + d_0 \,. \end{cases} \tag{10.29}$$

It is easy to see that all the relations in KKT conditions of this problem are verified by $\bar{\mathfrak{d}}$, except for the last equality, corresponding to (10.4). For the latter, observe that since $\breve{\pi}_T$ is the multiplier of the coupling constraint (10.16), and the variable p^0 is unconstrained in the problem of the extra player (10.28), the p^0-component of the optimality conditions for the game gives that $0 = -\sum_{k=1}^{N_T} P S_T^k \breve{q}_T^k - d_0 - \breve{\pi}_T$. □

Remark 1 (Uniqueness of Prices). With the explicit model, equilibrium prices will be unique if the linear program (10.26) has the unique solution. Likewise for the implicit model, which depends on the linear program (10.29). This problem can be further simplified, by eliminating the variable π_T, as follows:

$$\begin{cases} \min\limits_{\substack{\mu_P, \mu_T \geq 0 \\ \text{any } \pi_P}} \sum\limits_{i=1}^{N_P} \left\langle Z_P^i \breve{z}_P^i + Q_P^i \breve{q}_P^i - b_P^i, \mu_P^i \right\rangle + \sum\limits_{j=1}^{N_T} \left\langle Z_T^j \breve{z}_T^j + Q_T^j \breve{q}_T^j - b_T^j, \mu_T^j \right\rangle \\ \text{s.t.} \quad Z_P^{i\top} \mu_P^i = I_P^{i\prime}(\breve{z}_P^i), \quad Z_T^{j\top} \mu_T^j = I_T^{j\prime}(\breve{z}_T^j), \\ \qquad Q_P^{i\top} \mu_P^i + S_P^{i\top} \pi_P = c_P^{i\prime}(\breve{q}_P^i), \\ \qquad Q_T^{j\top} \mu_T^j - B_T^{j\top} \pi_P + S_T^{j\top} \left(\sum\limits_{k=1}^{N_T} P S_T^k \breve{q}_T^k + d_0 \right) = c_T^{j\prime}(\breve{q}_T^j) - \delta^j S_T^{j\top} P^\top S_T^j \breve{q}_T^j \,. \end{cases}$$

10.5 The European Network of Natural Gas

We now consider a network with a third kind of player, called *outsourcer*, in charge of modifying or transporting the product before the traders supply it to the end consumers.

Like before, producers only deal with traders and, therefore, solve problems (10.1). By contrast, traders now deal also with the outsourcer players, who charge a unitary price π_O for their activity. The exchange between the trader and the outsourcer player involves transformation of the product, represented by matrices $S^j_{T\to O}$, $B^j_{T\leftarrow O}$, S^k_O, B^k_O as schematically represented in Fig. 10.2, with the product flow.

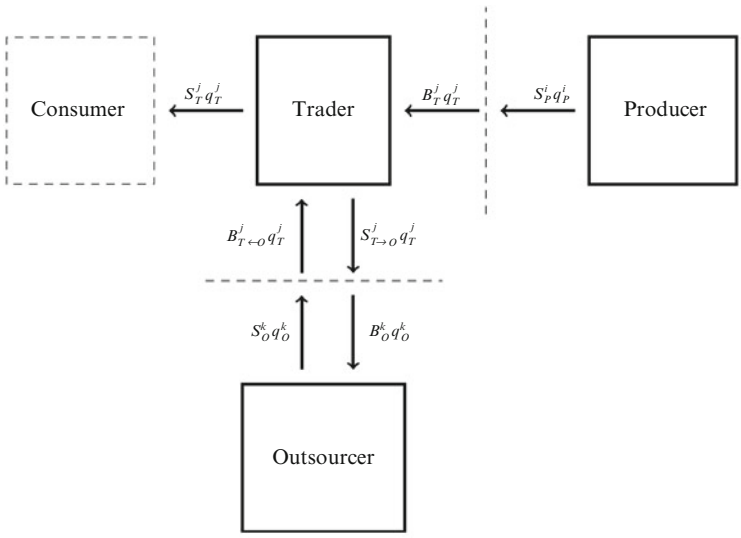

Fig. 10.2 Market flow

The jth trader problem (10.2) is modified accordingly:

$$\begin{cases} \max \left\langle S^j_T q^j_T, \pi_T \right\rangle - \left\langle B^j_T q^j_T, \pi_P \right\rangle - \left\langle S^j_{T\to O} q^j_T, \pi_O \right\rangle - c^j_T(q^j_T) - I^j_T(z^j_T) \\ \text{s.t.} \quad (z^j_T, q^j_T) \in X^j_T \,. \end{cases}$$

As for the outsourcing players, denoting once more the investment-operational decision variables of the kth agent by (z^k_O, q^k_O) and similarly for the costs and feasible set, the corresponding maximization problem is

$$\begin{cases} \max \left\langle B^k_O q^k_O, \pi_O \right\rangle - c^k_O(q^k_O) - I^k_O(z^k_O) \\ \text{s.t.} \quad (z^k_O, q^k_O) \in X^k_O \,. \end{cases} \tag{10.30}$$

To clear the market, in addition to (10.3) and (10.4), the exchange between traders and outsourcing players should be balanced and, hence,

$$\sum_{j=1}^{N_T} S^j_{T\to O}\bar{q}^j_T - \sum_{k=1}^{N_O} B^k_O \bar{q}^k_O = 0\,.$$

The additional balance $\sum_{k=1}^{N_O} S^k_O \bar{q}^k_O - \sum_{j=1}^{N_T} B^j_{T\leftarrow O}\bar{q}^j_T = 0$ is omitted, because it is often automatic from the condition above.

Our previous framework, starting in Sect. 10.2, covers the new network. This network, considered in [14] to analyze the market of natural gas in Europe (using an MCP formulation corresponding to the model in Sect. 10.2.2.1), is now considered as a test case in the numerical experience that follows.

10.5.1 Numerical Assessment

The full European gas network described in [14] covers 54 countries and 36 markets; the market has seven types of players representing producers, traders, and five different outsourcing activities. Specifically, there are 28 producers, 22 traders, 10 liquefiers, 15 re-gasifiers, 22 storage operators, 74 pipeline operators, and 36 marketers.

To illustrate the analysis that can be derived from the models presented above, we coded the MCP and game implicit models in Matlab (R2012a), using PATH [7, 12] to solve the variational problems. The runs were performed on a PC operating under Ubuntu 12.04–64 bit with a processor Intel Atom 1.80 GHz $\times$ 4 and 2 GB of memory.

The data in [14] gives a game problem with 4,620 variables and 488 constraints. We solved the equilibrium problem of the implicit model, with and without market power. In the first instance, the trader's problem $(10.2)_{\delta^j}$ has $\delta^j \equiv 0$. In the second, $\delta^j = 0.75$ for Russia, Norway, the Netherlands, and Algeria; and $\delta^j = 0.25$ for the Caspian Sea, Denmark, and the UK.

To ensure that the implementation is error-free, we first ran both formulations, that is, the game (10.23), (10.27), (10.28) and the MCP (10.15), and checked whether the corresponding output was alike. Table 10.1 summarizes the results.

Table 10.1 Output for the implicit model

Formulation	Market power	PATH residual	CPU (s)
Game	No	2E−08	36.7
MCP	No	7E−08	47.3
Game	Yes	7E−11	77.1
MCP	Yes	2.5E−11	201.5

The ∞-norms of the differences of the primal solutions obtained with both approaches were very small in all the cases. We observed larger differences in the dual components, in percentages ranging up to 6 % (for the competitive case, without market power). However, this is still an insignificant difference in this context, which allows us to conclude that the output of both formulations is indeed "the same" and the implementations are correct.

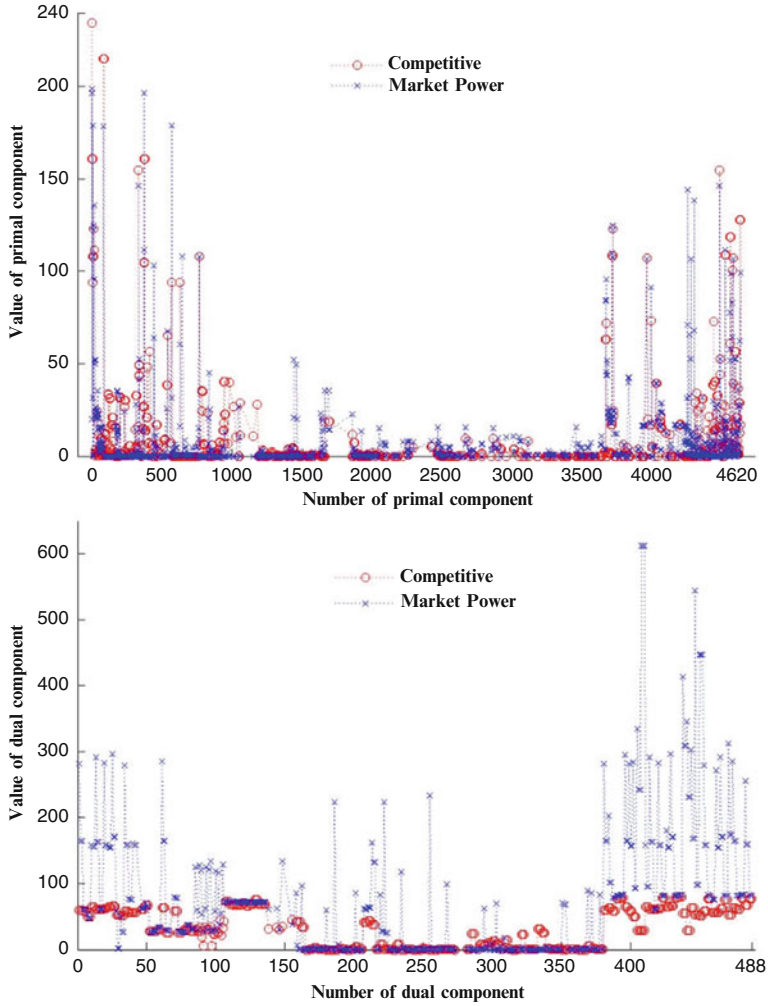

Fig. 10.3 Comparison of the primal and dual output

An interesting information in Table 10.1 is the CPU times. In general PATH was very fast, but solving time increased significantly for the MCP formulation

when there is market power. At this point, one could ask why this increase is of importance, given that the solution times were still within some minutes. The answer is that this increase, still significant in percentage terms, would blow up once stochasticity is introduced to the model. According to PATH final convergence report, when there is market power, the solver needed much more inner iterations to converge. We observed that when decreasing the solver precision from 10^{-8} to 10^{-6}, both formulations were again solved in about seventies.

A comparison of the results obtained with and without market power can be found in Fig. 10.3, whose top and bottom graphs correspond to the primal and dual output, respectively. The impact of market power is especially noticeable in the dual variables, corresponding to prices: in the bottom graph in Fig. 10.3 the red circles (competitive prices) are systematically lower than the blue crosses (market power). The graphs also show that the largest price increase is in the last components of the dual output, corresponding to the variable $\bar{\pi}_T$, i.e., to the remuneration of the traders.

10.6 Equilibrium for Stochastic Models

Realistic models for the energy industry often include *uncertainty*: for instance in (10.5), the actual electrical load may deviate from the predicted one due to random variations of temperature, switch off/on of local consumers, or daylight and similarly in (10.1), for the generation costs $c_P^i(\cdot)$ or the available resources defining the feasible sets X_P^i. To reflect such variations, a stochastic model of uncertainty must be built and the risk-averse decision process must be put in a suitable setting.

In what follows we no longer distinguish between producers, traders, and outsourcer players. Instead, we analyze a market with agents trying to maximize profit on a market regulated by coupling constraints or by a price cap. Accordingly, we unify the notation for problems (10.1)–(10.2) and consider that the agents solve $\max \langle \pi, S^i q^i \rangle - I^i(z^i) - c^i(q^i)$, which is equivalent to $\min I^i(z^i) + c^i(q^i) - \langle \pi, S^i q^i \rangle$.

10.6.1 Hedging Risk: The Setting

Consider the probability space defined by a measure $\mathbb{P}$ on a sample space Ω equipped with a sigma-algebra $\mathscr{F}$. Decision variables are now random functions in the space $L_p(\Omega, \mathscr{F}, \mathbb{P})$ for $p \in [1, +\infty)$, with dual $L_{p^*}(\Omega, \mathscr{F}, \mathbb{P})$ for $p^* \in (1, +\infty]$ such that $1/p + 1/p^* = 1$. We sometimes use the shorter notation L_p and L_{p^*} for these spaces, which are paired by the duality product

$$\langle x^*, x \rangle_{\mathbb{P}} = \int_{\Omega} \langle x^*(\omega), x(\omega) \rangle \, d\mathbb{P}(\omega) .$$

In the presence of uncertainty, a natural reaction of agents in the market is to hedge against undesirable events. For the ith agent, aversion to volatility is expressed by

a coherent (convex) risk measure $\rho^i(\cdot)$, assumed to be a proper function, as in [6, Chap. 6]. One possibility in the space $L_1(\Omega, \mathscr{F}, \mathbb{P})$ is to take the *Average Value-at-Risk* of level $1-\varepsilon_i$, a recent renaming of the *Conditional Value-at-Risk* [28]. Namely, given a confidence level $0 < 1-\varepsilon < 1$, if the random outcome $X \in L_1$ represents a loss (lower values are preferred), the measure is given by the expression

$$AV@R_\varepsilon(X) := \min_u \left\{ u + \frac{1}{1-\varepsilon}\mathbb{E}[X(\omega) - u]^+ \right\},$$

where $[\cdot]^+ := \max\{0, \cdot\}$ is the positive-part function and $\mathbb{E}(\cdot)$ denotes the expected-value function taken with respect to $d\mathbb{P}$. We consider the more general functions

$$\rho^i(X) := (1-\kappa_i)\mathbb{E}(X) + \kappa_i AV@R_{\varepsilon_i}(X), \tag{10.31}$$

depending on a given risk-aversion parameter $\kappa_i \in [0,1]$.

It is shown in [6, Theorem 6.4] that any proper coherent risk measure is in fact the support function of the domain of its conjugate; see also [24]. In particular (see [6, Theorem 6.4, and (6.69) in Exercise 6.16]), (10.31) has the dual representation

$$\rho^i(X) = \sup_{x^* \in X^*} \langle x^*, X \rangle_{\mathbb{P}}, \text{ where} \tag{10.32}$$

$$X^* := \left\{ x^* \in L_\infty(\Omega, \mathscr{F}, \mathbb{P}) : \begin{array}{l} 1-\kappa_i \le x^*(\omega) \le 1-\kappa_i+\kappa_i/\varepsilon_i \ a.e.\ \omega \in \Omega \\ \mathbb{E}(x^*) = 1 \end{array} \right\}.$$

10.6.2 Stochastic Mixed Complementarity Formulation

For convenience, from now on, we make two simplifying assumptions:

- The concept of stochastic equilibrium and its connections with a game formulation is examined for a market with agents maximizing profit as in (10.1), dropping sub-indices P throughout, using an explicit model (the analysis below remains valid for the implicit model too). Accordingly, the market-clearing relation (10.3) disappears; only a stochastic variant of (10.4) is in order. Incidentally, this is the framework considered in [9].
- The stochastic counterparts of the agents' problems are set in a two-stage framework. For example, in (10.23) the "investment" variables z_P^i are of the "here-and-now" type, to be decided before the uncertainty realizes. By contrast the "generation" variables q_P^i are of the type "wait and see": they are decided at a second stage, once ω becomes known, so q_P^i depends on ω. So, dropping the sub-index, the random vectors q^i belong to the space $L_p(\Omega, \mathscr{F}, \mathbb{P}; \mathbb{R}^{m^i})$, that is, $q(\omega) \in \mathbb{R}^{m^i}$ for all $\omega \in \Omega$, while the prices are in the dual space $L_{p^*}(\Omega, \mathscr{F}, \mathbb{P})$.

Given a price cap $PC \in L_{p^*}(\Omega, \mathscr{F}, \mathbb{P}; \mathbb{R}^{m^0})$, the complementarity formulation of stochastic equilibrium with risk aversion (considered also in [9]) is given below.

Find $\left((\bar{z}^i \in \mathbb{R}^{n^i})_{i=1}^N, (\bar{q}^i \in L_p)_{i=1}^N, \bar{q}^0 \in L_p, \bar{\pi} \in L_{p^*} \right)$ such that

$$\textbf{Risk-averse agents}\begin{cases}\min I^i(z^i)+\rho^i\Big(c^i(q^i(\omega),\omega)-\big\langle\pi(\omega),S^iq^i(\omega)\big\rangle\Big)\\ \text{s.t.}\quad (z^i,q^i(\omega))\in X^i(\omega)\qquad \text{a.e. } \omega\in\Omega,\end{cases}\tag{10.33}$$

$$\begin{aligned}\textbf{Coupling constraints}\quad & \sum_{i=1}^N S^iq^i(\omega)+q^0(\omega)=D(\omega)\qquad \text{a.e. } \omega\in\Omega\ \ (\text{mult.}\pi(\omega)),\\ \textbf{Price cap}\quad & 0\le q^0(\omega)\perp PC(\omega)-\pi(\omega)\ge 0\qquad \text{a.e. } \omega\in\Omega.\end{aligned}\tag{10.34}$$

When compared to (10.1), the agent's problem is now set as a minimization, because the risk-averse measure controls losses and not incomes. The objective function in (10.33) is in fact equivalent to the one considered in [9], taking into account that the investment functions I^i and the first-stage variables z^i are deterministic, recalling that risk measures are equivariant to translations.

10.6.3 Stochastic Variational Equilibria: Definition

Consider the following stochastic game.
Find $\tilde{\mathfrak{p}}=\Big((\tilde z^i\in\mathbb{R}^{n^i})_{i=1}^N,(\tilde q^i\in L_p)_{i=1}^N,\tilde q^0\in L_{p^*}\Big)$ solving the problems:

$$\textbf{Risk-averse agents}\begin{cases}\min I^i(z^i)+\rho^i\Big(c^i(q^i(\omega),\omega)\Big)\\ \text{s.t.}\quad (z^i,q^i(\omega))\in X^i(\omega)\qquad \text{a.e. } \omega\in\Omega,\\ \qquad S^iq^i(\omega)+\displaystyle\sum_{i\neq k=1}^N S^k\tilde q^k(\omega)+\tilde q^0(\omega)=D(\omega)\qquad \text{a.e. } \omega\in\Omega\,.\end{cases}\tag{10.35}$$

$$\textbf{Risk-averse player representing consumers}\begin{cases}\min \rho^0\Big(\big\langle PC(\omega),q^0(\omega)-D(\omega)\big\rangle\Big)\\ \text{s.t.}\quad q^0(\omega)\ge 0\qquad \text{a.e. } \omega\in\Omega,\\ \qquad \displaystyle\sum_{i=1}^N S^i\tilde q^i(\omega)+q^0(\omega)=D(\omega)\qquad \text{a.e. } \omega\in\Omega.\end{cases}\tag{10.36}$$

We define next the concept of variational equilibrium for this stochastic game. Recall that one is generally not interested in arbitrary Nash equilibria, but rather in VE defined as solutions to VIs derived from the game. In the general stochastic context like the one under consideration, instead of going via an explicit VI, we characterize VE using the Lagrange multipliers of the game coupling constraints.

Definition 1 (Stochastic VE). For a stochastic GNEP (10.35)–(10.36), the point $\bar{\mathfrak{p}}=\Big((\bar z^i)_{i=1}^N,(\bar q^i)_{i=1}^N,\bar q^0\Big)$ is a *variational equilibrium* if there exists a Lagrange multiplier $\bar\pi\in L_{p^*}$ associated to the coupling constraint,

$$\sum_{i=1}^{N} S^i q^i(\omega) + q^0(\omega) = D(\omega) \qquad \text{a.e. } \omega \in \Omega\,, \tag{10.37}$$

the same for all the players, such that $\bar{\mathfrak{p}}$ still solves the agents' problems after relaxing the coupling constraints (as in Proposition 1 below).

We now show that, under mild conditions, the concept is well defined.

Proposition 1 (Existence of Stochastic Multipliers). *For the game* (10.35)–(10.36)*, the following holds:*

1. *There exists* $\bar{\pi}^0 \in L_{p^*}(\Omega, \mathscr{F}, \mathbb{P}; \mathbb{R}^{m^0})$ *such that whenever* $\bar{q}^0$ *solves* (10.36)*, it also solves the relaxed problem*

$$\min_{q^0(\omega)\geq 0 \text{ a.e. } \omega} \rho^0\Big(\big\langle PC(\omega), q^0(\omega) - D(\omega)\big\rangle\Big) - \big\langle q^0(\omega), \bar{\pi}^0(\omega)\big\rangle_{\mathbb{P}}\,.$$

2. *Suppose for each problem* (10.35) *the functions* $I^i : \mathbb{R}^{n^i} \to \mathbb{R}$ *are smooth and convex, while* $c^i : \mathbb{R}^{m^i} \times \Omega \to \mathbb{R}$ *are random finite valued, lower semicontinuous, and convex for almost every* $\omega \in \Omega$*. Assume, in addition, that the sets* $X^i(\omega) \subset \mathbb{R}^{n^i+m^i}$ *are nonempty, closed, and convex, and some constraint qualification condition holds. If the function* $C^i : \mathbb{R}^{n^i} \times L_p(\Omega, \mathscr{F}, \mathbb{P}; \mathbb{R}^{m^i}) \to L_p(\Omega, \mathscr{F}, \mathbb{P}; \mathbb{R})$ *given by*

$$[C^i(z^i, q^i)](\omega) := c^i(z^i, q^i(\omega), \omega) \qquad \textit{is continuous and well defined,} \tag{10.38}$$

then there exists $\bar{\pi}^i \in L_{p^*}(\Omega, \mathscr{F}, \mathbb{P}; \mathbb{R}^{m^0})$ *such that whenever* $(\bar{z}^i, \bar{q}^i)$ *solves* (10.35)*, it also solves the relaxed problem*

$$\begin{cases} \min\ I^i(z^i) + \rho^i(C(z^i, q^i(\omega), \omega)) - \big\langle S^i q^i(\omega), \bar{\pi}^i(\omega)\big\rangle_{\mathbb{P}} \\ \text{s.t.}\ \ z^i \in \mathbb{R}^{n^i}, q^i \in L_p(\Omega, \mathscr{F}, \mathbb{P}; \mathbb{R}^{m^i})\,, \\ \qquad (z^i, q^i(\omega)) \in X^i(\omega) \qquad \text{a.e. } \omega \in \Omega\,. \end{cases} \tag{10.39}$$

Proof. Since the objective function in (10.36) satisfies (10.38) and a constraint qualification condition holds automatically for the feasible set, the first item is just a particular case of the second one. Accordingly, we prove the assertion for the problem

$$\begin{cases} \min\ I(z) + \rho(c(z, q(\omega), \omega)) \\ s.t.\ \ z \in \mathbb{R}^n, q \in L_p(\Omega, \mathscr{F}, \mathbb{P}; \mathbb{R}^m), \\ \qquad (z, q(\omega)) \in X(\omega) \qquad \text{a.e. } \omega \in \Omega\,, \\ \qquad Sq(\omega) = \tilde{D}_0(\omega) \qquad \text{a.e. } \omega \in \Omega\,, \end{cases} \tag{10.40}$$

corresponding to (10.35) without super-indices i (setting $\tilde{D}_0 := D_0 - \sum_{i\neq k} S^k \tilde{q}^k$). This problem is equivalent to

$$\min I(z) + \rho(C(z, q)) + \iota_{\mathscr{C}}(z, q) + \iota_{\mathbb{R}^n \times \mathscr{S}}(z, q)\,,$$

where $\iota_X(\cdot)$ denotes the indicator function of a set X (i.e., it returns zero for points in X and $+\infty$ otherwise) and where we defined the closed convex sets $\mathscr{C} := \{(z,q) \in \mathbb{R}^n \times L_p(\Omega,\mathscr{F},\mathbb{P};\mathbb{R}^m) : (z,q(\omega)) \in X(\omega) \text{ a.e. } \omega \in \Omega\}$ and $\mathscr{S} = \{q \in L_p(\Omega,\mathscr{F},\mathbb{P};\mathbb{R}^m) : Sq(\omega) = \tilde{D}_0(\omega) \text{ a.e. } \omega \in \Omega\}$. By (10.38) the objective function is well defined and, with our assumptions, it is convex. Therefore, $(\bar{z},\bar{q})$ solves the GE

$$0 \in \partial\left(I + \rho \circ C + \iota_{\mathscr{C}} + \iota_{\mathbb{R}^n \times \mathscr{S}}\right)(\bar{z},\bar{q}),$$

and the constraint qualification assumption yields that

$$\partial\left(I + \rho \circ C + \iota_{\mathscr{C}} + \iota_{\mathbb{R}^n \times \mathscr{S}}\right)(\bar{z},\bar{q}) = \partial\left(I + \rho \circ C + \iota_{\mathscr{C}}\right)(\bar{z},\bar{q}) + \partial \iota_{\mathbb{R}^n \times \mathscr{S}}(\bar{z},\bar{q}).$$

Since $\partial \iota_{\mathbb{R}^n \times \mathscr{S}}(\bar{z},\bar{q}) = \{0\} \times \mathscr{N}_{\mathscr{S}}(\bar{q})$, we have that

$$0 \in \partial\left(I + \rho \circ C + \iota_{\mathscr{C}}\right)(\bar{z},\bar{q}) + \{0\} \times \mathscr{N}_{\mathscr{S}}(\bar{q}). \tag{10.41}$$

We claim that for any $\bar{q} \in \mathscr{S}$ the normal cone is given by

$$\mathscr{N}_{\mathscr{S}}(\bar{q}) = \{\nu : \nu(\omega) = S^\top \pi(\omega) \quad \text{a.e. } \omega \in \Omega, \pi \in L_{p^*}(\Omega,\mathscr{F},\mathbb{P};\mathbb{R}^{m^0})\}.$$

The $\supset$ inclusion is straightforward. To see the converse one, first note that for any $q \in \mathscr{S}$ the identity $S(q(\omega) - \bar{q}(\omega)) = 0$ holds for a.e. $\omega \in \Omega$ (for simplicity, we omit the symbol a.e. $\omega \in \Omega$ below, noting that relations hold almost everywhere when appropriate.) Thus, $\bar{q} + \theta(q - \bar{q}) \in \mathscr{S}$ for any $\theta \in L_\infty(\Omega,\mathscr{F},\mathbb{P};\mathbb{R})$. By the definition of normal cone,

$$\langle \nu(\omega), \theta(\omega)(q(\omega) - \bar{q}(\omega))\rangle_{\mathbb{P}} \le 0,$$

and by [29, Corollary 1.9(e)], there exists $\theta \in L_\infty(\Omega,\mathscr{F},\mathbb{P};\mathbb{R})$ such that $|\theta(\omega)| = 1$ and $\langle \nu(\omega), \theta(\omega)(q(\omega) - \bar{q}(\omega))\rangle = |\langle \nu(\omega), q(\omega) - \bar{q}(\omega)\rangle|$. Therefore,

$$\int_\Omega |\langle \nu(\omega), q(\omega) - \bar{q}(\omega)\rangle| d\mathbb{P}(\omega) \le 0 \quad \Longrightarrow \quad \langle \nu(\omega), q(\omega) - \bar{q}(\omega)\rangle = 0.$$

In particular, for any $u \in \mathrm{Ker}(S)$ and $q(\omega) := \bar{q}(\omega) + u \in \mathscr{S}$, we have that $\langle \nu(\omega), u\rangle = 0$, which means that $\nu(\omega) \in [\mathrm{Ker}(S)]^\perp = \mathrm{Im}(S^\top)$.

As a result, there exists a function $\eta : \Omega \to \mathbb{R}^{m^0}$ such that $\nu(\omega) = S^\top \eta(\omega)$, and since $\nu \in L_{p^*}(\Omega,\mathscr{F},\mathbb{P};\mathbb{R}^m)$, the multiplier $\pi : \Omega \to \mathbb{R}^{m^0}$ in $L_{p^*}(\Omega,\mathscr{F},\mathbb{P};\mathbb{R}^{m^0})$ exists and is defined by $\pi(\omega) := [S^+]^\top \nu(\omega)$, where S^+ is the Moore–Penrose pseudo inverse. This establishes the claim, since

$$S^\top \pi(\omega) = S^\top [S^+]^\top \nu(\omega) = S^\top [S^+]^\top S^\top \eta(\omega) = S^\top \eta(\omega) = \nu(\omega).$$

In view of our claim, the inclusion (10.41) can be rewritten in the form

$$0 \in \partial\left(I + \rho \circ C + \iota_{\mathscr{C}}\right)(\bar{z},\bar{q}) - (0, S^\top \bar{\pi})$$

for some $\bar{\pi} \in L_{p^*}(\Omega,\mathscr{F},\mathbb{P};\mathbb{R}^{m^0})$, and the result follows. □

10.6.4 Relation Between Risk-Averse Games and MCP

We now are in a position to give the equivalent mixed complementarity counterpart of our risk-averse game.

Theorem 2 (MCP Formulation for the Risk-Averse Game). *In the setting of Proposition 1, suppose the risk-averse GNEP (10.35)–(10.36) has a variational equilibrium* $\bar{\mathfrak{p}} := \left((\bar{z}^i)_{i=1}^N, (\bar{q}^i)_{i=1}^N, \bar{q}^0\right)$*, and let* $(\bar{\mu}^i)_{i=1}^N$ *and* $\bar{\pi}$ *denote the respective* L_{p^*}*-multipliers for the endogenous constraints in (10.35) and the coupling constraints (10.37).*

Then the primal–dual pair $(\bar{\mathfrak{p}}, \bar{\mathfrak{d}})$ *with* $\bar{\mathfrak{d}} := (\bar{\mu}, \bar{\pi})$ *solves the risk-averse MCP derived from the following problems:*

$$\textbf{Risk-averse agents} \quad \begin{cases} \min I^i(z^i) + \rho^i\Big(c^i(q^i(\omega), \omega)\Big) - \big\langle \pi, S^i q^i \big\rangle_{\mathbb{P}} \\ \text{s.t.} \quad (z^i, q^i(\omega)) \in X^i(\omega) \qquad \text{a.e. } \omega \in \Omega, \end{cases} \tag{10.42}$$

Coupling constraints *as in (10.37)*

$$\textbf{Risk-averse price cap} \quad 0 \le q^0(\omega) \perp x_0^*(\omega) PC(\omega) - \pi(\omega) \ge 0 \qquad \text{a.e. } \omega \in \Omega$$

for x_0^* *solving*

$$\begin{cases} \min \displaystyle\sum_{i=1}^N \Big\langle \big\langle x_0^*(\omega) PC(\omega), S^i \bar{q}^i(\omega) \big\rangle \Big\rangle_{\mathbb{P}} \\ \text{s.t.} \quad \mathbb{E}(x_0^*) = 1 \\ \qquad 1 - \kappa_0 \le x_0^*(\omega) \le 1 - \kappa_0 + \dfrac{\kappa_0}{\varepsilon_0} \qquad \text{a.e. } \omega \in \Omega. \end{cases} \tag{10.43}$$

Proof. To derive a complementarity formulation, we first consider (10.35). By Proposition 1, $(\bar{z}^i, \bar{q}^i)$ solves the relaxed problem (10.39) with $\bar{\pi} = \bar{\pi}^i$, by Definition 1. Since the optimality conditions of the relaxed problem coincide with those of problem (10.42), the stated result for the agents follows.

In the case of consumers' representative, by Proposition 1 and Definition 1, $\bar{q}^0$ solves problem (10.36) as well as the relaxed problem

$$\min_{q^0 \ge 0} \rho^0\Big(\big\langle PC(\omega), q^0(\omega) - D(\omega) \big\rangle\Big) - \Big\langle \big\langle q^0(\omega), \bar{\pi}(\omega) \big\rangle \Big\rangle_{\mathbb{P}}.$$

We now show that the optimality conditions of the relaxed problem coincide with those of (10.43), together with the risk-averse price-cap condition. Since $PC \in L_{p^*}$ and $D \in L_p$, the affine operator $A : L_p \to L_p$ defined by

$$[A(q^0)](\omega) := \big\langle PC(\omega), q^0(\omega) - D(\omega) \big\rangle$$

is continuous and, hence, the optimality condition for the relaxed problem is

$$0 \in \partial\big(\rho^0 \circ A + i_{\ge 0}\big)(\bar{q}^0) - \bar{\pi} = \partial(\rho^0 \circ A)(\bar{q}^0) + \mathcal{N}_{\ge 0}(\bar{q}^0) - \bar{\pi}.$$

By the normal cone definition, there exist $g \in \partial(\rho^0 \circ A)(\bar{q}^0)$ and $\nu \in L_{p^*}$ such that

$$0 \le \bar{q}^0(\omega) \perp -\bar{\nu}(\omega) \ge 0 \quad \text{and} \quad 0 = g(\omega) + \bar{\nu}(\omega) - \bar{\pi}(\omega),$$

almost everywhere in Ω. To get an explicit expression for g above, we apply [35, Theorem 2.83] to compute the subdifferential $\partial(\rho^0 \circ A)(\bar{q}^0)$, recalling that the mapping A is affine and continuous, and the risk measure is increasing and finite-valued:

$$g \in \partial(\rho^0 \circ A)(\bar{q}^0) \iff g(\omega) = PC(\omega)s(\omega) \quad \text{for } s \in \partial\rho^0\Big(A(\bar{q}^0)\Big).$$

The definitions of the subdifferential and of the conjugate function give the equivalence $s \in \partial\rho^0(A(\bar{q}^0)) \iff A(\bar{q}^0) \in \partial\rho^{0^*}(s)$. By the dual representation (10.32), the conjugate of ρ^0 is the indicator function of the (convex and bounded) dual set X^*, that is, $\rho^0 = \iota^*_{X^*}$. Then, $\rho^{0^*} = \iota^{**}_{X^*} = i_{X^*}$. Since the subdifferential of the indicator function of a closed convex set is the normal cone of the set, by the definition of the normal cone, the subgradient $g \in \partial(\rho^0 \circ A)(\bar{q}^0)$ has components $g(\omega) = PC(\omega)s(\omega)$ for $s \in X^*$ satisfying $\langle A(\bar{q}^0), x^* - s\rangle_{\mathbb{P}} \le 0$ for all $x^* \in X^*$. So s maximizes $\langle A(\bar{q}^0), x^*\rangle_{\mathbb{P}}$ over X^*, and in view of (10.37), $s = \bar{x}^*_0$ from (10.43). The risk-averse price-cap condition follows from plugging $g(\omega) = PC(\omega)\bar{x}^*_0(\omega)$ in the optimality condition. □

Theorem 2 shows that, like in the deterministic framework, the stochastic game is equivalent to a complementarity model with risk aversion. Nevertheless, the stochastic MCP model is *not of the form* (10.33), where agents *hedge individually their profit*. Instead, a VE for the game (10.35)–(10.37) gives a stochastic equilibrium for a market that is cleared because (10.37) is satisfied and where the risk-averse agents are *remunerated in mean at a price that is controlled by a risk-averse price cap*.

In the game, aversion to risk is peculiar in the sense that agents hedge volatility by controlling only variations in the generation costs. In the game problem (10.42), the remuneration is taken in mean without hedging risk, while in the MCP (10.33)–(10.34) each agent tries to control the risk in their individual revenue. In the game the control of volatile prices is "delegated" to some higher instance. This is the same instance that caps the remunerations, only that now the cap is chosen adaptively, in a manner that is optimal for the market, in the sense of (10.43). By contrast, in the risk-averse MCP, the instance limiting prices only takes into account stochasticity but does not perceive the fact of capping prices as a risky action, perturbing the market.

Our final result shows that the three models become equivalent in a risk-neutral market.

Corollary 3 (Equivalence for Risk-Neutral Agents). *Suppose that for all the agents $\rho^i = \mathbb{E}$, the expected-value function. Then finding a variational equilibrium for the GNEP* (10.35)*,* (10.36)–(10.37) *is equivalent to solving the MCP* (10.33)–(10.34) *which is in turn equivalent to the MCP* (10.42)–(10.43)*.*

Proof. Straightforward from Theorem 2, as the expected-value function is recovered by setting $\kappa_i = 0$ in (10.31), with the singleton dual set $X^* = \{x^* \equiv 1\}$ in (10.32). In particular, a risk-neutral representative of the consumers can only take $\bar{x}^*_0 \equiv 1$, which yields the stochastic price cap from (10.34). The equivalence with the last MCP results from the linearity of the expected-value function. □

Concluding Remarks

Like it has been done in the deterministic case in Sect. 10.5.1, it would be interesting to analyze and compare the performances of the risk-averse game versus the risk-averse MCP on a numerical example. However, due to the positive-part function in (10.31), risk measures are not differentiable and for both models the GE mapping has multi-valued components. In this context, a direct application of a solver like PATH is no longer possible (and there is currently no other established software that can do the job). In [9], the MCP (10.33)–(10.34) is "solved" ignoring nondifferentiability issues and treating the mapping as if it were single-valued. This heuristic seems to produce sound results for the considered example, but cannot be regarded as a reliable solution method, of course. In order to handle nonsmoothness, some special technique should be used, for example, the approximation procedure in [21].

Finally, instead of handling uncertainty in two stages, a multistage setting can also be of interest. This, keeping in mind that multistage risk-averse models remain a delicate subject and involves intricate issues such as time consistency and information monotonicity; see [23]. Last but not least, and as discussed in [32, Sect. 5], risk-averse variants of sampling approaches like [25, 33] lack implementable stopping criteria. Multistage risk-averse models present numerous challenges already in an optimization framework; we refer to [30] and references therein for more details.

Acknowledgements Research of the second author is partially supported by grants CNPq 303840/2011-0, AFOSR FA9550-08-1-0370, and NSF DMS 0707205, as well as by PRONEX-Optimization and FAPERJ. The third author is partially supported by CNPq grant 302637/2011-7, by PRONEX-Optimization, and by FAPERJ.

References

1. Baldick R, Helman U, Hobbs B, O'Neill R (2005) Design of efficient generation markets. Proc IEEE 93(11):1998–2012
2. Barroso L, Carneiro R, Granville S, Pereira M, Fampa M (2006) Nash equilibrium in strategic bidding: a binary expansion approach. IEEE Trans Power Syst 21(2):629–638
3. Chung W, Fuller JD (2010) Subproblem approximation in Dantzig-Wolfe decomposition of variational inequality models with an application to a multicommodity economic equilibrium model. Oper Res 58:1318–1327
4. Conejo AJ, Carrión M, Morales JM (2010) Decision making under uncertainty in electricity markets. International series in operations research and management science. Springer, New York
5. David A, Wen F (2000) Strategic bidding in competitive electricity markets: a literature survey. In: Power engineering society summer meeting, 2000, vol 4. IEEE, Seattle, pp 2168–2173
6. Dentcheva D, Ruszczyński A, Shapiro A (2009) Lectures on stochastic programming. SIAM, Philadelphia

7. Dirkse SP, Ferris MC (1995) The PATH solver: a nommonotone stabilization scheme for mixed complementarity problems. Optim Methods and Software 5(2):123–156
8. Ehrenmann A, Neuhoff K (2009) A comparison of electricity market designs in networks. Oper Res 57(2):274–286
9. Ehrenmann A, Smeers Y (2011) Generation capacity expansion in risky environment: a stochastic equilibrium analysis. Oper Res 59(6):1332–1346
10. Facchinei F, Kanzow C (2010) Generalized Nash equilibrium problems. Ann Oper Res 175(1):177–211
11. Facchinei F, Pang JS (2003) Finite-dimensional variational inequalities and complementarity problems. Springer series in operations research, vol I. Springer, New York
12. Ferris MC, Munson TS (1999) Interfaces to PATH 3.0: design, implementation and usage. Comput Optim Appl 12:207–227
13. Fuller JD, Chung W (2005) Dantzig-Wolfe decomposition of variational inequalities. Comput Econ 25:303–326
14. Gabriel SA, Zhuang J, Egging R (2009) Solving stochastic complementarity problems in energy market modeling using scenario reduction. European J Oper Res 197(3):1028–1040
15. Harker PT, Pang JS (1990) Finite-dimensional variational inequality and nonlinear complementarity problems: a survey of theory, algorithms and applications. Math Program 48(2, Ser. B):161–220
16. Hiriart-Urruty JB, Lemaréchal C (1993) Convex analysis and minimization algorithms. No. 305–306 in Grund. der math. Wiss. Springer, New York
17. Hobbs BF, Pang JS (2004) Spatial oligopolistic equilibria with arbitrage, shared resources, and price function conjectures. Math Program 101:57–94
18. Hobbs B, Metzler C, Pang JS (2000) Strategic gaming analysis for electric power systems: an MPEC approach. IEEE Trans Power Syst 15(2):638–645
19. Hu X, Ralph D (2007) Using EPECs to model bilevel games in restructured electricity markets with locational prices. Oper Res 55(5):809–827
20. Luna JP, Sagastizábal C, Solodov M A class of Dantzig-Wolfe type decomposition methods for variational inequality problems. Math Program, Online first. doi:10.1007/s10107-012-0599-7
21. Luna JP, Sagastizábal C, Solodov M (2013) An approximation scheme for a class of risk-averse stochastic equilibrium problems. Optimization Online report, http://www.optimization-online.org/DB_HTML/2013/07/3941.html
22. Marcato R, Sagastizábal C (2007) Introducing environmental constraints in generation expansion problems. Numer Linear Algebra Appl 14:351–368

23. Pflug G, Pichler A (2011) On dynamic decomposition of multistage stochastic programs. Technical report, Optimization Online. http://www.optimization-online.org/DB_HTML/2011/11/3254.html
24. Pflug GC, Römisch W (2007) Modeling, measuring and managing risk. World Scientific, London
25. Philpott A, de Matos VL (2012) Dynamic sampling algorithms for multi-stage stochastic programs with risk aversion. European J Oper Res 218(2):470–483
26. Ralph D, Smeers Y (2006) EPECs as models for electricity markets. In: Power systems conference and exposition, 2006. PSCE '06, vol 2006. IEEE PES, Atlanta, pp 74–80
27. Robinson SM (1998) A reduction method for variational inequalities. Math Program 80(2, Ser. A):161–169
28. Rockafellar R, Uryasev S (2002) Conditional value-at-risk for general loss distributions. J Banking Financ 26(7):1443–1471
29. Rudin W (1987) Real and complex analysis, 3rd edn. McGraw-Hill Book Co., New York
30. Sagastizábal C (2012) Divide to conquer: decomposition methods for energy optimization. Math Program 134:187–222
31. Sagastizábal C, Solodov M (2012) Solving generation expansion planning problems with environmental constraints by a bundle method. Comput Manag Sci 9:163–182
32. Shapiro A (2011) Analysis of stochastic dual dynamic programming method. European J Oper Res 209(1):63–72
33. Shapiro A, Tekaya W (2011) Report for technical cooperation between Georgia Institute of Technology and ONS - Operador Nacional do Sistema Elétrico. Report 2: risk averse approach
34. Yin H, Shanbhag U, Mehta P (2011) Nash equilibrium problems with scaled congestion costs and shared constraints. IEEE Trans Automat Control 56(7):1702–1708
35. Zălinescu C (2002) Convex analysis in general vector spaces. World Scientific Publishing Co. Inc., River Edge

Chapter 11
Optimal Planning and Economic Evaluation of Trigeneration Districts

Maria Teresa Vespucci, Stefano Zigrino, Francesca Bazzocchi, and Alberto Gelmini

Abstract Trigeneration, or combined cooling, heat and power (*CCHP*), is the process by which electricity, heating and cooling are simultaneously generated from the combustion of a fuel. Trigeneration systems for serving the electricity, heating and cooling loads in residential districts are a possible solution to enhance energy efficiency, reduce fossil fuel consumption and increase the use of renewable energy sources in the residential sector. Technical, economical and financial issues have to be taken into account when planning a trigeneration system or when expanding an existing generation system. In this chapter a two-step decision support procedure is presented for analysing alternative system configurations. The first step is based on a mixed integer linear programming model that allows to describe the system components in great detail and computes the annual optimal dispatch of the distributed generation system with a hourly discretization, taking into account load profiles, fuel costs and technical constraints. The optimal dispatch is then used for the economic evaluation of the investment, taking into account prices of commodities, taxation, incentives and financial aspects. The procedure allows to compare alternative plant configurations and can be used as a simulation tool, for assessing the system sensitivity to variations of model parameters (e.g. incentives and ratio debt/equity).

11.1 Introduction

Trigeneration systems for serving the electricity, thermal and cooling loads in residential districts are a possible solution to enhance energy efficiency and to reduce fossil fuel consumption. They may include different kinds of generators

M.T. Vespucci (✉) • S. Zigrino
Department of Management, Economics and Quantitative Methods, University of Bergamo, Italy
e-mail: maria-teresa.vespucci@unibg.it; stefano.zigrino@unibg.it

F. Bazzocchi • A. Gelmini
Ricerca sul Sistema Energetico (RSE) SpA, Milano, Italy
e-mail: Francesca.Bazzocchi@rse-web.it; Alberto.Gelmini@rse-web.it

R.M. Kovacevic et al. (eds.), *Handbook of Risk Management in Energy Production and Trading*, International Series in Operations Research & Management Science 199,
DOI 10.1007/978-1-4614-9035-7_11,

(e.g. cogeneration units, boilers, electric heat pumps, gas absorption heat pumps, absorption chillers), as well as hot storages and ice storages. In particular, combined heat and power (*CHP*) plants, or cogeneration units, allow to recover the combustion heat generated during the production of electricity, obtaining a saving of primal energy, a decrease in production costs of the total energy required and a reduction of CO_2 emissions. The systems are connected to the national electric grid, in order to purchase and sell electricity when needed. Trigeneration systems also represent a chance to increase the use of renewable energy sources in the residential sector: indeed cogeneration units can benefit from the Green Certificate for the produced electricity, if biofuels are used instead of fossil fuel; they are moreover supported with the energy efficiency certificate (White Certificate), if they are qualified as high-efficiency systems.

In this chapter a decision support procedure is introduced for the configuration of distributed generation systems in residential districts, where various types of energy demands (electrical load, thermal loads at various temperatures, cooling load) have to be served. In the configuration process alternative solutions have to be compared, both from a technical and an economical point of view, taking into account the energy consumption profiles that vary along the day and along the year, due to the weather conditions. The decision support procedure consists of two steps, see [1] and [5]. In the first step, by solving the mixed integer linear programming model introduced in Sect. 11.2, the annual optimal dispatch of the distributed generation system is determined with an hourly discretization, taking into account technical constraints, load profiles and fuel costs. The optimal dispatch is then used for the economic evaluation of the investment, taking into account prices of commodities, taxation and financial aspects. Starting from *EBITDA* (Earnings Before Interest, Taxes, Depreciation and Amortization) and from investment and financial parameters, the net present value (*NPV*) of the investment, the payback time (*PBT*) and the internal rate of return (*IRR*) (see [6]) are computed. The decision support procedure allows to compare alternative plant configurations; it can also be used as a simulation tool, for assessing the system sensitivity to variations of parameter values. The decision support system is available as a web application (called *GDPint*) and can be freely accessed at `www.rds-web.it`. An interface allows the user to define the characteristics of the system components, the load profiles and the prices of the commodities.

A similar application is the "Distributed Energy Resources Customer Adoption Model" (*DER-CAM*) (see [10]), an economic and environmental model of customer-distributed energy resources adoption. The objective of this model is to minimize the cost of operating on-site generation and combined heat and power systems, either for individual customer sites or a micro-grid. To this aim *DER-CAM* addresses the following issues:

1. What is the lowest cost combination of distributed generation technologies that a specific customer can install?
2. What is the level of installed capacity of these technologies that minimizes cost?
3. How should the installed capacity be operated so as to minimize the total customer energy bill?

It is assumed that the customer wishes to install distributed generation so as to minimize the cost of the energy consumed on site. Consequently, it is possible to determine the technologies and the capacity the customer is likely to install and to predict when the customer will be self-generating electricity and/or transacting with the power grid and, likewise, when he will be purchasing fuel or using recovered heat. *DER-CAM* does not allow the user to describe the actual components of the system to be evaluated, as it requires the user to choose the system elements from a given database. *GDPint* instead allows the user either to select the components from a database or to define the characteristics of the actual elements in the system: for example, for cogeneration units the actual minimum uptime and downtime can be taken into account and efficiency can be defined as a function of load and air temperature. Also the economic evaluation is more detailed in *GDPint* and allows taking into account how the investment is financed (e.g. the debt/equity ratio and incentives).

In software *DCogEN* [2, 3, 4] the evaluation of cogeneration districts is based on a much simpler system optimization model, as only one single period is considered at a time; as a consequence, however, intertemporal constraints for modelling the energy levels of electric storages and the minimum uptime and downtime constraints of cogeneration units cannot be included in the system optimization model.

The chapter is organized as follows. In Sect. 11.2 the mixed integer linear programming model is introduced for determining the hourly dispatch of a distributed generation system that minimizes the total generation cost over the time horizon. In Sect. 11.3 the heuristic procedure is described for approximating the optimal solution of large dimensional instances. In Sect. 11.4 a case study is discussed in which decisions on investment in a trigeneration system for a residential district are supported by an extensive analysis, both from a technical and a financial point of view, of five different configurations; in Sect. 11.5 references to investment problems analysed by *GDPint* are given and future work is outlined.

11.2 The Model for the Optimal Hourly Dispatch of a Trigeneration System

In this section we introduce the mixed integer linear programming model for determining the annual optimal dispatch, with an hourly discretization, of all the resources in a trigeneration system. The economic optimization of the power dispatch takes into account the technical constraints of the system components, the time profiles of the loads and the prices of fuel and electricity. The trigeneration system may include different kind of generators (cogeneration units, boilers, electric heat pumps, gas absorption heat pumps, absorption chillers, etc.) for serving different kinds of loads (electrical load, thermal loads at various temperatures, cooling load). A set of binary parameters describes the system topology, i.e. the power flows from generators to storages and loads. The thermal loads may be served by different generators, such as boilers, electric heat pumps, gas absorption heat pumps, cogeneration units

and hot storages. The cooling load may be served by absorption chillers, reversible electric heat pumps, reversible gas absorption heat pumps and ice storages. The electrical load, which includes the electricity used by heat pumps and absorption chillers, may be generated by cogeneration units and/or purchased on the market, the system being connected to the national grid. Excess electricity production can be sold on the market.

A detailed model of each system component is considered. Each absorption chiller is characterized by capacity, electric coefficient of performance and thermal coefficient of performance. Each boiler is characterized by maximum heating rate and fuel consumption function. Each electric heat pump is characterized by heating capacity, cooling capacity (if reversible), energy efficiency ratio and coefficient of performance, both dependent on the air temperature. Each gas absorption heat pump is characterized by heating capacity, cooling capacity (if reversible), electric coefficient of performance and fuel consumption rate. Each cogeneration unit is characterized by its minimum and maximum electrical power outputs, minimum uptime and downtime, heat recovery function and fuel consumption function. For all generators operation and maintenance cost per output unit are given. Hot storages and ice storages are characterized by maximum stored energy, energy rate from source to load and loss coefficients. The hourly dispatch is computed so as to minimize the total costs minus the revenues from the sale of electricity to the grid.

The notation of the proposed model is provided below for quick reference.

11.2.1 Sets

H	Set of temperature levels of thermal loads, indexed by h
R	Set of absorption chillers, indexed by r
B	Set of boilers, indexed by b
P	Set of electric heat pumps, indexed by p
G	Set of gas absorption heat pumps, indexed by g
M	Set of cogeneration units, indexed by m
K	Set of ice storages, indexed by k
J	Set of hot storages, indexed by j
T	Set of hours, indexed by t

11.2.2 Parameters

- For absorption chiller $r \in R$:

C_r^O	Operation and maintenance cost per unit of thermal output
$\overline{\dot{Q}}_r^C$	Capacity
U_r^E	Electric coefficient of performance

$U^H_{r,h}$ Thermal coefficient of performance
$\alpha_{r,h}$ Binary parameter set to 1 if absorption chiller r can use h-temperature thermal power

- For boiler $b \in B$:

C^O_b Operation and maintenance cost per unit of thermal output
C^F_b Fuel specific cost
$\overline{\dot{Q}}^H_b$ Maximum heating rate
F_b Fuel consumption per unit of thermal output
$\alpha_{b,h}$ Binary parameter set to 1 if boiler b generates h-temperature thermal power

- For electric heat pump $p \in P$:

C^O_p Operation and maintenance cost per unit of thermal output
$\overline{\dot{Q}}^H_p$ Heating capacity
$\overline{\dot{Q}}^C_p$ Cooling capacity (if electric heat pump is reversible)
$EER_{p,t}$ Energy efficiency ratio (depends on the air temperature in hour t)
$COP_{p,t}$ Coefficient of performance (depends on the air temperature in hour t)
ρ_p Binary parameter set to 1 if electric heat pump p can generate cold
$\alpha_{p,h}$ Binary parameter set to 1 if electric heat pump p generates h-temperature thermal power

- For gas absorption heat pump $g \in G$:

C^O_g Operating and maintenance cost per unit of thermal output
C^F_g Fuel-specific cost
F^C_g Fuel consumption per unit of cooling output
F^H_g Fuel consumption per unit of thermal output
U^E_g Electric coefficient of performance
$\overline{\dot{Q}}^H_g$ Heating capacity
$\overline{\dot{Q}}^C_g$ Cooling capacity (if gas absorption heat pump is reversible)
ρ_g Binary parameter set to 1 if gas absorption heat pump g can generate cold
$\alpha_{g,h}$ Binary parameter set to 1 if gas absorption heat pump g generates h-temperature thermal power

- For cogeneration unit $m \in M$:

C^O_m Operation and maintenance cost per unit of power output
C^F_m Fuel-specific cost
$\underline{W}_m$ Minimum power
$\overline{W}_m$ Maximum power
t^U_m Minimum uptime

t_m^D Minimum downtime
$q_m^{(1)}$ Slope of the heat recovery function
$q_m^{(0)}$ Intercept of the heat recovery function
$F_m^{(1)}$ Slope of the fuel consumption function
$F_{m,t}^{(0)}$ Intercept of the fuel consumption function
$\alpha_{m,h}$ Binary parameter set to 1 if cogeneration unit m generates h-temperature thermal power

- For ice storage $k \in K$:

$\overline{Q}_k^C$ Maximum stored energy
$Q_{k,0}^C$ Energy stored at the beginning of the first hour
$\overline{\dot{Q}}_i^{C,in}$ Energy rate from source
$\overline{\dot{Q}}_k^{C,out}$ Energy rate to load
$l_k^{C,in}$ Input loss coefficient ($0 \leq l_k^{C,in} \leq 1$)
$l_k^{C,out}$ Output loss coefficient ($l_k^{C,out} \geq 1$)
l_k^C Tank loss coefficient ($0 \leq l_k^C \leq 1$)

- For hot storage $j \in J$:

$\overline{Q}_j^H$ Maximum stored energy
$Q_{j,0}^H$ Energy stored at the beginning of the first hour
$\overline{\dot{Q}}_j^{H,in}$ Maximum thermal input
$\overline{\dot{Q}}_j^{H,out}$ Maximum thermal output
$l_j^{H,in}$ Input loss coefficient ($0 \leq l_j^{H,in} \leq 1$)
$l_j^{H,out}$ Output loss coefficient ($l_j^{H,out} \geq 1$)
l_j^H Tank loss coefficient ($0 \leq l_j^H \leq 1$)
$\alpha_{j,h}^{in}$ Binary parameter set to 1 if the input of hot storage j is h-temperature thermal power
$\alpha_{j,h}^{out}$ Binary parameter set to 1 if the output of hot storage j is h-temperature thermal power

- Loads, outputs of non-dispatchable power plants and electricity prices in hour $t \in T$:

L_t^E Electrical load
L_t^C Cooling load
$L_{h,t}^H$ h-temperature thermal load
L_t^P Photovoltaic production
L_t^S Solar thermal production
α_h^S Binary parameter set to 1 if the solar thermal plant generates h-temperature thermal power
μ_t Purchase price of electricity
λ_t Sale price of electricity

- Data related to the electricity market:

$\overline{W}^A$	Maximum power that can be purchased from the grid
$\overline{W}^V$	Maximum power that can be sold to the grid

Finally, let $\hat{T}$ denote the subset of hours in which the electricity purchase price μ_t is less than the electricity sale price λ_t: for these hours constraints are introduced in the model so as to avoid arbitrage.

11.2.3 Decision Variables

The following symbols denote the decision variables pertaining to hour $t \in T$:

- For absorption chiller $r \in R$:

$\dot{Q}^C_{r,t}$	Cooling power production of absorption chiller r

- For boiler $b \in B$:

$\dot{Q}^H_{b,h,t}$	h-temperature thermal power production of boiler b

- For electric heat pump $p \in P$:

$\dot{Q}^C_{p,t}$	Cooling power production of electric heat pump p
$\dot{Q}^H_{p,h,t}$	h-temperature thermal power production of electric heat pump p

- For gas absorption heat pump $g \in G$:

$\dot{Q}^C_{g,t}$	Cooling power production of gas absorption heat pump g
$\dot{Q}^H_{g,h,t}$	h-temperature thermal power production of gas absorption heat pump g

- For cogeneration unit $m \in M$:

$\dot{Q}^H_{m,h,t}$	h-temperature thermal power production of cogeneration unit m
$W_{m,t}$	Power output of cogeneration unit m
$\gamma_{m,t}$	Status of cogeneration unit m (*on*, if $\gamma_{m,t} = 1$; *off*, if $\gamma_{m,t} = 0$)

- For ice storage $k \in K$:

$Q^C_{k,t}$	Energy stored in ice storage k at the end of the hour
$\dot{Q}^{C,in}_{k,t}$	Energy rate from source of ice storage k
$\dot{Q}^{C,out}_{k,t}$	Energy rate to load of ice storage k

- For hot storage $j \in J$:

$Q^H_{j,t}$	Energy stored in hot storage j at the end of the hour
$\dot{Q}^{H,in}_{j,h,t}$	h-temperature power output of hot storage j
$\dot{Q}^{H,out}_{j,h,t}$	h-temperature power input of hot storage j

- Variables related to exchanges on the electricity market in hour $t \in T$:

W_t^A	Electricity purchased on the market
W_t^V	Electricity sold on the market
θ_t	Binary variable used in "no-arbitrage" constraints (11.4) and (11.5)

11.2.4 The Mixed Integer Linear Programming Model

The optimal scheduling of the trigeneration plant is determined by solving the mixed integer linear programming model

$$\begin{aligned}\min \sum_{t\in T}\Bigg\{ & \sum_{r\in R}\left(C_r^O \dot{Q}_{r,t}^C\right)+\sum_{b\in B}\left[\left(C_b^F F_b + C_b^O\right)\left(\sum_{h\in H}\dot{Q}_{b,h,t}^H\right)\right]+ \\ & +\sum_{p\in P}\left[C_p^O\left(\dot{Q}_{p,t}^C+\sum_{h\in H}\dot{Q}_{p,h,t}^H\right)\right]+ \\ & +\sum_{g\in G}\left[\left(C_g^F F_g^C + C_g^O\right)\dot{Q}_{g,t}^C+\left(C_g^F F_g^H + C_g^O\right)\left(\sum_{h\in H}\dot{Q}_{g,h,t}^H\right)\right]+ \\ & +\sum_{m\in M}\left[C_m^F\left(F_{m,t}^{(0)}\gamma_{m,t}+F_m^{(1)}W_{m,t}\right)+C_m^O W_{m,t}\right]+\mu_t W_t^A-\lambda_t W_t^V\Bigg\}\end{aligned} \tag{11.1}$$

subject to the following constraints:

- For $t \in T$

$$\begin{aligned}\sum_{m\in M} W_{m,t}+L_t^P+W_t^A = L_t^E+\sum_{r\in R}U_r^E\dot{Q}_{r,t}^C+\sum_{p\in P}\left(\frac{\dot{Q}_{p,t}^C}{EER_{p,t}}+\frac{\sum_{h\in H}\dot{Q}_{g,h,t}^H}{COP_{p,t}}\right)+ \\ +\sum_{g\in G}\left[U_g^E\left(\dot{Q}_{g,t}^C+\sum_{h\in H}\dot{Q}_{p,h,t}^H\right)\right]+W_t^V\end{aligned} \tag{11.2}$$

- For $t \in \hat{T}$

$$\theta_t \in \{0,1\} \tag{11.3}$$

$$0 \le W_t^A \le \overline{W}^A(1-\theta_t) \tag{11.4}$$

$$0 \le W_t^V \le \overline{W}^V\theta_t \tag{11.5}$$

- For $t \in T$

$$\sum_{r\in R}\dot{Q}_{r,t}^C+\sum_{p\in P}\rho_p\dot{Q}_{p,t}^C+\sum_{g\in G}\rho_g\dot{Q}_{g,t}^C+\sum_{i\in I}\dot{Q}_{i,t}^{C,out}\ge L_t^C+\sum_{i\in I}\dot{Q}_{i,t}^{C,in} \tag{11.6}$$

- For $h \in H$ and $t \in T$

$$\alpha_h^S L_t^S + \sum_{b\in B} \alpha_{b,h} \dot{Q}_{b,h,t}^H + \sum_{p\in P} \alpha_{p,h} \dot{Q}_{p,h,t}^H + \sum_{g\in G} \alpha_{g,h} \dot{Q}_{g,h,t}^H + \sum_{m\in M} \alpha_{m,h} \dot{Q}_{m,h,t}^H + \\ + \sum_{j\in J} \alpha_{j,h}^{out} \dot{Q}_{j,h,t}^{H,out} \geq L_{h,t}^H + \sum_{r\in R} \alpha_{r,h} U_{r,h}^H \dot{Q}_{r,t}^C + \sum_{j\in J} \alpha_{j,h}^{in} \dot{Q}_{j,h,t}^{H,in} \tag{11.7}$$

- For $r \in R$ and $t \in T$

$$0 \leq \dot{Q}_{r,t}^C \leq \overline{\dot{Q}}_r^C \tag{11.8}$$

- For $b \in B$ and $t \in T$

$$0 \leq \sum_{h\in H} \dot{Q}_{b,h,t}^H \leq \overline{\dot{Q}}_b^H \tag{11.9}$$

- For $p \in P$ and $t \in T$

$$0 \leq \dot{Q}_{p,t}^C \leq \overline{\dot{Q}}_p^C \tag{11.10}$$

$$0 \leq \sum_{h\in H} \dot{Q}_{p,h,t}^H \leq \overline{\dot{Q}}_p^H \tag{11.11}$$

$$\frac{\dot{Q}_{p,t}^C}{\overline{\dot{Q}}_p^C} + \frac{\sum_{h\in H} \dot{Q}_{p,h,t}^H}{\overline{\dot{Q}}_p^H} \leq 1 \tag{11.12}$$

- For $g \in G$ and $t \in T$

$$0 \leq \dot{Q}_{g,t}^C \leq \overline{\dot{Q}}_g^C \tag{11.13}$$

$$0 \leq \sum_{h\in H} \dot{Q}_{g,h,t}^H \leq \overline{\dot{Q}}_g^H \tag{11.14}$$

$$\frac{\dot{Q}_{g,t}^C}{\overline{\dot{Q}}_g^C} + \frac{\sum_{h\in H} \dot{Q}_{g,h,t}^H}{\overline{\dot{Q}}_g^H} \leq 1 \tag{11.15}$$

- For $m \in M$ and $t \in T$

$$\gamma_{m,t} \in \{0,1\} \tag{11.16}$$

$$\sum_{\tau=t+1}^{\min(t+t_m^U-1,|T|)} \gamma_{m,\tau} \geq \min\left(t_m^U - 1, |T| - t\right)(\gamma_{m,t} - \gamma_{m,t-1}) \tag{11.17}$$

$$\sum_{\tau=t+1}^{\min(t+t_m^D-1,|T|)} (1-\gamma_{m,\tau}) \geq \min\left(t_m^D - 1, |T| - t\right)(\gamma_{m,t-1} - \gamma_{m,t}) \tag{11.18}$$

$$\underline{W}_m \gamma_{m,t} \leq W_{m,t} \leq \overline{W}_m \gamma_{m,t} \tag{11.19}$$

$$0 \leq \sum_{h\in H} \dot{Q}_{m,h,t}^H \leq q_m^{(0)} \gamma_{m,t} + q_m^{(1)} W_{m,t} \tag{11.20}$$

- For $k \in K$ and $t \in T$

$$Q_{k,t}^C = \left(1 - l_k^C\right) Q_{k,t-1}^C + l_k^{C,in} \dot{Q}_{k,t}^{C,in} - l_k^{C,out} \dot{Q}_{k,t}^{C,out} \tag{11.21}$$

$$0 \leq Q_{k,t}^{C} \leq \overline{Q}_{k}^{C} \tag{11.22}$$

$$0 \leq \dot{Q}_{k,t}^{C,in} \leq \overline{\dot{Q}}_{k}^{C,in} \tag{11.23}$$

$$0 \leq \dot{Q}_{k,t}^{C,out} \leq \overline{\dot{Q}}_{k}^{C,out} \tag{11.24}$$

- For $j \in J$ and $t \in T$

$$Q_{j,t}^{H} = \left(1 - l_{j}^{H}\right) Q_{j,t-1}^{H} + l_{j}^{H,in} \sum_{h \in H} \alpha_{j,h}^{in} \dot{Q}_{j,h,t}^{H,in} - l_{j}^{H,out} \sum_{h \in H} \alpha_{j,h}^{out} \dot{Q}_{j,h,t}^{H,out} \tag{11.25}$$

$$0 \leq Q_{j,t}^{H} \leq \overline{Q}_{j}^{H} \tag{11.26}$$

$$0 \leq \sum_{h \in H} \alpha_{j,h}^{in} \dot{Q}_{j,h,t}^{H,in} \leq \overline{\dot{Q}}_{j}^{H,in} \tag{11.27}$$

$$0 \leq \sum_{h \in H} \alpha_{j,h}^{out} \dot{Q}_{j,h,t}^{H,out} \leq \overline{\dot{Q}}_{j}^{H,out} \tag{11.28}$$

The objective function (11.1), to be minimized, represents the total cost, in a typical year, for satisfying the electrical, cooling and thermal loads, minus the revenues from the sold electricity. The net costs related to hour t consist of seven terms: the first term represents the operation and maintenance costs of the absorption chillers; the second term expresses the fuel cost and the operation and maintenance costs of the boilers; the third term represents the operation and maintenance costs for generating cooling and thermal power by electric heat pumps; the fourth term represents the fuel costs and the operation and maintenance costs of the gas absorption heat pumps; the fifth term expresses the fuel costs and the operation and maintenance costs of the cogeneration units; the sixth term represents the cost for purchasing electricity from the market; and the seventh term represents the revenues from electricity sold into the market. In the fifth term the fuel-specific cost C_{m}^{F} of cogeneration unit m is multiplied by the fuel consumption in hour t, given by the affine function F_{m}^{MI} of power output $W_{m,t}$

$$F_{m}^{MI}(W_{m,t}) = \begin{cases} F_{m,t}^{(0)} + F_{m}^{(1)} W_{m,t} & \text{if } \underline{W}_{m} \leq W_{m,t} \leq \overline{W}_{m} \\ 0 & \text{if } W_{m,t} = 0, \end{cases} \tag{11.29}$$

where the intercept $F_{m,t}^{(0)} > 0$ depends on the air temperature in hour t.

Constraint (11.2) imposes the electric balance in every hour t. The total electricity supply is given by the production of cogeneration units and the output of photovoltaic plants. The electricity demand consists of four terms: the first term is the electrical load; the second term represents the electricity required by the absorption chillers for generating cooling power; the third and fourth terms express the electricity required by the electric heat pumps and by the gas absorption heat pumps. This constraint determines the amount W_{t}^{A} of electricity to be purchased from the market, if production is not sufficient to satisfy the hourly demand, or the amount W_{t}^{V}

of electricity to be sold into the market, if production exceeds the hourly demand. Constraints (11.3), (11.4) and (11.5) state that in every hour $t \in \hat{T}$, in which the purchase price μ_t is smaller than the sale price λ_t, electricity cannot be simultaneously purchased and sold ("no-arbitrage" constraints).

Constraint (11.6) guarantees that the cooling load is served in hour t, i.e. the cooling power generated by absorption chillers, electric heat pumps and gas absorption heat pumps, plus the output of ice storages, is required not to be less than the sum of the cooling load and of the input of ice storages.

Constraint (11.7) requires the h-temperature thermal load to be served in hour t. The binary parameters α_h^S, $\alpha_{r,h}$, $\alpha_{b,h}$, $\alpha_{p,h}$, $\alpha_{g,h}$, $\alpha_{m,h}$, $\alpha_{j,h}^{out}$ and $\alpha_{j,h}^{in}$ identify the devices (absorption chillers, boilers, electric heat pumps, gas absorption heat pumps, cogeneration units, solar thermal plants and hot storages, respectively) used for serving the h-temperature thermal load. The total production plus the output of hot storages cannot be less than the sum of the load, of the thermal power used by the absorption chillers and of the input of hot storages.

Constraint (11.8) imposes that the cooling power generated by absorption chiller r in hour t is nonnegative and bounded above by its capacity. Analogous restriction is expressed by constraints (11.10) and (11.13) for electric heat pump p and gas absorption heat pump g, respectively.

Constraint (11.9) states that the total thermal power generated by boiler b in hour t is nonnegative and bounded above by its capacity. Constraints (11.11) and (11.14) express analogous restrictions for electric heat pump p and gas absorption heat pump g, respectively.

Constraint (11.12) guarantees that electric heat pump p cannot simultaneously generate both cooling and thermal powers. Analogous restriction is expressed by constraint (11.15) for the gas absorption heat pump g.

The status of cogeneration unit m is represented by the binary variable $\gamma_{m,t}$ defined in constraint (11.16). Constraint (11.17) imposes that if cogeneration unit m is started up in hour t, it must be *on* either for the minimum uptime, if $t + t_m^U - 1 < |T|$, or until the last hour $|T|$, otherwise. Analogously, constraint (11.18) imposes that if cogeneration unit m is shut down in hour t, it must be *off* either for the minimum downtime, if $t + t_m^D - 1 < |T|$, or until the last hour $|T|$, otherwise. Constraint (11.19) states that in hour t the power production is between the minimum and the maximum power output, if its status is *on*, otherwise is 0. The associated thermal output is guaranteed by constraint (11.20) to be nonnegative and bounded above by the maximum thermal output, if its status is *on*, otherwise is 0. The maximum thermal output in hour t is given by the heat recovery function of cogeneration unit m, which is the affine function Q_m^{MI} of power output $W_{m,t}$

$$Q_m^{MI}(W_{m,t}) = \begin{cases} q_m^{(0)} + q_m^{(1)} W_{m,t} & \text{if } \underline{W}_m \le W_{m,t} \le \overline{W}_m \\ 0 & \text{if } W_{m,t} = 0, \end{cases} \tag{11.30}$$

with $q_m^{(0)} > 0$.

The energy stored in ice storage k at the end of hour t, which is required by constraint (11.22) to be nonnegative and bounded above by the maximum storable energy, must satisfy the balance constraint (11.21). Constraints (11.23) and (11.24) impose that the input energy rate and the output energy rate of ice storage k in every hour t are nonnegative and bounded above by their maximum values, respectively. Analogously, the energy stored in hot storage j at the end of hour t, which is required by constraint (11.26) to be nonnegative and bounded above by the maximum storable energy, must satisfy the balance constraint (11.25). Constraints (11.27) and (11.28) guarantee that the thermal input and the thermal output of the hot storage j in every hour t are nonnegative and bounded above by their maximum values, respectively.

11.3 A Heuristic Procedure for Large Instances

The solution of the mixed integer linear programming model for the optimal annual dispatch requires to consider the set T of cardinality $|T| = 8760$, because of the hourly discretization. The computational effort may therefore become a substantial issue, as the cardinality of the sets H, R, B, P, G, M, K, J and $\hat{T}$ increases. In this section a heuristic procedure is introduced for approximating the optimal solution. The procedure consists of the three steps, A, B and C, described below.

In step A the following mixed integer linear programming model is solved:

$$\begin{aligned}\min \sum_{t\in T}\Bigg\{ & \sum_{r\in R}\left(C_r^O \dot{Q}_{r,t}^C\right) + \sum_{b\in B}\left[\left(C_b^F F_b + C_b^O\right)\left(\sum_{h\in H}\dot{Q}_{b,h,t}^H\right)\right] + \\ & + \sum_{p\in P}\left[C_p^O\left(\dot{Q}_{p,t}^C + \sum_{h\in H}\dot{Q}_{p,h,t}^H\right)\right] + \\ & + \sum_{g\in G}\left[\left(C_g^F F_g^C + C_g^O\right)\dot{Q}_{g,t}^C + \left(C_g^F F_g^H + C_g^O\right)\left(\sum_{h\in H}\dot{Q}_{g,h,t}^H\right)\right] + \\ & + \sum_{m\in M}\left[C_m^F\left(F_m^{(1)} + \frac{F_{m,t}^{(0)}}{\overline{W}_m}\right)W_{m,t} + C_m^O W_{m,t}\right] + \mu_t W_t^A - \lambda_t W_t^V\Bigg\}\end{aligned} \tag{11.31}$$

subject to

- For $m \in M$ and $t \in T$

$$0 \le W_{m,t} \le \overline{W}_m \tag{11.32}$$

$$0 \le \sum_{h\in H}\dot{Q}_{m,h,t}^H \le \left(q_m^{(1)} + \frac{q_m^{(0)}}{\overline{W}_m}\right)W_{m,t} \tag{11.33}$$

- Constraints (11.2) to (11.15) and (11.21) to (11.28)

The mixed integer linear programming model (11.31), (11.32), (11.33), (11.2) to (11.15) and (11.21) to (11.28) is based on the following assumptions:

1. The fuel consumption of cogeneration unit $m \in M$ is proportional to the power produced; therefore, in the fifth term of (11.31) the linear fuel consumption function
$$\left(F_m^{(1)} + \frac{F_{m,t}^{(0)}}{\overline{W}_m}\right) W_{m,t} \tag{11.34}$$
is used.
2. The cogenerator heat recovery is proportional to the power produced; therefore, in constraint (11.33) the upper bound to thermal output is expressed by the linear heat recovery function
$$\left(q_m^{(1)} + \frac{q_m^{(0)}}{\overline{W}_m}\right) W_{m,t}. \tag{11.35}$$
3. The minimum power output of cogeneration unit $m \in M$ is zero, as expressed by constraint (11.32).
4. The cogeneration units are not subject to minimum up time and minimum downtime constraints; therefore, constraints (11.17) and (11.18) are neglected.

The four hypotheses above imply that the binary variables $\gamma_{m,t}$ are no longer needed, which allow definition (11.16) to be neglected. Let $W_{m,t}^*$ denote the optimal power production of cogeneration unit m in hour t determined by the approximated model.

In step B the binary parameter $\gamma_{m,t}^*$, for $m \in M$ and $t \in T$, is first assigned the value $\gamma_{m,t}^* = 1$, if $W_{m,t}^* \geq \underline{W}_m$, or $\gamma_{m,t}^* = 0$, otherwise. The values of parameters $\gamma_{m,t}^*$ are then suitably redefined, if necessary, in order to satisfy minimum up time and minimum downtime constraints.

In step C the objective function (11.1) is minimized subject to constraints (11.2) to (11.15) and (11.21) to (11.28), where variables $\gamma_{m,t}$, for $m \in M$ and $t \in T$, are assigned the values $\gamma_{m,t}^*$ computed in step B.

11.4 The Economic Evaluation of the Trigeneration System

After the simulation of one year of optimized operation of the trigeneration district, the procedure computes the $EBITDA$ (Earnings Before Interest, Taxes, Depreciation and Amortization) that depends on the optimal values of the decision variables W_t^A, W_t^V, $\gamma_{m,t}$, $W_{m,t}$, $\dot{Q}_{b,h,t}^H$, $\dot{Q}_{g,t}^C$ and $\dot{Q}_{g,h,t}^H$ resulting from the optimization model

$$
\begin{aligned}
EBITDA = & \sum_{t\in T}\left[\mu_t\left(L_t^E + \frac{L_t^C}{COP^{rif}}\right) + \frac{C_b^F}{EER^{rif}}\left(\sum_{h\in H} L_{h,t}^H\right)\right] + \\
& + \sum_{t\in T}\left(\lambda_t\, W_t^V - \mu_t\, W_t^A\right) + \\
& - \sum_{m\in M}\left[C_m^{fix} + C_m^F\left(\sum_{t\in T} F_{m,t}^{(0)}\gamma_{m,t}\right) + \left(C_m^F F_m^{(1)} + C_m^O\right)\left(\sum_{t\in T} W_{m,t}\right)\right] + \\
& - \sum_{b\in B}\left[C_b^{fix} + \left(C_b^F F_b + C_b^O\right)\left(\sum_{t\in T}\sum_{h\in H}\dot{Q}_{b,h,t}^H\right)\right] + \\
& - \sum_{g\in G}\left[C_g^{fix} + \left(C_g^F F_g^C + C_g^O\right)\left(\sum_{t\in T}\dot{Q}_{g,t}^C\right)\right] + \\
& - \sum_{g\in G}\left[\left(C_g^F\, F_g^H\, + C_g^O\right)\left(\sum_{t\in T}\sum_{h\in H}\dot{Q}_{g,h,t}^H\right)\right] + \\
& - \sum_{r\in R} C_r^{fix} - \sum_{p\in P} C_p^{fix} - \sum_{k\in K} C_k^{fix} - \sum_{j\in J} C_j^{fix}.
\end{aligned}
\tag{11.36}
$$

The first term of *EBITDA* in (11.36) is the value of total energy load:

- The electrical load is evaluated at the electricity purchase price.
- The value of the cooling load is the cost of the electricity needed to satisfy it by an electric heat pump of coefficient of performance COP^{rif}.
- The value of the total thermal load is the cost of the fuel needed to satisfy it by a boiler of efficiency EER^{rif}.

The reference values COP^{rif} and EER^{rif} are chosen by the user. If the expansion of an existing plant is under evaluation, the parameters COP^{rif} and EER^{rif} may be assigned the values characterizing the corresponding components in the existing plant: this allows the expanded plant to be compared, in terms of reduction of generation costs, with respect to the existing one. The second term of *EBITDA* in (11.36) represents the net revenues from selling electricity to the grid. The third and fourth terms represent the fixed and variable costs of cogeneration units and boilers, respectively. The fifth and sixth terms represent fixed and variable costs of gas absorption heat pumps. The last four terms are the fixed costs of absorption chillers, electric heat pumps, ice storages and hot storages, respectively.

Cooling and thermal loads, L_t^C and $L_{h,t}^H$, as well as the intercept $F_{m,t}^{(0)}$ of the fuel consumption function of cogeneration unit m, depend on the air temperature in hour t, which is not known with certainty when investment decisions have to be taken. Further uncertainties are related to fuel prices and market electricity prices. For these uncertainties to be taken into account explicitly, a stochastic programming model should be solved, with uncertainties of the above-mentioned model parameters represented by a scenario tree. In real instances, however, the solution of the stochastic optimization problem corresponding to the problem under study requires

a computational time not compatible with an on-line use of the decision support procedure. Planned future work will be concerned with computing lower and upper bounds of the optimal objective function value of the stochastic model: this will be done by solving a set of stochastic subproblems, each one with a very small number of scenarios (e.g. two or three), and by evaluating, at different levels of information, how the deterministic solution performs in the stochastic framework, following [7] and [8], respectively. In the current version of the decision support procedure the user can introduce information related to the variability of weather conditions, which influences load and efficiency, by assigning up to three sets of yearly data for air temperatures (in order to represent three possible situations: warm, cold and average) with the associated probability of occurrence. Analogously, the user can introduce information related to the variability of prices by assigning up to three sets of yearly data and the associated probability of occurrence. An average *EBITDA* is computed if variability of input parameters is taken into account. Also, in order to better describe the expansion of a trigeneration system, the evolution in time of the system configuration may be taken into account and the optimal annual dispatch is determined for each configuration of the system, e.g. a transitional configuration in the first year, in which a reduced energy load is served, and a final configuration. For each configuration the corresponding average *EBITDA* is computed.

The methodology used to assess the profitability of the trigeneration system is based on the analysis of annual cash flows, in accordance with the instructions contained in [6].

11.5 Case Studies

In [11] the decision support system has been used for assessing the profitability of a new investment in a CHP system fed by biofuel and for comparing it with a natural gas boiler. In such a comparison several aspects have been taken into account, like incentives, volatility of biofuel prices, higher investment costs of CHP plants and a low load factor (i.e. concentration in few hours) of the thermal load in the residential district. The optimal dispatch of the system has therefore to be determined considering the irregular profiles of space heating and domestic hot water demands, the technical features of different generators (boilers and *CHP*s), like minimum power output, flexibility constraints and efficiency, and the variability of commodity prices (for electric energy) every hour. The use of a biofuel in distributed power plant, with capacity not greater than 1 MW, allows to apply for either the all-inclusive feed-in tariff or the Green Certificates (see [9]). The sensitivity of the investment profitability has been analysed under different hypotheses of biofuel price. The evaluation has been performed using both community and extra-community bio-oils, which have different prices and benefit from different incentive mechanisms.

In this chapter we consider a residential district of typical dimensions in Northern Italy, consisting of 359 flats of small size (50 m^2 on average), with a total surface of 18 000 m^2 and volume 52 000 m^3, to be served by a trigeneration system for

providing space heating, cooling and domestic hot water (*DHW*). The annual energy loads and their peaks are shown in Table 11.1. The total thermal load consists of two parts: demand of district heating and demand of domestic hot water.

Table 11.1 Data of electrical, cooling and thermal load

	$[kWh]$	
Total annual electrical load	298 546	$\sum_{t\in T} L_t^E$
Electrical load peak	255	$max_t\ L_t^E$
Total annual cooling load	188 387	$\sum_{t\in T} L_t^C$
Cooling load peak	253	$max_t\ L_t^C$
Total annual thermal load (district heating + *DHW*)	2 118 456	$\sum_{t\in T} \sum_{h\in H} L_{h,t}^H$
Thermal load peak (district heating + *DHW*)	1 301	$max_t\ \sum_{h\in H} L_{h,t}^H$

The system configurations to be compared, which are defined in Table 11.2, are subsets of the following set of components:

- Turbec T100 gas microturbine, with rated electrical power 105 *kW* and rated thermal power 167 kW
- Condensing boiler, with rated thermal power 978 kW
- Condensing boiler, with rated thermal power 300 kW
- Condensing boiler, with rated thermal power 600 kW
- 4000 L tank, with thermal power 250 kW (input/output) and capacity 200 kWh, for hot storage
- Array of 15 non-reversible gas absorption heat pumps GAHP-A-LT, each of 41.7 kW rated thermal power, with total thermal power 625.5 kW
- Array of 15 reversible gas absorption heat pumps GAHP-AR-LT, with total thermal power 562.5 kW and total cooling power 253.5 kW
- Array of reversible electric heat pumps, with total thermal power 550 kW and total cooling power 400 kW
- Ammonia-water absorption chiller, with cooling power 88.6 kW
- 5370 L ice storage, with cooling power 250 kW and capacity 500 kWh

In Case 1 the thermal load for space heating and domestic hot water is satisfied by the reversible electric heat pumps, which also supply the cooling load, and by two condensing boilers (boilers 1 and 2, with total power 1278 *kW*). The hot water for the *DHW* demand is stored in the tank. Electricity is purchased on the market. The optimization model determines which source of heat will serve the thermal load in each hour, taking into account the hourly electricity price, the air temperature and the load levels. Table 11.3 shows that in the optimal solution about 25% of the thermal load is supplied by the electric heat pumps, which are more convenient than boilers when the electricity price is low and the thermodynamic cycle that depends on the air temperature is efficient. In this solution the total cost for providing space heating and *DHW* is less than the total cost of producing the required thermal power by a small boiler located in each flat (Fig. 11.1).

Table 11.2 System configurations

	Electric power (kW)	Heating rate (kW)	Cooling rate (kW)	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
Hot storage		250		x	x	x	x	x	x
Boiler 1		978		x	x	x	x	x	x
Boiler 2		300		x	x	x	x		x
Boiler 3		600						x	
Reversible electric HP		550	400	x	x	x	x		
Cogeneration unit	105	167			x		x	x	
GAHP		625.5				x	x		
Ice storage			250					x	x
Absorption chiller			88.6					x	
Reversible GAHP		562.5	253.5						x

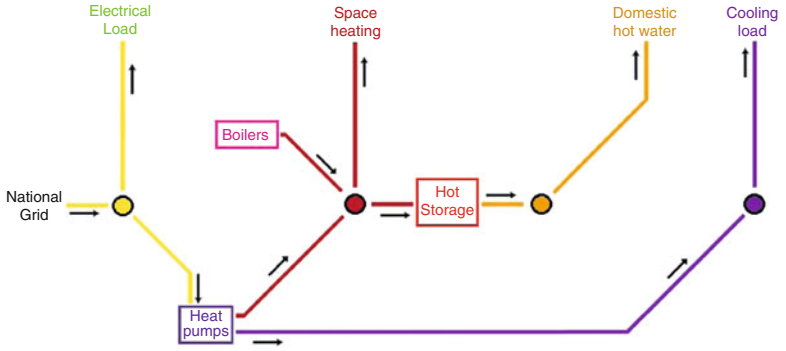

Fig. 11.1 System configuration in Case 1

In Case 2 the cogeneration unit is added to the system considered in Case 1. The cooling load is served by the reversible electric heat pumps. Electricity may be either purchased on the market or produced by the cogeneration unit. Reversible electric heat pumps, condensing boilers and the cogeneration unit satisfy the thermal load for space heating and domestic hot water. In the optimal solution the reversible electric heat pumps supply the thermal load in hours in which the electricity price is low and the thermodynamic cycle that depends on the air temperature is efficient. The electricity load, as well as the electricity used by the electric heat pumps, is supplied by the market when the price is low, otherwise by the cogeneration unit: this happens in many hours during the summer, as well as in hours with high electricity prices—in these cases the cogeneration unit satisfies a portion of the electrical load and the thermal power output is either used for serving the *DHW* demand or stored in the tank for use in subsequent hours, if it exceeds the thermal load. In some hours it is convenient to generate electrical power by the cogeneration unit, even if the recovered thermal power is wasted: this corresponds to the thermal surplus reported in Table 11.3, which is about 0.75% of the total cogenerated heat. In the optimal

solution the cogeneration unit works for 4,751 equivalent hours per year, yielding an expected life of 14 years for the microturbine, the total number of working hours being 66,000 h (Fig. 11.2).

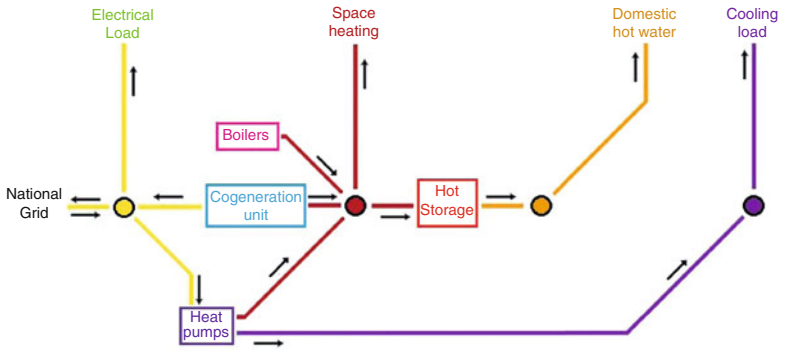

Fig. 11.2 System configuration in Case 2

In Case 3 the array of non-reversible gas absorption heat pumps is added to the system considered in Case 1. Table 11.3 shows that the thermal load is mainly satisfied by gas absorption heat pumps, as they use less natural gas than boilers: this is due to the fact that gas absorption heat pumps have a much higher Gas Utilization Efficiency than boilers (up to 150%), because a part of the thermal energy is taken from the air. The electric heat pumps are used almost exclusively for the cooling load (see Table 11.4) (Fig. 11.3).

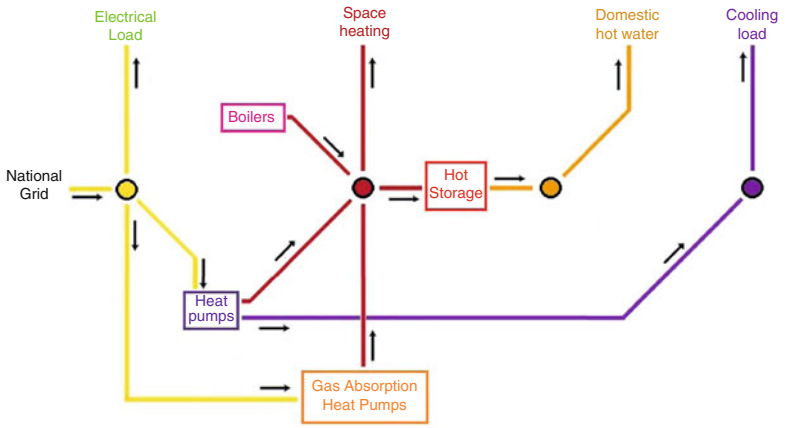

Fig. 11.3 System configuration in Case 3

In Case 4 both the cogeneration unit and the array of non-reversible gas absorption heat pumps are added to the system considered in Case 1. The model suggests

that the reversible electric heat pumps satisfy the thermal load in hours when the electricity price is low and the air temperature allows an efficient thermodynamic cycle. The electricity load, as well as the electricity used by the electric heat pumps, is supplied by the market when the price is low, otherwise by the cogeneration unit, which produces the required electricity, while the recovered heat is used for satisfying the thermal load. As shown in Table 11.3 most of the thermal power demand is satisfied by gas absorption heat pumps, which use less natural gas than boilers; a small amount is supplied by the most convenient source among heat pumps and boilers (depending on hourly electricity price, air temperature and load levels) and the remaining part by the cogeneration unit. Table 11.5 shows that most of the electricity is provided by the cogeneration unit. High volumes of purchased and sold electricity are due to the hourly market prices that in some hours make it convenient to either sell or purchase electricity. In the optimal solution the cogeneration unit works only for 3,465 hours per year, yielding an expected life of about 20 years for the microturbine. This system configuration requires the highest investment among the six considered, but yields the highest *EBITDA* (see Table 11.7) (Fig. 11.4).

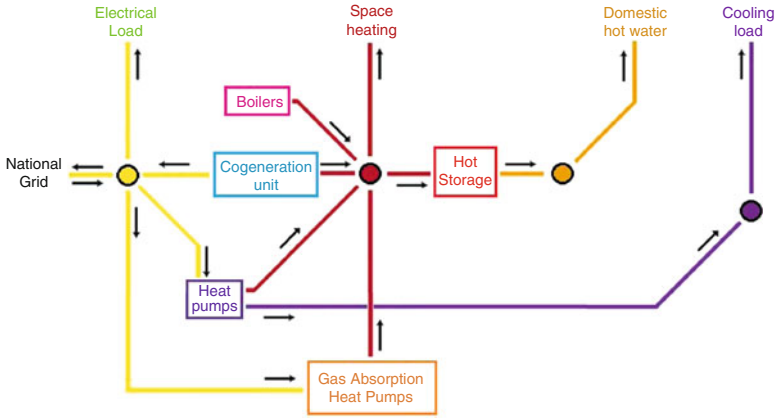

Fig. 11.4 System configuration in Case 4

In Case 5 the trigeneration system consists of the cogeneration unit, two condensing boilers (boilers 1 and 3, with total power 1578 kW), the hot water tank for supplying *DHW* demand and the absorption chiller coupled with the ice storage, as absorption chillers (unlike electrical heat pumps) can produce very low temperatures (e.g. −33°C). In the absence of electric heat pumps, boiler 3, with 600 kW maximum heating rate, is used instead of boiler 2, with 300 kW maximum heating rate, in order to guarantee satisfying the peak demand of thermal power. This configuration is useful when it is preferred to reduce the amount of electricity purchased from the market (see Table 11.5). In the optimal solution the cogeneration unit works for 6 055 equivalent hours per year, yielding an expected life of 11 years for the microturbine (Fig. 11.5).

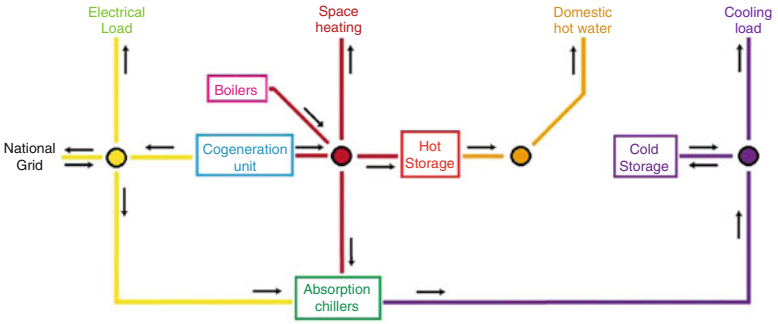

Fig. 11.5 System configuration in Case 5

In Case 6 the configuration differs from the one in Case 3 as the array of reversible electric heat pumps is no longer included in the system and the cooling load is served by the array of reversible gas absorption heat pumps, coupled with the ice storage. In this configuration reversible gas absorption heat pumps serve both the thermal and the cooling loads, obtaining a decrease of the electrical load and requiring much lower investment cost with respect to Case 3. On the other hand, the cooling power production by reversible gas absorption heat pumps is less convenient than by reversible electric heat pumps (see the annual consumption of natural gas and electricity in Table 11.6); therefore, a lower *EBITDA* than in Case 3 is obtained, as shown in Table 11.7.

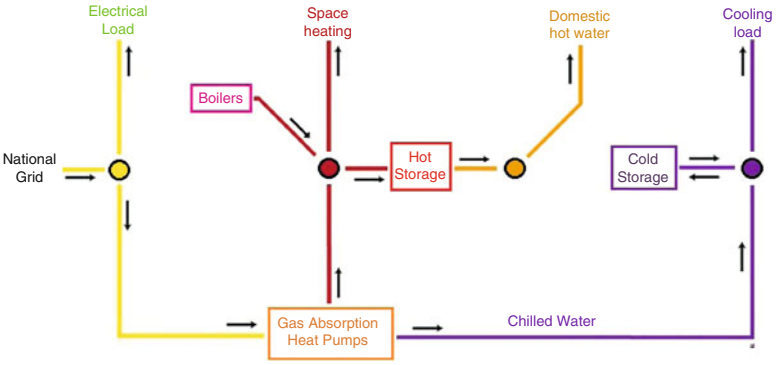

Fig. 11.6 System configuration in Case 6

In Tables 11.3, 11.4 and 11.5 the results obtained by the optimal dispatch model in the six configurations are reported in aggregated form. The total energy supplied by every generator and the total energy of every usage in the year is reported for every case. A generator is not included in the trigeneration system if the sign "–" appears in the corresponding column (Fig. 11.6).

Table 11.3 Thermal power: total annual supplies and uses

	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
Supplies:						
Boilers	1578.7	1079.3	217.1	104.7	1477.1	385.6
Electric heat pumps	563.0	302.9	38.6	17.9	–	–
Cogeneration unit	–	766.8	–	559.3	976.9	–
Gas absorption heat pumps	–	–	1885.3	1465.6	–	1775.4
Uses:						
Thermal load	2118.5	2118.5	2118.5	2118.5	2118.5	2118.5
Hot storage losses	23.2	24.8	22.5	23.6	24.6	22.5
Absorption chiller	–	–	–	–	279.3	–
Thermal surplus	0.0	5.8	0.0	5.5	1.7	0.0

All values are in MWh and refer to a year of operation

Table 11.4 Cooling power: total annual supplies and uses

	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
Supplies:						
Reversible electric heat pumps	188.4	188.4	188.4	188.4	–	–
Reversible gas absorption heat pumps	–	–	–	–	–	190.0
Absorption chiller	–	–	–	–	195.7	–
Uses:						
Cooling load	188.4	188.4	188.4	188.4	188.4	188.4
Ice storage losses	–	–	–	–	7.3	1.6

All values are in MWh and refer to a year of operation

Table 11.5 Electricity: total annual supplies and uses

	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
Supplies:						
Electricity market (purchase)	489.6	123.3	415.4	165.7	43.8	348.2
Cogeneration unit	–	498.9	–	363.9	635.7	–
Uses:						
Electrical load	298.5	298.5	298.5	298.5	298.5	298.5
Electricity market (sale)	0.0	189.6	0.0	127.8	372.0	0.0
Electric heat pumps	191.0	134.1	76.1	71.5	–	–
Gas absorption heat pumps	–	–	40.7	31.7	–	49.6
Absorption chiller	–	–	–	–	9.1	–

All values are in MWh and refer to a year of operation

Based on [6], the analysis of annual cash flows is performed in order to assess the profitability of the different configurations of the trigeneration district. An industrial life of 20 years is considered, assuming that the worn parts of the cogeneration plant are replaced in Cases 2 and 5. In Table 11.7 the equity C_0, assumed to cover 50% of the total investment, the $EBITDA$ of the reference year, the net

Table 11.6 Annual consumption of natural gas and electricity and sold electricity

	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
Natural gas consumption	1714.6	2776.8	1397.5	2175.2	3624.4	1708.9
Electricity consumption	489.5	432.6	415.4	401.7	307.6	342.5
Sold electricity	0.0	189.6	0.0	127.8	372.0	0.0

All values are in *MWh* and refer to a year of operation

present value (*NPV*), the internal rate of return (*IRR*), the payback time (*PBT*) and the average debt service coverage ratio (*DSCR*) are reported for each configuration and in Fig. 11.7 the actualized cash flows in the time horizon are shown for the six configurations.

Table 11.7 Economic indicators

	C_0 (k€)	*EBITDA* (k€)	*NPV* (k€)	*IRR* (-)	*PBT* (years)	*DSCR* (-)
Case 1	55.9	58.9	357.4	0.748	1.5	9.36
Case 2	110.9	73.2	306.5	0.490	2.3	5.86
Case 3	145.9	84.9	427.3	0.409	2.9	5.17
Case 4	200.9	92.3	416.5	0.327	3.6	4.08
Case 5	98.9	22.5	48.0	0.157	9.7	2.02
Case 6	11.9	70.6	358.4	0.418	2.8	5.27

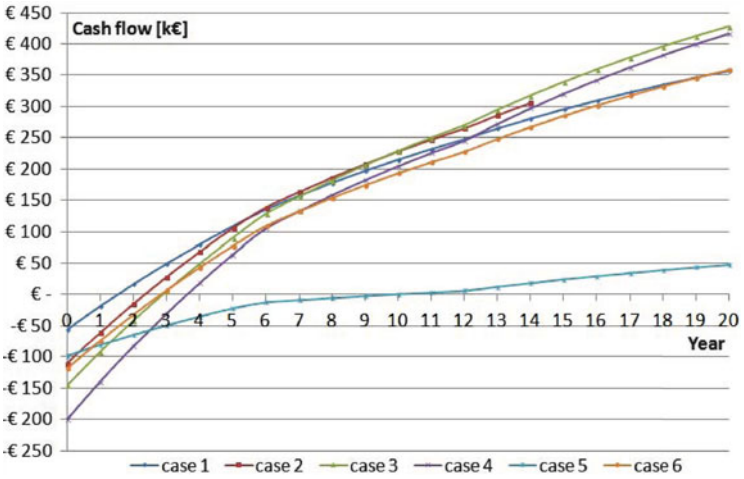

Fig. 11.7 Cash flows comparison for the six configurations

In configuration 1 the use of more efficient boilers and electric heat pumps than those installed in the reference system yields a reduction of generation costs that allows the investment to be paid in about 2 years, as shown in Fig. 11.7 by the intersection of the blue line with the abscissa line. In configuration 2 the addition of the cogeneration unit requires higher investment costs; on the other hand generation costs are further reduced, which yields a payback time of 2.5 years. In this case, moreover, a much lower electrical load is satisfied by the national grid. In configuration 3 the investment cost is higher than in Case 2, as gas absorption heat pumps are more costly than the cogeneration unit. Gas absorption heat pumps allow large reductions in consumption of natural gas and electricity (see Table 11.6). In this case the largest net present value is also achieved. Furthermore this solution also allows satisfying national and European regulations that require a prescribed fraction of production from renewable energy sources in satisfying thermal loads. Configuration 4 combines the advantages of configurations 2 and 3, as it allows achieving the largest energetic and economical savings, with an investment of 800ï£¡ per flat. Configuration 5 has the highest payback time; however, it can be of interest when the proposed configuration is an expansion of an existing system, in which some components are already available (for instance the cogeneration unit) and new elements need to be added, in order to increase the energy load served. In this case the reduced investment cost would result in the light-blue line in Fig. 11.7 to be translated above and therefore in a shorter payback time. In configuration 6 a good level of profitability is achieved, while limiting the electrical consumption. If there are limitation on the maximum electric power to be purchase from the grid, this is the appropriate solution. In the absence of such limitation, configurations 3 or 4 appear to be the best. Configuration 4 requires a higher investment, but it allows to satisfy the energy loads with a lower fuel consumption and is therefore preferable if fuel prices increase.

11.6 Conclusions

In this chapter we have presented the procedure *GDPint* for the evaluation of investments in new trigeneration systems or in the expansion of an existing distributed generation system, taking into account both technical and financial aspects. The optimal dispatch model allows describing the system components in great detail and the economic evaluation allows the users to compare several financial structures. A case study is discussed in which six alternative configurations are compared. The software tool, which can be freely accessed at `www.rds-web.it`, has been extensively used by different kinds of stakeholders (power producers, banks, investors, etc.) as well as power plant engineers and regulation authority, providing a very useful feedback regarding the details to be taken into account in the procedure, as well as the information to be provided to the user for evaluating the investment decisions. Maintenance work is required to keep the tool up to date, with particular reference to the legislation on incentives and taxation. Further work is planned towards a more sophisticated representation of uncertainty, although limited by the quite large dimension of the problem to be solved.

References

1. Armanasco F, Brignoli V, Marzoli M, Perego O, Scagliotti M, Campanari S, Colombo L, Silva P (2006) Analisi tecnico-economica e sperimentazione di sistemi cotrigenerativi. Deliverable RdS - W.P. 1.1 06007178, RSE (www.rse-web.it)
2. Campanari S, Colombo L, Silva P (2007) Performance assessment of cogeneration systems for industrial district applications. In: Proceedings of the American Society of Mechanical Engineers (ASME) Turbo Expo 2007, Montreal. ASME Paper no. GT2007-27659, doi:10.1115/GT2007-27659
3. Campanari S, Manzolini G, Silva P (2008) A multi-step optimization approach to distributed cogeneration systems with heat storage. In: Proceedings of the American Society of Mechanical Engineers Turbo Expo 2008, Berlin. ASME Paper no. GT2008-51227, doi:10.1115/GT2008-51227
4. Campanari S, Manzolini G, Silva P (2009) Comparison of detailed and simplified optimization approaches for the performance simulation of gas-turbine cogeneration plants. In: Proceedings of the American Society of Mechanical Engineers Turbo Expo 2009, Orlando. ASME Paper no. GT2009-60114, doi:10.1115/GT2009-60114
5. Chemelli C, Gelmini A, Marciandi M (2008) Gendisplan: applicativo software per la definizione fuori linea dell'ottimo tecnico-economico nell'impiego di risorse energetiche afferenti ad una microrete. Deliverable RdS 08005854, RSE (www.rse-web.it)
6. EPRI. Technical assessment guide. Technical Report TR-102276-V1R7, Electric Power Research Institute, 1993
7. Maggioni F, Allevi E, Bertocchi M (2012) Measures of information in multistage stochastic programming. In: Stochastic programming for implementation and advanced applications (STOPROG-2012)
8. Maggioni F, Wallace SW (2012) Analyzing the quality of the expected value solution in stochastic programming. Ann Oper Res 200(1):37–54
9. MSE. Decreto legge sull'incentivazione della produzione di energia elettrica da fonti rinnovabili, http://www.sviluppoeconomico.gov.it/pdf-upload/decreto181208.pdf. Ministero dello Sviluppo Economico, 2009.
10. Stadler M, Cardoso G, Bozchalui M, Sharma R, Marnay C, Siddiqui A, Groissböck M (2012) Microgrid modeling using the stochastic distributed energy resources customer adoption model der-cam. Conference presentation lbnl-5937e, Lawrence Berkeley National Laboratory
11. Vespucci MT, Zigrino S, Bazzocchi F, Gelmini A (2011) A software tool for the optimal planning and the economic evaluation of residential cogeneration districts. In: Delimar M (ed) Proceeding of the 8th International Conference on the European Energy Market (EEM 2011), Zagreb, 2011

Chapter 12
Renewable Energy and Its Impact on Power Markets

David Wozabal, Christoph Graf, and David Hirschmann

Abstract The widespread introduction of renewable energy production is transforming electricity markets all around the globe. The changes are often hard to anticipate for market participants and the resulting uncertainty about future market conditions, policy regimes, technologies, and prices makes participation in these markets risky. In this article, we focus on changes induced by the growing capacities of wind power and photovoltaic electricity production. We highlight some aspects of power markets that are currently changing fundamentally due to increased capacities in these technologies. In particular, we discuss technological development, predictability and stochastic modeling of wind and solar output, policy issues pertaining to subsidies for renewable energies, and effects on the electricity prices on spot markets. We illustrate our findings using data from Germany and the Californian electricity market.

12.1 Introduction

Like players in any other industry, firms operating in the commodity market face a range of uncertainties about future developments which influence profits and thereby introduce risk in the planning process. In the energy sector, structural change

D. Wozabal (✉)
Institute for Finance, Banking and Insurance, Vienna University of Economics and Business, Welthandelsplatz 1, 1020 Vienna, Austria
e-mail: david.wozabal@wu.ac.at

C. Graf • D. Hirschmann
Institute of Business Administration, University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria
e-mail: christoph.graf@univie.ac.at; david.hirschmann@univie.ac.at

R.M. Kovacevic et al. (eds.), *Handbook of Risk Management in Energy Production and Trading*, International Series in Operations Research & Management Science 199,
DOI 10.1007/978-1-4614-9035-7_12,

is mainly driven by the concern about the emission of climate active gases and the resulting global warming with its possible adverse effects. Another driver is the geostrategic concern about oil dependence of many industrialized nations.

The result is a global move towards renewable energies in its various forms which concerns all the major fields of energy consumption including electricity, heating/cooling, and transport. According to [60], in 2011, 71% of the newly installed power generation capacities in the EU are renewables, while in the USA this number is at 39%. Furthermore, from 2009 to 2011 the global amount of produced biodiesel and ethanol increased by 18% and 20%, respectively, and capacities for solar heating increased by 52% in the same time. In most industrialized countries this change is heavily supported by policy schemes: currently 92 countries/states support renewable power generation with feed-in tariffs and 72 countries/states introduced quota systems for renewable power generation (see [60]), and the EU countries adopted a cap and trade emission trading system to limit the amount of carbon dioxide emissions.

In this article, we focus on liberalized electricity markets which in many countries fundamentally changed due to the increased deployment of renewable generation capacities as well as regulatory frameworks intended to promote investment in clean generation technology. To narrow down the scope of the article, we mostly take the view of electricity producers. This choice is motivated by the fact that the expansion of renewable power generation affects the decision problem of producers more than that of consumers, whose decision space with regard to the purchase of electricity is most often fairly limited.

We do not consider specific risk management models but rather try to give an overview of the different ways the promotion of renewable energy production introduces uncertainty into the decision of electricity producers and retailers on all levels of planning. In particular, the impact of increased deployment of renewables induces short-run as well as long-run uncertainty in the electricity markets:

- Short-run (several hours to several months): Most renewable sources of electricity have in common that they are intermittent, i.e., their electric output at any given point in time depends on environmental factors that are outside the control of the owner of the plant and therefore have to be considered random. Typical examples are wind farms producing electricity dependent on wind speed, solar panels whose output depends on the cloud cover, or hydro plants which are dependent on the amount of precipitation. As is evident from the examples, this uncertainty is mainly pertaining to the output of already existing plants and is driven by weather-related factors. While accurate weather models are available, prediction of intermittent infeed based on these models is always subject to uncertainty and thereby introduces risk in the short-term planning of electricity generators. We investigate this topic in more detail in Sect. 12.2.2.
- Long-run (one year to several years): Long-term uncertainty in the electricity price induced by renewables stems mainly from uncertain political conditions and regulatory frameworks on the one hand and technological development on the other hand. Political uncertainties include direct measures to disincentive the emission of carbon dioxide as well as schemes to support different types

of renewable energy in the form of feed-in tariffs, tax incentives, or fixed quotas for renewable energy production. Technological developments change the relative prices and therefore the attractiveness of certain technologies, either through innovations in the technology itself or through cost savings in the production. Both of these factors influence the long-term appeal of a particular technology and thereby the investment decisions in the electricity market. This in turn introduces long-term uncertainty about the production mix and therefore about average electricity prices as well as other characteristics of the price process such as price variability and seasonality.

Short-run uncertainties mostly affect operational and tactical decisions of plant usage, i.e., how much energy is produced on a given day, how to bid on various markets, or how much water to withhold in hydroelectric storages. Long-run uncertainties, on the other hand, affect investment decisions in power plants and infrastructure. For these decisions changing price levels as well as shifts in the price patterns play a crucial role as long-run price levels and volatility influences investment decisions for peak technology, storage devices, and the grid. It is crucial that firms understand these changes and the associated risks.

The aim of this paper is to discuss several of the abovementioned aspects including different drivers for long-term developments and their impact on the electricity market and the short-term stochastics of intermittent technologies. For this purpose, we focus on the technologies which, as we argue in Sect. 12.2, will be the most disruptive to the power markets in the coming years, namely wind power and solar power.

We leave out some important topics to keep the length of the paper at bay:

- The electric grid: The extensive buildup of renewable generation resources leads to challenges for the electric grid in some countries. The reason for this is that the location of renewable energy production often does not coincide with the regions where electricity is consumed, which may lead to considerable stress for existing transmission lines. This is, for example, the case in Germany, where the majority of the wind energy is produced in the coastal regions of the north sea and the Baltic sea while consumption is centered in the south of Germany. Another challenge is that grids are not designed to receive feed-in in low-voltage subgrids, which leads to problems in case of large-scale decentralized production by wind turbines or solar panels not connected to high-voltage lines; see [6].
- Storages, smart grid with demand response: To achieve a high penetration of intermittent renewable energy, it is necessary to be able to control the demand for energy either by promoting technologies that make a demand response by consumers possible (smart grid) or by the buildup of large capacities of electricity storages.

This paper is organized as follows: In Sect. 12.2, we outline the recent development of the most promising renewable technologies and discuss the cost of different technologies as well as predictability of solar and wind power output and statistical models for these quantities. Since most technologies for renewable power generation are currently not profitable, we discuss various subsidy schemes

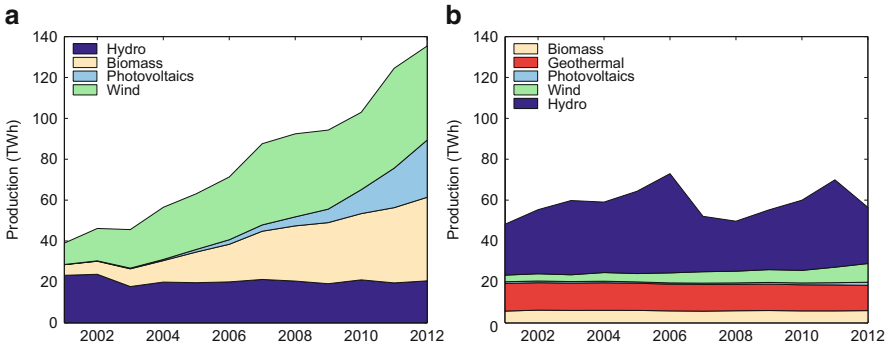

Fig. 12.1 Renewable electricity production in (**a**) Germany and (**b**) California in the years 2001–2012. Sources: [2, 11, 16, 23]

in Sect. 12.3. We also contrast the programs in California and Germany as examples of two fundamentally different approaches encouraging the expansion of renewable generation capacities. Sect. 12.4 is devoted to the effects of renewables on the power price. Section 12.5 concludes the paper.

12.2 Technologies

The most promising technologies for renewable electricity production are hydropower, burning of biogas and biomass, photovoltaics, wind energy, and, in some areas, geothermal production. See Fig. 12.1 for the development of energy production from renewable sources in Germany and California in the last 12 years.

As opposed to the other listed technologies, hydropower is a mature and well-understood technology which is extensively used for many decades. As a result the overall potential for further projects is rather low in most countries and therefore the impact of hydropower on the electricity markets of the future is limited.

Biogas is generated by anaerobic digestion of either residential or agricultural waste, wood, or plants that are grown solely for this purpose. While plants that burn biogas or biomass to produce electricity operate largely carbon dioxide neutral, they otherwise work on the same basis as conventional gas plants. In particular, generation is fully controllable and the plant has significantly positive marginal costs of production.

In this article, we focus on technologies that have the most disruptive and transformative potential for electricity markets, namely, wind and solar power. Both technologies are known for decades but were not employed on a mass scale for production of power until recently. Concerns about climate active gases lead to intensified efforts to produce clean, i.e., carbon neutral, electricity. The resulting subsidy

schemes (see Sect. 12.3) lead to an increase in the amount of installed capacity in the USA as well as in some European countries, as is illustrated in Fig. 12.1 for Germany and the state of California.

Wind power and solar power both combine three distinctive features that make them relevant for risk management:

1. There is a large potential for further growth.
2. The marginal costs for production are essentially zero.
3. Production is intermittent, i.e., dependent on random weather conditions that cannot be influenced by the owner of the plant.

From a technical standpoint, the capacity of solar power as well as wind power is quite large; see for an assessment of worldwide potential [37, 45, 46]. Clearly, the future growth potential is dependent on many factors such as subsidy schemes, technological progress, and the price of fossil fuels. All these long-term developments influence investment decisions of companies and private individuals and therefore induce uncertainty that has to be considered in the strategic planning of electricity-producing companies.

Note that the combination of the points 2 and 3 above adds a certain degree of unpredictability to electricity markets with a high penetration of wind and solar power: 2 implies that electricity produced will always be fed into the grid whenever prices are positive and 3 implies that it is not known when exactly how much electricity will be produced. Hence, these characteristics introduce risk in short-term planning. We will discuss these aspects in more detail in Sect. 12.2.2 and in Sect. 12.4.

12.2.1 Cost Structure of Solar and Wind Power

The main driving factor for the success of a specific power generation technology is its cost of production. Hence, while many countries subsidize renewable power generation, the hope is that after some time the respective technologies mature in a way that they can compete with traditional technologies. However, calculating production costs is not straightforward, since the total costs of building and operating a plant over a long time horizon has to be compared with the amount of power that can be produced during the lifetime of the plant. The cost per MWh of produced electricity obtained this way is usually called *levelized cost of electricity* (LCOE) . LCOE is a frequently reported figure, since it serves as the basis for many arguments and policy decisions for example about the subsidies for renewable energies.

As hinted above, the calculation of LCOE is usually based on strong assumptions including specifications for the following inputs:

1. The *overnight capital costs* of a plant, i.e., the cost of construction not taking into account any interest rates during construction. These costs heavily depend on the corresponding legal and environmental requirements, as well as on the cost of land, labor, and construction material.

2. The *capacity factor*, which indicates how many hours a year the plant can produce at peak capacity. While quantifying the capacity factor is less of a problem for conventional plants, it poses a challenge for intermittent technologies for which the capacity factors vary with environmental conditions like average wind speeds or the amount of sunlight in a specific region.
3. The discount rate, which is used to make costs that occur at different times comparable. Usually the discount rate is calculated as a *weighted average cost of capital*, requiring assumptions on the equity to debt ratio, the internal rate of return of the investor, and the interest rates for loans over a long period of time.
4. Variable costs of production like variable operations and maintenance (O&M) costs and, for conventional plants, the price of coal, gas, or uranium required to produce one MWh of electricity.
5. The lifetime of the project, possibly a yearly loss of efficiency (as is the case with solar panels) as well as a salvage value (possibly negative) for the components of the plant after the end of operation.

Different studies use different assumptions regarding the above variables and therefore come to widely varying LCOE estimates—especially for renewable source of electricity, for which reliable long-term figures are often not available and technological progress leads to rapidly varying prices. See for example [8, 13] for a critical review of the LCOE literature regarding photovoltaics.

For the purpose of this paper, we calculate the LCOE of wind and solar power and compare it to figures of a combined cycle gas turbine (CCGT) and an advanced pulverized coal plant (APC). We use the LCOE calculator provided by the *National Renewable Energy Laboratory* [1] and input parameters from the recent study by the [74]. The results are summarized in Table 12.1. Note that the calculation routine was augmented to include a yearly degeneration rate, i.e., a decrease in maximum capacity, to accurately reflect the technical specifications of photovoltaic panels.

At the time of writing, gas as well as coal prices were rather low; hence, there is a considerable input price risk and the costs of these technologies might very well rise during the lifetime of the plant, making energy produced from coal or gas potentially more expensive. To keep our results independent of policy schemes, possible costs for CO_2 emission rights, a factor that raises the LCOE of conventional production, as well as measures that make renewables more attractive like tax credits or subsidized loans are not taken into account.

As the name suggests, the calculation is purely cost based and does not consider the revenue that can be made by different technologies. The revenue that can be made for 1 MWh of produced electricity varies between technologies mainly because of the different flexibility and thereby the ability to react to changing prices on electricity exchanges. Typically this flexibility is the highest for gas-fired plants, followed by coal plants and nuclear power stations. Intermittent technologies, which cannot control output at all, can be considered as entirely inflexible. Note that the cost advantage of the CCGT plant over the APC plant is due to the high capacity factor of the former. A capacity factor of 0.85 for a CCGT plant is technologically feasible but not economical in most markets, because of the high marginal cost of production (see Sect. 12.4.1).

[1] See http://www.nrel.gov/analysis/tech_lcoe.html.

As can be seen from the results in Table 12.1, conventional production has a clear cost advantage over utility-scale solar production and offshore wind power, which is not yet a mature technology. However, the most advanced renewable technology, onshore wind production, seems to be competitive at current prices.

Utility-scale solar power is clearly the most expensive way to produce electricity and cannot compete with other technologies in terms yet. However, we also calculated the LCOE of a classical rooftop installation which is nowadays very popular in some countries due to generous subsidies. The results show that due to a lower cost of capital for private investors, which we assume to be 5%, and lower overnight capital cost, the LCOE for rooftop installations of photovoltaic panels is substantially more competitive than utility-scale installations and comes close to the costs of conventional technologies.

Table 12.1 LCOE estimates for wind power, solar power, a combined cycle gas turbine (CCGT), and an advanced pulverized coal plant (APC) in EUR

	Solar		Wind		Fossil fuel	
	Rooftop	Utility	Onshore	Offshore	CCGT	APC
Lifetime	30	30	30	30	30	30
Interest	5%	9%[a]	9%[a]	9%[a]	9%[a]	9%[a]
Overnight capital cost	1,770,000[b]	3,735,000	1,593,893	3,812,214	651,145	2,064,885
Capacity factor	0.25[c]	0.25[c]	0.34[c]	0.37[c]	0.85[c]	0.87[c]
Fixed O&M costs	21,183	21,183	30,191	56,489	11,733	28,855
Variable O&M costs	0	0	0	0	0.00252[d]	0.0034[d]
Heat rate (Btu / KWh)	0	0	0	0	6,430	8,800
Fuel cost (EUR/MMBtu)	0	0	0	0	7.73[e]	3.51[f]
Degeneration (per a)	1%[g]	1%[g]	0	0	0	0
LCOE (EUR / MWh)	**69.97**	**193.07**	**62.23**	**131.91**	**62.34**	**64.50**

All data that is not discussed in the footnote above is taken from [74]

[a] See [79]

[b] http://www.solaranlagen-portal.com/, roof installation in Bavaria, 50m^2 panels, south-east facing roof, output of 6.16 Wp (13.05.2013)

[c] IEA 2018 projections; see http://www.eia.gov/forecasts/aeo/er/electricity_generation.cfm.

[d] National Renewable Energy Laboratory: median of estimates for 2012 from studies in 2008–2012

[e] NCG Natural Gas Futures 2014, Source: EEX (13.05.2013)

[f] BP Statistical Review of World Energy 2011, prices for 2011

[g] See [13]

Another advantage of small rooftop solar installations is that the reference price for competitiveness is not the wholesale price for electricity but rather the retail price paid by household customers, which in most countries amounts to more than double the wholesale price because of grid fees, taxes, and, in some countries, subsidies for renewables. If the cost for production is lower than the retail price for electricity, it

pays off for households to consume their own production and thereby a photovoltaic installation might be competitive even if the costs exceed the production costs of other technologies.

The point in time where the LCOE from a technology equals the price of electricity in the grid is called *grid parity*. Breyer and Gerlach [14] project that in 2020 grid parity for photovoltaic production will be reached in most industrialized countries and [63] estimate that in 2016 it would be profitable for a typical German household to install photovoltaic panels for own consumption without any subsidies, even if only 25% of the energy can be directly consumed and the rest is sold for wholesale prices to the grid. This, combined with the fact that the price for solar panels is rapidly decreasing, makes growth of the installed capacities in the coming years very likely—even in countries with less generous subsidy schemes.

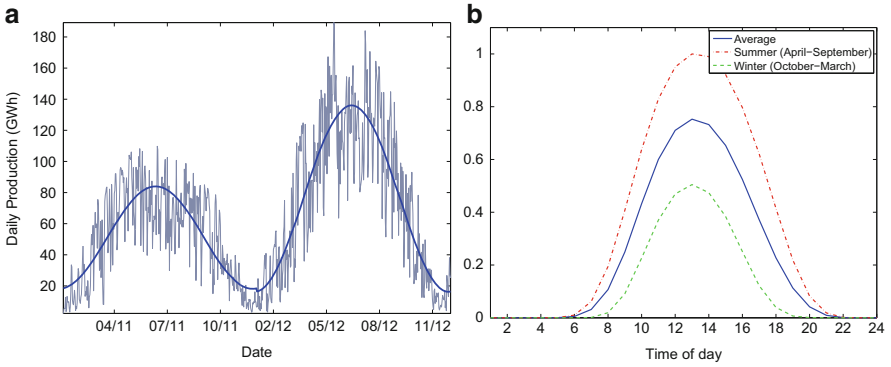

Fig. 12.2 (**a**) History of daily photovoltaic generation in Germany for the years 2011–2012 and trigonometric trend. (**b**) Daily infeed pattern from German production for the years 2011–2012

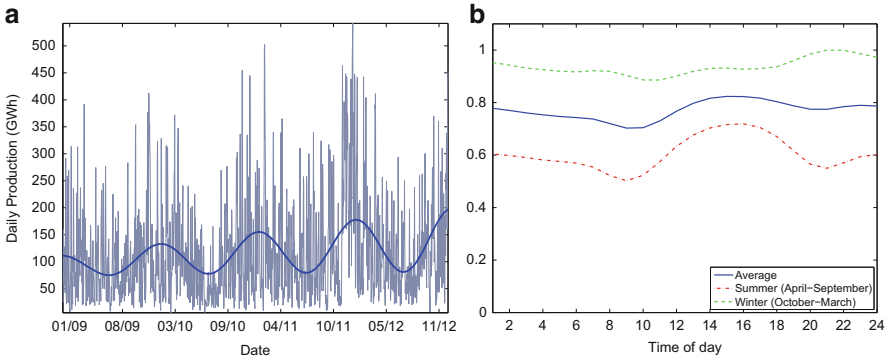

Fig. 12.3 (**a**) History of daily production of wind power in Germany for the years 2009–2012 and trigonometric trend. (**b**) Daily infeed pattern from German production for the years 2009–2012

12.2.2 Trend, Stochastics, and Predictability

Since wind and solar power are intermittent technologies, it is important to have predictive models for electricity production from these technologies. Such a model optimally consists of two parts: predictors as well as a stochastic model for the forecasting error. Both of these components are essential when planning plant dispatch and we will discuss them separately in this section.

We start with predictive models and an investigation of trends and seasonality for wind and solar power using data provided by the four German *Transmission System Operators* (TSO). The energy output of these technologies is mainly dependent on weather conditions and as such is subject to yearly seasonality. Figure 12.2 (a) shows the daily quantities of produced solar power over the last two years in Germany. Clearly, there is a pronounced seasonality with nearly no production in midwinter and peak production in midsummer. The trend is represented by a simple trigonometric model with installed capacities as an additional variable and an interaction term between capacity and seasonality, whose coefficients were estimated by least squares. As can be seen from the plot, the model captures the seasonality and the impact of additional capacity quite well, which is also reflected by the relatively high R^2 of 74%. The daily variations for summers and winters are depicted in Fig. 12.2b.

Figure 12.3 shows a similar analysis for wind power production. Figure 12.3a depicts the trend in wind infeed from 2009 till 2012. It can be seen from the trend analysis that the yearly seasonality is much less pronounced than in the case of solar energy with slightly higher production during winters than in summers. However, this yearly seasonality is superimposed by large daily variations in wind output. Looking at Fig. 12.3b, we can see that also daily seasonality is much flatter for wind energy than for solar power. Moreover, we remark that the pattern for wind and solar energy production might vary considerably between different regions of the world.

Summarizing, we can say that solar production follows a much more regular trend than wind energy production and hence it is easier to incorporate into medium-term planning models. Another interesting finding is that wind power also fluctuates strongly at a yearly level, i.e., there seem to be more and less windy years. This is illustrated by Fig. 12.5, where it can be seen that in Germany, production in the years 2007 to 2010 stagnated although there was a large increase in capacities. Likewise in California wind energy production constantly rose in the years 2004 to 2010, although capacities stayed constant.

Table 12.2 MAPE and RMSE (in MWh) of wind and solar forecasts for the year 2012 in Germany

	MAPE			RMSE		
	15 min	60 min	Daily	15 min	60 min	Daily
Wind power	0.29	0.28	0.18	281	1,112	18,488
Photovoltaics	0.40	0.46	0.17	276	1,069	10,625
Wind + photovoltaics	0.20	0.19	0.12	350	1,388	21,804

Seasonality captures medium-term trends and expected output weeks or months ahead. Equally important is the forecasting of the output for the next 24 h to several days. For these short horizons accurate models are available that take forecast wind speeds and cloud cover as well as the specifics of the plant into account. A review of these rather technical meteorological models is beyond the scope of this paper and we refer to [25] for a survey.

To assess the quality of short-term predictions, we use the 24-hour-ahead forecasts for wind and solar power published by the four German TSOs for theirs respective grid zones and compare them with the actual values for the year 2012. We calculate the *root mean squared error* (RMSE) and the *mean absolute percentage error* (MAPE), defined as

$$\text{RMSE} = \sqrt{n^{-1}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}, \quad \text{MAPE} = n^{-1}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right|$$

where $(\hat{y}_i)_{i=1}^n$ are the forecast and $(y_i)_{i=1}^n$ are the actual values. To avoid numerical problems with the computation of MAPE for the solar power forecasts, we excluded observations with a solar power output of less than 1 MWh.

The results summarized in Table 12.2 show that the absolute error, expressed by RMSE, is on average slightly above 1 GW on an hourly level for both technologies summing up to a daily forecasting error of 18.5 GWh and 10.6 GWh for wind and solar, respectively. The relative error, measured by the MAPE, suggests that predictions for both wind and solar power are prone to rather large errors on shorter time scales, whereby predictions for solar power are much worse in this respect with an average percentage error of 46% on an hourly scale. However, it also has to be remarked that the MAPE of the solar power output drops to similar levels as the MAPE for wind power if hours with less than 10 MW are excluded from the calculation of the MAPE.

Before the large-scale introduction of renewable energy, the uncertainty about the residual demand, i.e., the load that has to be covered by controllable plants, basically resulted from uncertainties about the exact demand. Since the total demand for electricity can be forecast with a MAPE of about 2% (see for example [68]), the electricity networks were much more predictable. Systems with a high penetration of renewables require large capacities for balancing energy, which typically is provided by storage plants, fast gas or oil turbines, or changes in output of already-running base-load plants. We, therefore, also calculate the errors in the prediction of the sum of wind and solar power, since this quantity is important when calculating the residual demand which in turn influences the power price on competitive markets (see Sect. 12.4): the sum of solar and wind output can be forecast with lower relative error than any of the two individual time series. However, using Spearman's rank-order coefficient, we cannot reject the null hypothesis of independence of the forecast errors of the two technologies at any reasonable level of confidence. Hence, we conclude that forecast errors do not seem to cancel out to a greater extent than what would be expected for the sum of two independent random variables. The figure for the RMSE of the sum shows that the demand for balancing energy

caused by volatile renewable energy is on average 1.4 GW on a 15-minute level, a fluctuation that does not seem to be too big considering the size of the market of around 60 GW on average. Looking at the worst 2% in terms of forecasting errors, i.e., the 1% and 99% quantiles of the 15-minute error distribution, we find errors of 3.7 GW and 3.9 GW, respectively.

Having discussed seasonality and forecast errors, we now turn to distributional models for renewable energy infeed, which are needed to assess risk in electricity markets. To model the stochastics of the power output, most authors first model the stochastics of weather conditions and then transfer these random conditions to random power outputs using the technical specifications of the system. In the case of wind, wind speeds are usually modeled as either normally distributed [31, 40] or following a Weibull distribution [38]. The wind speed is translated to power output by a polynomial power curve (with cutoffs), i.e., a function mapping the wind speed to the amount of power produced; see for example [38]. For solar energy a similar approach is followed: in the first step a stochastic model for the so-called clearness index [34] is developed and then this quantity is translated to a random power output using the specifications and the location of the system.

More elaborate models that also take into account the time dynamics and reliability of the systems using Markov-chain models can be found in [5, 84]. The joint stochastics of wind and solar power infeed is considered in [70].

12.3 Regulatory Frameworks to Support Renewable Energy

From an economic perspective, the main line of justification to support renewables through public policies are the effects of negative externalities which occur from producing electricity with conventional technologies, e.g., the negative environmental effects from burning fossil fuels like the emission of greenhouse gases. Another argument for promotion of renewable energy sources is the reduction of dependency on fossil fuels and the security of supply in the energy sector.

According to standard microeconomic theory, efficient market equilibria are distorted when facing widespread externalities. Since the value of public goods like clean air or the protection from climate change is not captured in the marginal cost of the suppliers, energy prices are lower than what they should be. In the following subsections, we elaborate on two ways of dealing with negative externalities: direct and indirect mechanisms.

The most efficient way of internalizing an externality would be by means of *directly* taxing emissions or introducing a permit market.[2] However, the unpopularity of taxes or permit supplements is one reason why indirect mechanisms play a large role around the globe.

[2] The Coase theorem provides bargaining as a solution to cope with externalities in a market. However, property has to be defined in order to trade the externality. Furthermore, for the theorem to be applicable, the absence of transaction costs is necessary. In the case of a public good like environmental protection, property rights are poorly defined and also transaction costs would be severe.

Indirect mechanisms are tools to support (renewable) technologies rather then tackling the negative externality directly. Given sufficient support, renewable technologies are deployed leading to a reduction of the market share of conventional technologies. Hence, these mechanisms *indirectly* cope with the negative externality. Furthermore, the subsidy of certain technologies may trigger dynamic learning effects, which are hard to achieve in direct schemes, as the latter favors the deployment of technologies that are currently most cost-effective. In the following subsections, we describe the two mechanisms in more detail and provide some empirical evidence of implemented indirect mechanisms with a focus on Germany and California.

12.3.1 Direct Mechanisms

Direct mechanisms can be divided in price-based instruments, where the regulating authority directly influences the price of the externality, and quantity-based instruments where the amount of the produced externality is restricted. The question whether prices or quantities are the more efficient instrument to control for negative externalities was first raised by [78]. While in the case without uncertainty about the marginal cost function both instruments yield efficient results, the situation changes when introducing uncertainty about marginal costs, in which case neither instrument yields an optimum ex post. Which instrument eventually can be considered second best in terms of efficiency depends on the shape of the cost and benefit curves [78]. In case of a flat marginal cost curve, the application of a price instrument can have severe consequences on the desired quantity if the marginal costs are estimated wrongly.

The classical textbook example of a price-based instrument is an environmental tax á la Pigou [55]. An application of this instrument is the gasoline tax which targets the externality of burning fossil fuels in the transportation sector and exists in many countries around the world for decades.

The most prominent examples in the class of quantity-based instruments are cap and trade mechanisms in which tradeable permits are allocated to firms. Such trading schemes which are used to price clean air and correct for the negative externalities have been introduced in recent years in some economies.

Although direct mechanisms are considered the most efficient solution, there are two practical reasons why they are difficult to implement. First, the exact valuation of the avoided damage or the value of the public goods preserved necessary to design an optimal tax or an optimal permits market is not possible and second the unpopularity of taxes in general. While the former could be solved by an uncertainty premium, the latter seems to be the more severe reason. Price increases through taxes or permit supplements are, therefore, politically difficult to implement also because of the interlinkage of energy prices and macroeconomic performance of a country. A global consensus on how to economically treat pollutants stemming from burning fossil fuel does not exist.

The European Union Emissions Trading System (EU-ETS) serves as an example for problems associated with the design of a direct mechanism. The EU-ETS is a carbon market which exists since 2003 and initially served as a role model for other countries around the world. However, at the moment the EU-ETS fails to provide proper price signals. The recent price decline (20 EUR/t in 2011 to 5 EUR/t in 2013) is partly due to the economic crisis and partly due to an oversupply of free allowances. This makes the construction of heavily polluting coal plants more attractive than that of cleaner gas plants. This development emphasizes the importance of market design, the knowledge about the value of the public good, and political influence [69].

12.3.2 Indirect Mechanisms

In addition to direct mechanisms described in the previous section, many countries implemented subsidy schemes in order to promote (renewable) technologies. Although this strategy is at most third best in terms of efficiency, it is politically more feasible. Another stream of arguments for supporting renewable technologies concern the stimulation of dynamic learning processes required to reduce costs of renewables to guarantee security of electricity supply in the future at a reasonable price. Electricity production from renewable sources is fairly cheap at the margin; however, subsidy schemes are in many cases the only alternative to encourage investors to bear the high fixed cost of the development of renewables especially in the emerging phase. An increased diffusion of these technologies ensures benefits from dynamic learning effects [4] and eventually helps to make them competitive [47].

In the category of indirect mechanisms there exist many ways to support the diffusion of renewables, for example, investment in research and development (see for instance [1] who argue that an optimal environmental policy involves both carbon taxes and research subsidies when conventional and renewable technologies are sufficiently substitutable), financial support by providing investment subsidies, or tax credits. Yet, the most common support schemes are based on generation and are either price-based or quantity-based instruments.

The so-called feed-in tariffs (FiT) fall into the category of price-based instruments, which typically oblige electric utilities to buy all renewable generated electricity at a price determined by public authorities. This fixed price is usually higher than the market price and moreover guaranteed for a specified period of time. A variation of the FiT is the premium model, where a fixed premium is added to the market price. This premium is again determined by public authorities and guaranteed for a specified period. The level of the market premium should, theoretically, be set at the level of the external costs of conventional power generation. From an investor's perspective, the FiT is clearly preferred over the premium system since market risk induced by volatile prices is excluded. The cost of the market premium or the FiT likewise is generally borne by the final consumers of electricity [15, 30, 47, 62].

In contrast to price-based instruments like the FiT, there exist also quantity-based instruments. The most prominent example in this category is a renewable energy certificate market. In the USA, this quantity-based scheme is called renewable portfolio standard (RPS) and in the UK, renewable obligation (RO) . In this type of scheme, producers or retailers are required to supply a certain percentage of electricity from renewables or otherwise pay a penalty which is redistributed to the renewable energy suppliers. The supply from renewables in total has to meet a certain quota set by the government. Certificates are issued by renewable electricity generators who sell electricity directly on the wholesale market and additionally benefit from selling certificates on the renewable energy certificate market. Trading in certificate markets activates market forces and therefore, theoretically, leads to efficient outcomes [30, 47].

Without further design restrictions, the certificate market does not discriminate between technologies which subsequently can lead to unfair competition between renewable technologies at different stages of development. Therefore, the result would be a rather homogeneous technology mix, i.e., investors would prefer technologies which already experienced high learning rates like wind power to technologies which might need further stimuli in order to become competitive.

Two major criteria influence the investor's decision: risk and profitability. Supporting schemes for renewables are aiming at decreasing risks and increasing profitability. Mitchell et al. [50] consider three types of risk to be relevant for investors: price risk, volume risk, and balancing risk. Volume risk refers to the risk of not being able to sell all the produced electricity, whereas balancing risk is the risk associated to the deviation from a prespecified production schedule.

In the standard design of an FiT, there is no price and volume risk, since typically the total generation from renewables is fed into the system at a fixed price. Note that balancing risk may occur under feed-in schemes depending on their exact design [19, 33, 50]. For example, in the original FiT design in Spain, generators had to pay a fixed penalty for deviating from the schedule [43]. However, it is more common that generators are not exposed to balancing risk in feed-in regimes.

In the market-based certificate trading scheme, without any fine-tuning, quantity risk and price risk are present. The price risks enter through the volatility of electricity prices and the prices of certificates, which fluctuate, due to the intermittent nature of renewable technologies. Volume risk [50] is induced by the quota, which fixes an overall volume for renewable generation. Such a quota does not guarantee that the individual renewable electricity producer's output will be bought. Balancing risk is typically also present in market-based certificate trading schemes. However, in practice, there are workarounds for the balancing requirement, for example, California introduced more lenient balancing rules for producer included in the Participating Intermittent Resource Program [33].

From the risk perspective of an investor, an FiT is favored compared to the market-based support scheme, though it also has some drawbacks. For example, it is seen to be more prone to political risk, i.e., if costs escalate, public support decreases and therefore also political support vanishes. This has been the case in Spain, where after a change of government all support schemes for building of new

renewable capacities were suspended in 2012 [61] and the support scheme were changed to the disadvantage of owners of existing plants. Note that the risk evaluations under the different subsidy schemes hold only under the pure design of the policies described above. In practice, there are various policy design variations, intending to absorb price risks, for example, through introducing cap and floor prices in a feed-in premium system or floor prices for quota schemes [44].

Two important questions have not been tackled so far: which policy instrument is more effective in terms of deployment and which one is more cost-efficient? The literature considers FiTs to be more effective than quantity-based instruments, meaning that they are able to quickly deploy significant capacities of renewables [47, 67]. The most obvious reason for this is the risk reduction of investors as described above. With respect to efficiency the literature is not so clear. Menanteau et al. [47] point out that administrative costs of implementing FiTs are rather low, but also note that quantity-based instruments like the RPS are preferred when the slope of the marginal cost curve is flat. Fagiani et al. [24] analyze cost-efficiency and cost-effectiveness of certificate mechanisms compared to FiTs. They simulated the hypothetical future Spanish electricity market including both policy tools. They conclude that FiTs could obtain higher efficiency than certificate markets if tariffs are well calculated. This means that policy maker's estimation of marginal cost curves need to be accurate; otherwise, the FiT could lead to over- or underinvestment and, thus, inefficiency. Hence, their simulation model confirms the theory by [78]. The ultimate efficiency of an instrument, however, depends on the exact design of the support scheme.

The issue of mixing direct and indirect policies to achieve the goal of emission reduction is controversially discussed in economic literature. Most of the debate on policy mixes for one target is based on [71] who states that only one policy instrument per target is optimal. Böhringer and Rosendahl [12] point out that mixing direct and indirect mechanisms may have unintended consequences. In particular, they analyze the impacts of overlapping quotas for emissions and renewables. The promotion of renewables through quotas leads to a price decrease in the permit market, thus making coal more attractive compared to cleaner gas. Philibert [54] confirms that overlapping direct and indirect policy instruments are likely to increase the cost of achieving the single-objective in the short-run. However, in the long-run, when large-scale deployment of renewables establishes its role in emission mitigation, cost will be considerably lower because of the support in the early phase of deployment [54]. Moreover, supporters of indirect mechanisms emphasize that mitigating emission is not the only target; it has, rather, multiple targets like the security of supply at reasonable prices in the future.

12.3.3 Empirical Evidence of Indirect Mechanisms

Around the globe, FiTs as supporting device for renewables are dominating. By early 2012, FiTs were in place in at least 65 countries and 27 states while quotas or

renewable portfolio standards were in use in 18 countries and at least 53 other jurisdictions [60, 67]. In the USA, 29 states, Washington DC, and two territories have implemented RPSs [20]. In the European Union, more than 20 of the 27 member states use the feed-in system (either premium or tariff) as their main support instrument in 2012. In the last years member states which traditionally used quota systems like Belgium, Italy, and the UK have introduced FiT for small-scale installations [9, 58, 83]. Figure 12.4a shows a map with the prevailing policy tools implemented by the European member states. Spain is marked in light yellow since the former price-based policies to support renewables are currently suspended. Figure 12.4b shows the distribution of states in the USA which have implemented RPSs or have renewable goals.

Germany and California are seen as among the front-runners concerning the deployment of renewable energy in Europe and in the USA, respectively. The 2020 policy goal of the European Union (EU) states that 20% of EU gross final energy consumption is produced by renewable energy sources [22]. To reach this target, the German government decided to reach a share of 35% generation of electricity from renewables by 2020 [9]. In 2011, the Californian government decided on a goal of 33% of electricity generated from renewable energy resources by 2020 [16]. While the targets of Germany and California are quite similar the policies to achieve these goals are very different. Germany opted for an FiT while California implemented the RPS.

In California, the first version of the RPS was established in 2002 and accelerated in 2006 by requiring that 20% of electricity retail sales be served by renewable energy resources by 2010. In 2011, the RPS was amended to mandate 33% production from renewables by 2020 [16, 41].

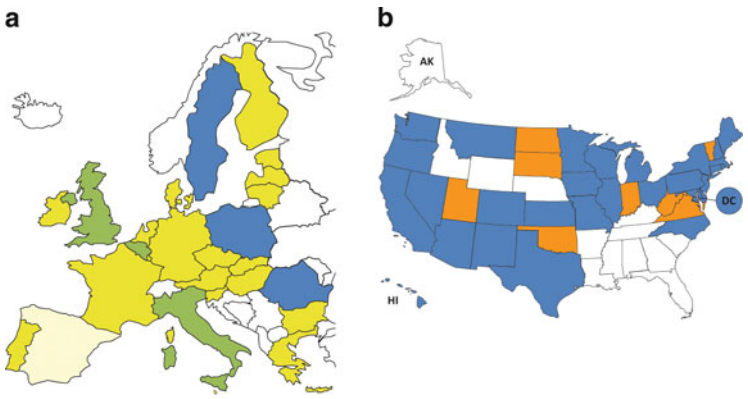

Fig. 12.4 Renewable policy schemes. (**a**) European Union: member states with feed-in system (yellow), quota system (*blue, dark*), and mixed system (*green*) by December 2012. Source: [9, 58, 61]. (**b**) United States of America: states with renewable portfolio standard (*blue*) and renewable portfolio goal (*orange*), by March 2013. Source: [20]

The German Renewable Energy Sources Act (EEG), enacted in the year 2000, was the start of Germany's efforts towards large-scale deployment of renewable energy production capacities. To this day, the EEG has been modified a few times also in order to decelerate the massive buildup of capacity in wind and photovoltaics (see Figs. 12.1 and 12.5, Panel a). This was done by lowering degression rates, which were originally designed to exert cost pressure on manufactures through changes in remuneration for new plants.

The EEG is deemed to be effective since the price, quantity, and balancing risk is eliminated for investors. However, in the recent past, a debate about the efficiency of the German model started because the cost borne by end users saw a dramatic rise from 0.2 ct/kWh in 2000 to 5.28 ct/kWh in 2013 [21, 48]. Therefore, in a recent update of the EEG, Germany implemented incentives for renewable energy producers to switch to more market-based support schemes. In particular, Germany now offers a market premium model as an alternative to the FiT, which allows operators to sell their production directly on the market and receive a premium dependent on the average market price from the government. Moreover, the government pays a management premium in order to compensate for losses due to forecasting errors induced by the intermittent nature of most of the renewable technologies, mitigating balancing risk [10]. The recent modification of the EEG also addresses the lack of controllability of installed capacities through the FiT mechanism.

Figures 12.1 and 12.5a show the transformation of the German electricity supply. Although California produced more electricity from renewables than Germany in 2001, Germany has increased its supply by approximately a factor of 4.5 from 2001 to 2012 while California has increased by a factor of 1.15. This development supports the effectiveness of the German subsidy scheme. In order to compare the numbers in Figs. 12.1 and 12.5, notice that Germany has more than twice as much habitants as California, and also the electricity consumption of Germany with 594.5 GWh compared to 278.3 GWh of California was about two times higher in 2012 [11, 16]. Figure 12.5 depicts the production and installed capacities of photovoltaics and wind in Germany and California. In 2012, Germany was able to supply about 23% [11] of the demand from all renewable technologies, while for California this figure is at 20% [3].

It has to be noted that generation from photovoltaics and wind so far only add a small part to fulfill the demand (about 12% in Germany and 4% in California); however, these two technologies showed considerable dynamics in installed capacity in recent years which is expected to continue in the years to come. As it can be seen in Fig. 12.1, Panel b, production from hydro, geothermal, and biomass plays a major role in California. In the case of Germany, also biomass experienced dynamic growth in the recent years, while the potential of hydro seems to be exhausted (see Fig. 12.1a).

[3] Data on California's hydro generation contains also production from pumped hydro storages, which are normally not counted as renewable.

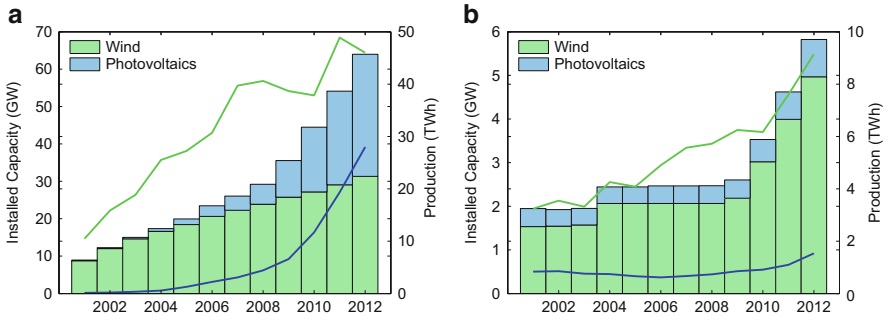

Fig. 12.5 Wind energy and photovoltaics in (**a**) Germany and (**b**) California. Bars indicate installed capacity; lines indicate production. Source: [11, 16, 23]

12.4 Price Effects

Since renewables introduce additional uncertainty to electricity markets, one would also expect to see this effect in the prices. Correspondingly, a lot of research is done to quantify the impact that renewables have on electricity spot market prices.

One can distinguish short-run and long-run effects on the prices, as mentioned in Sect. 12.1. It is widely agreed on that in the short-run, an increasing share of renewables will lower the average price. At the same time, literature indicates that price volatility will increase.

The long-run effects are more difficult to estimate since strategic decisions of producers, policies, and technological changes will determine the power generation mix and consequently the future electricity prices; see [53, 77, 80].

In the following subsections, we introduce a simple static equilibrium model for price formation on electricity markets, which we subsequently use to discuss shot-run and long-run effects of renewables.

12.4.1 The Merit Order—A Static Model

The merit order is a helpful theoretical concept to model how spot market prices form. It is a way of ranking sources of electricity production. Each utility and its capacity are lined up, ordered by their marginal production costs, from lowest to highest, resulting in the so-called merit order. Ignoring renewables, uranium and lignite plants are found to the left end of the line, whereas gas and oil plants are found at the right end. Figure 12.6 shows an example of a merit order for the German market.

Often, the term merit order is used not only to refer to the actual ranking of the plants but also to the resulting curve that relates quantity (x-axis) to marginal costs (y-axis). Therefore, the merit order can be interpreted as the electricity supply curve.

Even though in general, the merit order is discontinuous, to facilitate analysis of the merit order effect, a continuous curve is assumed in most of the literature. Furthermore, it is usually assumed that the curve is convex; see, e.g., [7, 57, 65] for examples of merit order curves in US markets and [32, 59] for the German market.

The demand for electricity is rather inelastic. Most consumers do not base their electricity consumption on current prices, since they usually buy their electricity for a fixed tariff which is independent of short-term fluctuations on spot markets. Furthermore, it is difficult for these consumers to shift demand to a different point in time and there is no possibility to store electricity economically. So far, the only tools to vary demand from customers are interruptible contracts which exist mostly between producers and industry. In future, smart grids could be used for demand management where different price signals influence consumption patterns. Critics state that the effect will be limited as a big share of consumption is still very inelastic: demand management can only be used for appliances like dishwashers, refrigerators, washing machines, and possibly air-conditioning. The situation could change if more storage capacities were introduced, e.g., hydrogen storages or strategic use of batteries of parked electric cars.

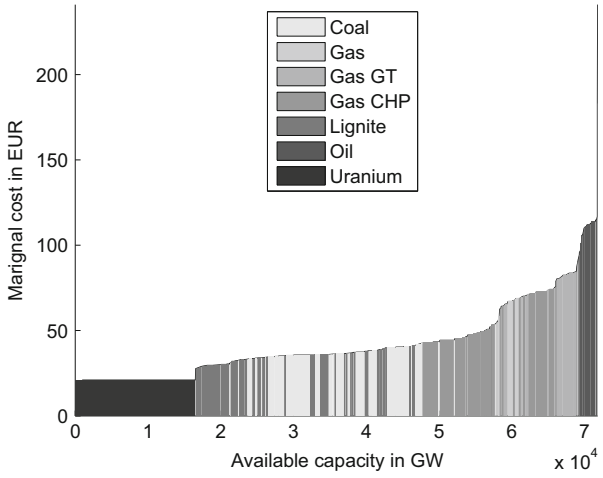

Fig. 12.6 Constructed merit order for Germany (conventional plants only), November 17, 2010. See [26]

Assuming perfect competition and neglecting time dependencies and grid constraints, the merit order determines the spot price. To cover a certain demand, the marginal cost curve of the industry, represented by the merit order, is intersected with the inelastic demand curve. The point of intersection determines the electricity price and all the plants with marginal costs below that price are switched on to cover the demand.

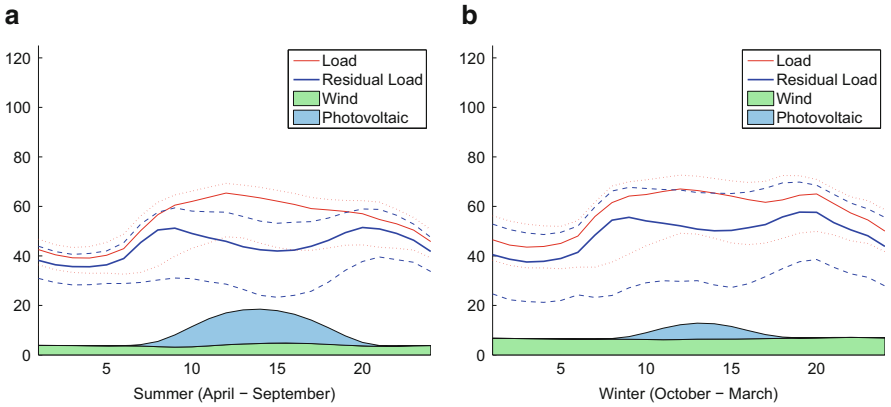

Fig. 12.7 Demand and residual demand (approximated by loads) for Germany in (**a**) summer and (**b**) winter in the year 2012. *Solid lines* indicate hourly medians, whereas *dotted lines* indicate the corresponding 0.05 and 0.95 quantiles. Wind and photovoltaic infeeds are hourly means. All quantities are MWh. Source: Data published by German TSOs

12.4.2 Renewables and the Residual Demand

Renewable sources enter the merit order typically at the left end because of their low marginal cost and therefore will almost always dispatch. Only the demand not covered by renewables will be covered by conventional energy sources with positive marginal costs. Another way to see this is that demand for conventional energy is lowered by renewable energy production. In the static demand and supply framework, this is equal to a shift in the demand. The shifted demand, that is, the remaining demand for conventional energy, is the so-called residual demand.

In this context it is interesting to note that while wind is fed into the grid directly, photovoltaics can effect demand in two ways: either the produced energy is fed into the grid as well or households satisfy their own demand. In either case, the demand for conventional energy is lowered.

For an example of the behavior of the residual demand, see Fig. 12.7. The figure shows the influence of renewables on demand in Germany in 2012. The thin, solid line is the load, used as a proxy for demand. On the bottom of the figure, the amount of wind and solar energy is indicated. The difference between load and production from renewables is a proxy for residual demand and is indicated by the dark, solid line. For the loads, the hourly median and the 0.05 and 0.95 quantiles are shown; for renewables the hourly average production is depicted at the bottom of the graph. The following can be noted:

- The shape and level of the (residual) demand changes with the introduction of renewables. Wind shifts the entire residual demand curve downwards, since wind infeed is almost constant in the mean. Photovoltaics, on the other side, change the shape of the residual demand curve. Since the infeed from photovoltaic production is concentrated in the hours with peak consumption and therefore peak price,

additional solar capacities dramatically lower the prices during these hours. One result of this development is that the price differences between day and night hours, on which pumped hydro storage facilities rely to make profits, are shrinking considerably. However, these developments may give rise to a morning and a late evening peak in future. This shows that renewables can influence price patterns in a nontrivial way.

- The distribution of the residual demand changes. The quantiles show how the distribution becomes broader. This indicates that renewables are an additional source of uncertainty.

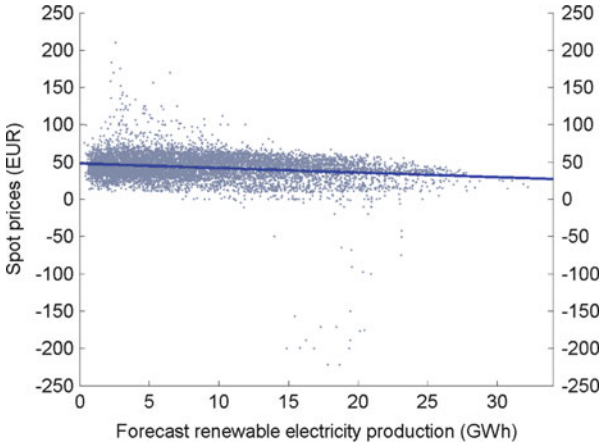

Fig. 12.8 Renewable energy production and spot market prices for Germany in the year 2012. The *solid line* indicates the regression line. Source: Renewable electricity production published by German TSOs and spot prices published daily by the EEX

12.4.3 Short-Run Effects

As described above, the residual demand is the demand curve shifted by the infeed of renewable energy. Since the residual demand is always smaller than the demand, the price is always lower due to the upward slopping shape of the merit order. Therefore, the average price decreases.

This effect was shown in many theoretical and empirical works. Traber and Kemfert [72] calibrate a computational model for Germany and find price decreases due to infeed of wind power. Similar results are derived in [64] by means of agent-based modeling. Although they find almost no price reduction in low load hours, they report a reduction up to EUR 36/MWh in hours of peak load. Nicholson et al. [52] find that wind generation mildly, but significantly, reduces the market price in the four-zone Electricity Reliability Council of Texas market. For Denmark, [51] report a decrease in spot prices if wind penetration increases.

Jónsson et al. [39] point out that the spot market prices are not influenced by the actual wind infeed but rather by the wind forecasts. This can be attributed to the fact that electricity spot markets are usually settled one day before delivery. In Fig. 12.8 the day-ahead forecasts for renewable electricity production are plotted against the spot market prices. One can see that low forecasts correlate with high prices, whereas high forecasts typically yield lower prices.

Another interesting effect, which can also be observed in Fig. 12.8, is the surge of negative spot prices which mainly arise in hours of high infeed of renewables. In these hours utilities are actually paid to consume electricity or lower their production. However, negative prices are not necessarily related to an oversupply of renewables. Andor et al. [3] show that negative prices were also observed in the German market even when there was no renewable energy traded. The reason is that some suppliers accept negative prices due to ramping costs and opportunity costs of their power plants. However, it is widely accepted that the increase in renewables has increased the number of negative price events. It is also interesting to note that the surge of negative spot prices is strongly related to the type of subsidy scheme. For example, under a renewable energy certificate market (see Sect. 12.3.2), renewable energy utilities will feed in at negative costs, as long as the price of the certificates is higher than the loss they make. In an FiT scheme producers will always feed in their energy, regardless of the market price.

Whereas in literature lower average prices in the short-run are well documented, price variance is not studied as extensively. Price variance is an important risk factor and an input for financial modeling of electricity products, for consumer contract pricing, or for the market strategy of a pump storage facility.

The following oddity, reported by [18] for Australia, indicates how important price fluctuations can be for producers: during September 2008 and August 2010, prices were dominated by 25 extreme peaks exceeding AUD 1000/MWh. These prices contributed 45% and 63% of the spot market revenue during September 2008 until August 2009 and September 2009 until August 2010, respectively, although representing just 0.25% and 0.42% of the 30-min trading intervals.

In general, it is concluded in literature that with a higher share of renewables, price variability increases, e.g., [17, 27, 36, 81]. The main point is often that the range and the variability of residual demand increases and that this carries over to the prices. Ketterer [42] studies the influence of wind on German electricity prices and price variance in particular using a GARCH model and concludes that variable wind power increases volatility.

Note that short-run effects may also depend on the time between market bids and delivery due to better weather predictions for shorter time horizons [35] as well as the functioning of balancing markets [75] and the availability of storage to react to deviations from load predictions.

12.4.4 Long-Run Effects

So far, we have studied the short-run effects of renewables on prices using the merit order with a fixed set of power plants. Matters are more difficult for long-run effects. The reason is that producers react to market developments and adapt their portfolio mix dynamically, which leads to a change in the shape of the merit order [53, 80].

Weber and Woll [77] point out that an increasing share of renewables can reduce the degree of capacity utilization of base-load plants. As base-load plants have high fixed costs, producers may favor the building of peak-load electricity generation plants. Green and Vasilakos [28] conduct a numerical study for the UK and come to similar conclusions. They show that when maximizing social welfare, the capacity of base-load conventional power generators would fall, but the amount of flexible gas-fired generators would increase. Steggals et al. [66] recognize that due to higher uncertainty about load factors of peak plants and depressed average prices, investment decisions become increasingly complicated. Nevertheless, they also find that for the future portfolio mix, gas-fired plants are more likely to be built due to their lower capital costs.

In a portfolio mix with a high share of peak-load plants, it is more likely that an expensive peak-load plant determines the market clearing price if demand exceeds production of renewables. Weber and Woll [77] argue that this could offset the price-reducing effect of wind energy, where [49] show that this can eventually increase average prices and price variance.

It is also interesting to note how renewable energy affects base-load plants. Troy et al. [73] report that cycling (starting/stopping operation or varying load levels) due to varying wind infeed causes accelerated deterioration of components. One consequence, they argue, could be that generators alter their bid strategies in order to minimize cycling damage which would affect the market prices.

Contrary to the studies mentioned above, [72] argue that currently incentives for investments into flexible gas-fired plants are missing: wind energy production reduces their load factors overproportionally and consequently, earnings are low at current prices. Woo et al. [82] find decreasing revenues for gas-fired plants in Texas if wind energy production increased further. Therefore, [72] propose to implement a policy under which wind energy producers have to take the balancing risk. Such a policy would not only affect the supply curve but also open up new market perspectives: wind producers may want to sell the risk of varying infeeds or trade more actively on the balancing market.

In general, policies and subsidy schemes can have a great impact on the future supply curve and therefore on the prices. As an example we mention the effect of CO_2 prices: if CO_2 prices were sufficiently high, then this could interchange the position of coal- and gas-fired plants which would fundamentally change the merit order.

Yet another scenario would be the introduction of large-scale storage capacities or the alternative use of oversupply. Green et al. [29] consider the long-run impact on electricity prices and on the optimal capacity mix in the UK, if hydrogen was produced in times of excess wind production. The produced hydrogen could

be used either to generate electricity, for transport, or for combined heat and power. The authors find that the presence of hydrogen production via electrolysis would raise average price of electricity and that in the cost-minimizing capacity mix, nuclear plants would replace gas-fired power stations, as demand could be spread more evenly.

Finally, an alternative approach to study the future portfolio mix is to consider strategic interaction among producers. A stream of literature models investment decisions in capacities in a game-theoretic setting. The models are formulated as Cournot or Stackelberg games and are solved numerically. Although they help to understand the strategic interactions, these models are rather stylized and do not permit detailed predictions of the portfolio mix. For an overview of these types of models see [56, 76].

12.5 Conclusion

In this article we tried to give an overview of the most important drivers of the fundamental changes that can be observed on many electricity markets nowadays. In particular, we analyzed the technological and economic impact of increased capacities in wind and solar power, dealing with issues of levelized costs of different technologies as well as the predictability and stochastic modeling of renewable production. Furthermore, we reviewed the current status of subsidy schemes for renewables and contrasted FiT policies as they are applied in Germany with RPS, which are more popular in the USA. We also used a stylized market model, the merit order, to investigate the short-run and long-run impact of increased infeed of intermittent renewable electricity in a day-ahead spot market.

We conclude the paper by emphasizing that the next decade will probably bring about further far-reaching changes in the electricity markets and there is scope for a lot more research to help market participants understand the developments and the risk they are taking in their investment as well as in their operational decisions. In particular, we mention the topic of forecasting the output of renewables and the econometric modeling of the pertaining random fluctuations, the issue of grid stability, demand response and the need for large-scale electricity storage, the efficiency of different kind of subsidy schemes, as well as concrete stochastic planning models for investment and operational decisions taking into account the abovementioned uncertainties.

References

1. Acemoglu D, Aghion P, Bursztyn L, Hemous D (2012) The environment and directed technical change. Am Econ Rev 102(1):131–166

2. AG Energiebilanzen eV (2013) Bruttostromerzeugung in Deutschland von 1990 bis 2012 nach Energieträgern. URL http://www.ag-energiebilanzen.de
3. Andor M, Flinkerbusch K, Janssen M, Liebau B, Wobben M (2010) Negative Strompreise und der Vorrang Erneuerbarer Energien. Zeitschrift für Energiewirtschaft 34:91–99
4. Arrow KJ (1962) The economic implications of learning by doing. The Rev Econ Stud 29(3):155–173
5. Arwade S, Lackner M, Grigoriu M (2011) Probabilistic models for wind turbine and wind farm performance. J Sol Energy Eng 133(4):9. doi:10.1115/1.4004273
6. Azmy A, Erlich I (2005) Impact of distributed generation on the stability of electrical power system. In: Power Engineering Society General Meeting, 2005. IEEE, vol 2, pp 1056–1063
7. Barlow M (2002) A diffusion model for electricity prices. Math Finance 12(4):287–298
8. Bazilian M, Onyeji I, Liebreich M, MacGill I, Chase J, Shah J, Gielen D, Arent D, Landfear D, Zhengrong S (2013) Re-considering the economics of photovoltaic power. Renew Energ 53:329–338
9. BMU (2012) Renewable energy sources in figures. National and International Development. Technical report German Ministry for the Environment, Nature Conservation, and Nuclear Safety (BMU)
10. BMU (2013a) Eckpunkte der EEG Novelle sowie sonstige Neuerungen für erneuerbare Energien. URLhttp://www.erneuerbare-energien.de/die-themen/gesetze-verordnungen/erneuerbareenergien-gesetz/eckpunkte-der-eeg-novelle, German Ministry for the Environment, Nature Conservation, and Nuclear Safety (BMU)
11. BMU (2013b) Entwicklung der erneuerbaren Energien in Deutschland im Jahr 2012. Technical report German Ministry for the Environment, Nature Conservation, and Nuclear Safety (BMU)
12. Böhringer C, Rosendahl KE (2010) Green promotes the dirtiest: on the interaction between black and green quotas in energy markets. J Regul Econ 37(3):316–325
13. Branker K, Pathak M, Pearce J (2011) A review of solar photovoltaic levelized cost of electricity. Renew Sustain Energ Rev 15(9):4470–4482
14. Breyer C, Gerlach A (2010) Global overview on grid-parity event dynamics. In: Proceedings of the 25th EUPVSEC/ WCPEC-5. Valencia
15. Butler L, Neuhoff K (2008) Comparison of feed-in tariff, quota and auction mechanisms to support wind power development. Renew Energ 33(8):1854–1867
16. California Energy Commission (2013) California Electricity Statistics & Data. URL http://energyalmanac.ca.gov/electricity/electric_generation_capacity.html
17. Chao H (2011) Efficient pricing and investment in electricity markets with intermittent resources. Energy Pol 39(7):3945–3953

18. Cutler NJ, Boerema ND, MacGill IF, Outhred HR (2011) High penetration wind generation impacts on spot prices in the Australian national electricity market. Energy Pol 39(10):5939–5949
19. Dinica V (2006) Support systems for the diffusion of renewable energy technologies?an investor perspective. Energy Pol 34(4):461–480
20. DSIRE (2013) Database of state incentives for renewables and efficiency. URL http://www.dsireusa.org
21. EEG/KWK-G (2013) EEG-Umlage. URL http://www.eeg-kwk.net/de/EEG-Umlage.htm, informationsplattform der deutschen Übertragungsnetzbetreiber
22. European Commission (2009) Directive 2009/28/EC of the European Parliament and of the Council of 23 April 2009 on the promotion of the use of energy from renewable sources and amending and subsequently repealing Directives 2001/77/EC and 2003/30/EC (Text with EEA relevance). Official J Eur Union 562009 L 140:16–60
23. Eurostat (2013) Energy database. URL http://epp.eurostat.ec.europa.eu
24. Fagiani R, Barquín J, Hakvoort R (2013) Risk-based assessment of the cost-efficiency and the effectivity of renewable energy support schemes: Certificate markets versus feed-in tariffs. Energy Pol 55:648–661
25. Giebel G (2011) The state-of-the-art in short-term prediction of wind power: A literature overview. Technical report, SafeWind
26. Graf C, Wozabal D (2013) On the efficiency of the epex day-ahead spot market. Technical report, University of Vienna
27. Green R, Vasilakos N (2010a) Market behaviour with large amounts of intermittent generation. Energy Pol 38(7):3211–3220
28. Green R, Vasilakos N (2010b) The Long-Term Impact of Wind Power on Electricity Prices and Generating Capacity. Technical report, University of Birmingham, Department of Economics, Discussion Paper 11–09
29. Green R, Hu H, Vasilakos N (2011) Turning the wind into hydrogen: The long-run impact on electricity prices and generating capacity. Energ Pol 39(7):3992–3998
30. Haas R, Panzer C, Resch G, Ragwitz M, Reece G, Held A (2011) A historical review of promotion strategies for electricity from renewable energy sources in EU countries. Renew Sustain Energ Rev 15(2):1003–1034
31. Haghifam M, Omidvar M (2006) Wind farm modeling in reliability assessment of power system. In: Probabilistic Methods Applied to Power Systems, 2006. PMAPS 2006. International Conference on, pp 1–5
32. He Y, Hildmann M, Herog F, Andersson G (2012) Modeling the Merit Order Curve of the EEX Market. Accepted for publication in IEEE Transaction on Power Systems
33. Hiroux C, Saguan M (2010) Large-scale wind power in european electricity markets: Time for revisiting support schemes and market designs? Energy Pol 38(7):3135–3145

34. Hollands K, Huget R (1983) A probability density function for the clearness index, with applications. Sol Energ 30(3):195–209
35. Holttinen H (2005) Optimal electricity market for wind power. Energ Pol 33(16):2052–2063
36. Jacobsen HK, Zvingilaite E (2010) Reducing the market impact of large shares of intermittent energy in Denmark. Energ Pol 38(7):3403–3413
37. Jacobson M, Archer C (2012) Saturation wind power potential and its implications for wind energy. Proceedings of the National Academy of Sciences of the United States of America 109(39):15,679–15,684
38. Jin T, Tian Z (2010) Uncertainty analysis for wind energy production with dynamic power curves. In: 2010 IEEE 11th International Conference on probabilistic methods applied to power systems (PMAPS), pp 745–750
39. Jónsson T, Pinson P, Madsen H (2010) On the market impact of wind energy forecasts. Energ Econ 32:313–320
40. Karki R, Hu P, Billinton R (2006) A simplified wind power generation model for reliability evaluation. IEEE Trans Energ Convers 21(2):533–540
41. Keppley JM (2012) A Comparative Analysis of California and German Renewable Energy Policy: Actors and Outcomes. The Josef Korbel J AdvInternational Stud 4:1–26
42. Ketterer JC (2012) The impact of wind power generation on the electricity price in Germany. Ifo Working Paper Series Ifo Working Paper No. 143, Ifo Institute for Economic Research at the University of Munich
43. Klessmann C, Nabe C, Burges K (2008) Pros and cons of exposing renewables to electricity market risks—A comparison of the market integration approaches in Germany, Spain, and the UK. Energ Pol 36(10):3646–3661
44. Klessmann C, Rathmann M, de Jager D, Gazzo A, Resch G, Busch S, Ragwitz M (2013) Policy options for reducing the costs of reaching the european renewables target. Renew Energ 57:390–403
45. MacKay D (2009) Sustainable energy: without the hot air. UIT Cambridge Limited, Cambridge
46. Marvel K, Kravitz B, Caldeira K (2013) Geophysical limits to global wind power. Nat Clim Change 3(2):118–121
47. Menanteau P, Finon D, Lamy ML (2003) Prices versus quantities: choosing policies for promoting the development of renewable energy. Energ Pol 31(8):799–812
48. Mennel T (2012) Das Erneuerbare-Energien-Gesetz – Erfolgsgeschichte oder Kostenfalle? Wirtschaftsdienst 92(1):17–22
49. Milstein I, Tishler A (2011) Intermittently renewable energy, optimal capacity mix and prices in a deregulated electricity market. Energ Pol 39(7):3922–3927
50. Mitchell C, Bauknecht D, Connor P (2006) Effectiveness through risk reduction: a comparison of the renewable obligation in England and Wales and the feed-in system in Germany. Energ Pol 34(3):297–305
51. Munksgaard J, Morthorst PE (2008) Wind power in the Danish liberalised power market—Policy measures, price impact and investor incentives. Energ Pol 36(10):3940–3947

52. Nicholson E, Rogers J, Porter K (2010) The Relationship between Wind Generation and Balancing-Energy Market Prices in ERCOT: 2007–2009. Technical report National Renewable Energy Laboratory (NREL)
53. Nicolosi M (2012) The Economics of Renewable Electricity Market Integration. An Empirical and Model-Based Analysis of Regulatory Frameworks and their Impacts on the Power Market. Ph.D thesis, University of Cologne
54. Philibert C (2011) Interactions of policies for renewable energy and climate. IEA Energy Papers 2011/6, OECD Publishing
55. Pigou AC (1932) The Economics of welfare, 4th edn. Macmillan, London
56. Pineau PO, Rasata H, Zaccour G (2011) Impact of some parameters on investments in oligopolistic electricity markets. Eur J Oper Res 213(1):180–195
57. Pirrong C, Jermakyan M (2008) The price of power: The valuation of power and weather derivatives. J Bank Finance 32(12):2520–2529
58. Ragwitz M, Winkler J, Klessmann C, Gephart M, Resch G (2012) Recent developments of feed-in systems in the EU – A research paper for the International Feed-In Cooperation. Technical report German Ministry for the Environment, Nature Conservation, and Nuclear Safety (BMU)
59. Redl C, Haas R, Huber C, Böhm B (2009) Price formation in electricity forward markets and the relevance of systematic forecast errors. Energ Econ 31(3):356–364
60. REN21 (2012) Renewables 2012: Global status report
61. res-legal (2013) Renewable energy policy database and support. URL http://www.res-legal.eu/
62. Ringel M (2006) Fostering the use of renewable energies in the European Union: the race between feed-in tariffs and green certificates. Renew Energ 31(1):1–17
63. Schleicher S, Tappeser R (2012) How renewables will change electricity markets in the next five years. Energ Pol 48:64–75
64. Sensfuß F, Ragwitz M, Genoese M (2008) The merit-order effect: A detailed analysis of the price effect of renewable electricity generation on spot market prices in Germany. Energ Pol 36(8):3086–3094
65. Skantze P, Gubina A, Ilic M (2000) Bid-based stochastic model for electricity prices: the impact of fundamental drivers on market dynamics. MIT e-lab report
66. Steggals W, Gross R, Heptonstall P (2011) Winds of change: How high wind penetrations will affect investment incentives in the GB electricity sector. Energ Pol 39(3):1389–1396
67. Stokes LC (2013) The politics of renewable energy policies: The case of feed-in tariffs in Ontario, Canada. Energ Pol 56:490–500
68. Taylor J, McSharry P (2007) Short-term load forecasting methods: An evaluation based on european data. IEEE Trans Power Syst 22(4):2213–2219
69. The Economist (2013) Carbon trading: ETS, RIP? The Economist
70. Tina G, Gagliano S (2008) Probability analysis of weather data for energy assessment of hybrid solar/wind power system. In: Proceedings of 4th IASME/WSEAS International Conference on Energy, Environment, Ecosystems and Sustainable Development. EEESD08. pp 217–223

71. Tinbergen J (1952) On the theory of economic policy. North Holland, Amsterdam, Chapters 4 and 5
72. Traber T, Kemfert C (2011) Gone with the wind? – Electricity market prices and incentives to invest int thermal power plants under increasing wind energy supply. Energ Econ 33(2):249–256
73. Troy N, Denny E, O'Malley M (2010) Base-Load Cycling on a System With Significant Wind Penetration. IEEE Trans Power Syst 25(2):1088–1097
74. US Energy Information Administration (2013) Updated capital cost estimates for utility scale electricity generating plants. Technical report, EIA
75. Vandezande L, Meeus L, Belmans R, Saguan M, Glachant JM (2010) Well-functioning balancing markets: A prerequisite for wind power integration. Energ Pol 38(7):3146–3154
76. Ventosa M, Baíllo A, Ramos A, Rivier M (2005) Electricity market modeling trends. Energy Policy 33(7):897–913
77. Weber C, Woll O (2007) Merit-Order-Effekte von Erneuerbaren Energien - Zu schön um wahr zu sein?, eWL Working Paper No. 01/07
78. Weitzman ML (1974) Prices vs. quantities. The Rev Econ Stud 41(4):477–491
79. Wiser R, Lantz E, Bolinger M, Hand M (2012) Recent Developments in the Levelized Cost of Energy from U.S. Wind Power Projects. Technical report, National Renewable Energy Laboratory
80. Wissen R, Nicolosi M (2008) Ist der Merit-Order-Effekt der erneuerbaren Energien richtig bewertet? Energiewirtschaftliche Tagesfragen 58(1):110–115
81. Woo C, Horowitz I, Moore J, Pacheco A (2011) The impact of wind generation on the electricity spot-market price level and variance: The Texas experience. Energ Pol 39(7):3939–3944
82. Woo CK, Horowitz I, Horii B, Orans R, Zarnikau J (2012) Blowing in the wind: vanishing payoffs of a tolling agreement for natural-gas-fired generation of electricity in texas. The Energ J 33(1):207–230
83. Woodman B, Mitchell C (2011) Learning from experience? The development of the renewables obligation in England and Wales 2002?2010. Energ Pol 39(7):3914–3921
84. Youcef Ettoumi F, Sauvageot H, Adane AEH (2003) Statistical bivariate modelling of wind using first-order markov chain and weibull distribution. Renew Energ 28(11):1787–1802

Chapter 13
Copula-Based Hedge Ratios for Renewable Power Generation

Audun Nordtveit, Kim T. Watle, and Stein-Erik Fleten

Abstract The electricity price and production volume determine the revenue of a renewable electricity producer. Feed-in variations to power plants and high price volatility result in significant cash flow uncertainty. A copula-based Monte Carlo model is used to relate price and production volume and to find optimal hedge ratios through minimization of risk measures such as variance, hedge effectiveness, cash flow at risk, and conditional cash flow at risk. In our case study, all risk measures argue for an optimal hedge ratio between 35 and 60% of expected production. The highest risk reduction is achieved by the use of forward contracts with long time to maturity but at the expense of a low risk premium. Conversely, short-term futures and forwards only provide marginal risk reduction, but can yield attractive positive risk premiums. These findings underline the importance of distinguishing the use of derivative contracts for speculation and hedging purposes, through positions in short-term and long-term contracts, respectively.

13.1 Introduction

With increasing use of renewable sources in the deregulated electricity markets, power producers are faced with production volume risk caused by varying feed-in. This comes in addition to price risk. In this chapter we develop a copula-based approach to the simultaneous price and production risk for renewable electricity producers.

A. Nordtveit
Department of Industrial Economics and Technology Management, Norwegian University of Science and Technology (NTNU), Trondheim, Norway
e-mail: Audun.nordtveit@gmail.com

K.T. Watle • S.-E. Fleten (✉)
NTNU, Trondheim, Norway
e-mail: kim.thomassen.watle@gmail.com; stein-erik.fleten@iot.ntnu.no

R.M. Kovacevic et al. (eds.), *Handbook of Risk Management in Energy Production and Trading*, International Series in Operations Research & Management Science 199,
DOI 10.1007/978-1-4614-9035-7_13,

We will use a Norwegian hydropower producer as a case; however, the analysis is general enough to be relevant for, e.g., a wind power producer in Spain or solar power in Germany.

Copula is a statistical tool which has recently received much attention in the financial literature and is popular in practice. Genest, Gendron [20] show that from 2000 to 2005 the number of documents published on copula theory per year increased by a factor of nine. According to their survey, the financial industry is by far the field where copulas have been applied most frequently, due to their advantages in modeling non-normal returns and dependency between extreme values of assets.

It is interesting to extend the copula approach from its traditional financial applications to commodity markets. In addition, it can be of interest for a renewables producer to have an alternative financial approach to the traditional optimization method for risk management purposes.

It would be great to be able to report that copula-based analysis provides significant changes in hedge ratios compared to benchmark methods. This is not the case in our study, however; calculations using historical data (not copulas) recommend about the same hedge ratios, which hover around 50% across different maturities. That is, the recommendation is to sell of around half of the expected future production, using month, quarter, and year contracts. This cuts the risk in half, using variance and value-at-risk-related measures.

This chapter proceeds along the following lines: Sect. 13.2 treats more thoroughly how risks faced by hydropower producers can be measured, modeled, and managed. In Sect. 13.3 the hedge ratios obtained from historical price and production data are considered. The derivation of the copula-based Monte Carlo model is explained in Sect. 13.4. Hedge ratio results from the simulation for various risk measures are then obtained and discussed in Sect. 13.5. Finally, Sect. 13.6 concludes.

13.2 Market and Institutional Background

Price and production volume are identified to be the main risk factors faced by hydropower producers. Measurement and management of these risks are first discussed. Subsequently, important elements in hedging decisions such as taxation questions and risk premium are treated. Finally, the copula framework used to connect the two identified risk factors is presented.

13.2.1 Measuring Operational Risks

Variance in return, value at risk (VaR), and conditional value at risk (CVaR) [29] are risk measures commonly used by financial companies, but have also been introduced in nonfinancial firms and in the commodity literature. These risk measures are often used to evaluate and find optimal hedging strategies. Fleten et at. [18] consider a hydropower producer and use VaR, CVaR, and standard deviation of the producer's

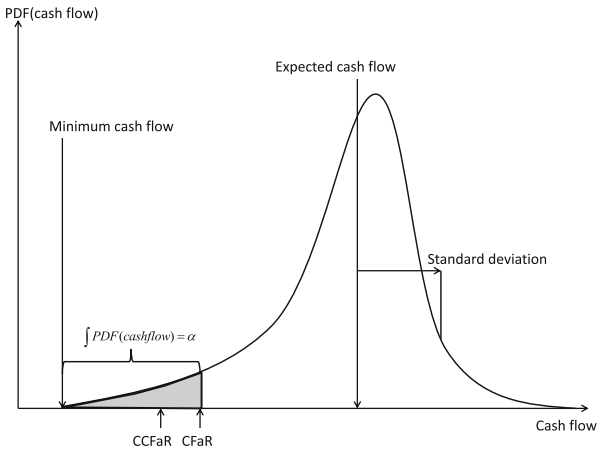

Fig. 13.1 Illustration of standard deviation in cash flow, CFaR and CCFaR. PDF (cash flow) represents the probability density function of a cash flow distribution. From the figure it appears that tail events affect standard deviation, CFaR, and CCFaR differently. Long, fat tails will not affect the standard deviation a lot, but the CFaR and especially the CCFaR will take much lower values for these extreme scenarios. CFaR and CCFaR are thus good measures for the downside risk

revenue as risk measures to obtain optimal hedging positions. VaR is also used as a risk measure in [35] to find a producer's optimal power portfolio. Unger [34] and Kettunen et al. [22] use CVaR.

The variance approach is relatively easy to implement in a model where a hydropower producer's cash flow volatility depends on the price risk and production uncertainty. Ederington's hedging effectiveness measure, e, defined in (13.1), can be used for a comprehensive comparison of the variance reduction achieved in hedged power portfolios with different hedge ratios to the variance of an unhedged portfolio [14].

$$e = \frac{Var(U) - Var(H)}{Var(U)} = 1 - \frac{Var(H)}{Var(U)} \tag{13.1}$$

In (13.1), $Var(U)$ and $Var(H)$ is the variance of the unhedged and hedged positions, respectively. The Ederington hedging effectiveness measure gives the percentage reduction in variance achieved by the hedged portfolio. One shortcoming of the variance risk measure is that it may give misleading results for asymmetrical and non-normal distributions which are common in power portfolios [28]. This results in higher possibilities of extreme undesirable outcomes in the cash flow.

CFaR and CCFaR[1] are based on the VaR framework and measure the downside risk in the cash flow. They may therefore be better suited than variance to describe risk for asymmetrical distributions. $CFaR_\alpha$ is defined as the α-quantile of the distribution of the cash flow, π. Thus, α is the confidence level as represented in (13.2). CFaR and CCFaR are illustrated in Fig. 13.1.

$$\alpha = Prob(\pi \leq CFaR_\alpha), \tag{13.2}$$

[1] Cash flow at risk and conditional cash flow at risk.

Standard values of acceptable cash flow threshold values are $\alpha = 1$ %, 5%, or 10%. The $CFaR_\alpha$ represents the threshold cash flow value such that α % of possible cash flow outcomes over a given time horizon are equal or below this value. The choice of the threshold value, α, reflects the risk aversion of a company. By reducing α a firm is more reluctant to accept uncertainty in its cash flow. The appropriate α value for a hydropower producer will be elaborated further in Sect.13.5.4. CCFaR is used to measure the expected value of the cash flow when it is known to be equal or lower than the $CFaR_\alpha$ value. The definition of CCFaR is given in (13.3):

$$CCFaR_\alpha = E[\pi | \pi \leq CFaR_\alpha] \tag{13.3}$$

13.2.2 Available Hedging Instruments

The system spot price at Nord Pool is the price obtained in a supply-demand equilibrium in the market without considering transmission grid congestions and capacity constraints. Transmission bottlenecks give rise to different local zonal market prices. Hedging price risk by using futures is a well-discussed topic in the literature [15, 19]. In addition to futures, a power producer can use several other derivative contracts to hedge the price risk, including contracts for difference (to hedge zonal price risks), swing contracts, and options.

Futures and forwards in the electricity market differ from the contracts in the financial market, since they are delivered over a period instead of on a specific day. These power derivatives are therefore comparable to financial swaps [6]. Both futures and forwards traded at the Eltermin market are standardized contracts with denomination in EUR per MWh, and the system spot price is the underlying of these contracts. Futures contracts consist of daily and weekly agreements and are rolling contracts for the next 6 weeks [27]. They are marked-to-market with daily settlement of the change in the market price in the trading period. The difference between the price on the last trading day, called closing price, and the system spot price is used to calculate the settlements in the delivery period. Forward contracts are settled in the same way as futures, but have no marked-to-market settlement in the trading period. Profits/losses are accumulated in the trading period and realized when the delivery period ends. Due to no margin requirements prior to delivery, the liquidity of these long-term contracts is higher than the liquidity of futures [9]. Forward contracts have monthly, quarterly, and yearly delivery periods.

Contracts for differences are the third type of derivative contracts traded at Nord Pool. These agreements are used to hedge the price differences between local areas and the system spot price caused by congestions in the transmission grid. A hydropower producer sells the electricity for the local area price, which not necessarily equals the system spot price. Thus hedging with just swaps will not eliminate all price risk. By using CfDs in combination with swaps it is possible to create a perfect price hedge. The liquidity of these contracts is however low and only traded for five of the thirteen local areas in the Nordic power market. Options available at Nord

Pool are European-style calls and puts with quarterly and annual forward contracts as the underlying. These contracts are useful because they offer several strategies to hedge variation in prices [27]. Nevertheless, options are not widely traded and may be expensive to use in hedging policies due to high transactions costs. For a more thorough description of different derivative instruments used in electricity risk management see [12].

Sanda et al. [30] analyzed the hedging policies of twelve different Norwegian hydropower firms. According to their study futures and forward contracts have the highest traded volume and are the most commonly used hedging derivatives. These findings and the low liquidity in both CfDs and options argue for the consideration of only swaps for hedging decisions in this chapter. Still, these products do not necessarily give a perfect price hedge alone.

13.2.3 Taxation Influences the Hedging Decision of a Hydropower Producer

Hydropower producers in Norway are subjected to four different taxes: income tax, resource rent tax, natural resource tax, and property tax. The resource rent tax is 30% of spot revenues for plants of a certain age, and it is directly determined by the spot price. As a result, the resource rent tax may influence the hedging strategy of a producer since deviations between the spot price and the hedged price are transformed into a relative tax gain or loss. The resource rent tax is calculated from the power plants' production sales value individually, where operating costs, concession costs, property tax, depreciation costs, and a nontaxed revenue are deduced from the calculated revenue.

Other taxes are less sensitive to hedging decisions in the sense that they are either fixed, as the natural resource tax of 13 NOK/MWh of the average production over the last seven years, or calculated as a percentage of the revenue such as the income tax and the property tax of 28% and 0.2–0.7%, respectively. Since the property tax is deductible from the resource rent tax, the total tax paid by an unhedged producer is $28\% + 30\% = 58\%$ of the sales value when costs are ignored. For a hedged producer this number is somewhat different dependent on its hedging performance and hedging level.

A cash flow after tax portfolio model for Norwegian hydropower producers, which utilizes swaps to hedge price risk and includes taxation issues, can be developed to find optimal hedge ratios. The revenue after tax of the hedged portfolio, Π, is defined in (13.4), but neglects the variable and fixed costs faced by hydropower producers. The transaction and margin costs in trading swaps are also ignored.

$$\Pi = [(P - H\bar{P})S + H\bar{P}F](1 - T_C) - PST_{RR} \tag{13.4}$$

In (13.4), P represents the actual production volume, $\bar{P}$ the expected production volume, S the spot price, F the swap price, H the hedge ratio, and T_C and T_{RR} are corporate and resource rent tax, respectively.

The variance in profit after tax of a hedged portfolio is given by (13.5). The $Var(F)$ term is set equal to zero since the swap price is locked when a producer enter a swap agreement.

$$\begin{aligned} Var[\Pi] =& (1-T_C)^2[Var(PS) + (H\bar{P})^2 Var(S)] \\ &+ (T_{RR})^2 Var(PS) \\ &+ 2(1-T_C)T_{RR}[H\bar{P}Cov(PS,S) - Var(PS)] \\ &- 2(1-T_C)^2 H\bar{P}Cov(PS,S) \end{aligned} \tag{13.5}$$

The risk reduction achieved in variance in revenue after tax depends on the chosen hedge ratio, H, of the individual hydropower producer. By minimizing (13.5) with respect to H, the optimal hedge ratio, H^*, is obtained:

$$\begin{aligned} &\frac{\partial Var(\Pi)}{\partial H} = 0 \\ &\rightarrow H^* = (1 - \frac{T_{RR}}{1-T_C})\frac{Cov(PS,S)}{\bar{P}Var(S)} \end{aligned} \tag{13.6}$$

By assuming no uncertainty in the production volume, $E[P] = \bar{P} \rightarrow Cov(PS,S) = \bar{P}Var(S)$, the hedge ratio expression in (13.6) simplifies to (13.7):

$$H^*_{Tax-neutral} = 1 - \frac{T_{RR}}{1-T_C} \tag{13.7}$$

The hedge ratio $H^*_{Tax-neutral}$ developed in (13.7) states that hydropower producers should hedge 58.3% of their expected production volume. Sanda et al. [30] derive the same hedge ratio for a Norwegian hydropower producer, which means that 58.3% of expected production must be sold in derivative contracts to obtain a fully hedged power portfolio.

One shortcoming with (13.5) to (13.7) is that the variance in swaps is set equal to zero, thus neglecting the possible effect of these contracts' term structure on the variance. To deal with this shortcoming, one might generate price and production scenarios and use (13.4) directly to measure the risk in the resulting cash flow scenarios.

13.2.4 Effects of Hedging Strategies for Hydropower Producers

Norwegian hydropower producers experience a negative relationship between electricity prices and production and pay a resource rent tax on spot revenues. These factors reduce and set an upper bound for the optimal hedging level well below 100%, as shown for the tax-neutral portfolio in Sect. 13.2.3.

Fleten et al. [18] argue that "the main reason for the negative correlation between price and hydropower production in the Norwegian market is that the market is regional, and 99% of the electricity production comes from hydropower." The inflow to the water reservoirs is the main factor determining the production volume, and reservoir inflow depends on precipitation. Local precipitation is correlated with national precipitation, so periods with high water reservoir levels or water reservoir shortages often occur synchronously for all hydropower companies in Norway [18]. Dry and cold or wet and warm periods often tend to coincide within the Nordic countries. Electricity consumption depends on the need for residential heating in Norway. Consequently, the demand for power by customers and production willingness among producers often mismatch. Thus, price and production tend to be negatively correlated. The negative correlation works as a natural hedge and decreases the hydropower producers' variance in revenue. Further, this limits their incentive to invest in derivative contracts to hedge price risk.

Hydropower producers' hedging policies vary with their risk aversion, with risk averse producers hedging large parts of their expected production. Multiple optimization methods have been developed using both static and dynamic hedging approaches to investigate different hedging strategies and find optimal hedge ratios. Fleten et al. [18] develop an optimization model to examine the performance of static hedge positions for hydropower producers. They find that the use of forwards to hedge price risk significantly reduces the revenue risk with just a minor decrease in revenue. It is also shown that hedging costs are higher when producers uses contracts with long time to maturity.

Sanda et al. [30] find evidence of an extensive risk management practice among Norwegian hydropower companies. An interesting discovery is that hedging reduces the downside risk in cash flow, measured by CFaR, in ten out of twelve firms. Surprisingly, derivative investments contribute significantly to the firms' profit without any substantial decrease in cash flow variance. This finding is explained by the prevalent use of selective hedging, meaning incorporating own market views in hedging decisions.

13.2.5 Connection Between Electricity Spot and Swap Prices

Electricity is a non-storable commodity, and therefore the usual cost-of-carry relationship in finance is not applicable [7, 10, 23, 24]. The risk premium approach has emerged as a method to investigate the spot-forward price relationship. Fama, French [17], Longstaff , Wang [23] and Adam , Fernando [1] define the risk premium as in (13.8):

$$R(t,T) = F(t,T) - E_t[S(T)] \tag{13.8}$$

where $F(t,T)$ is the forward price at time t with delivery at time T, $E_t[S(T)]$ is the expected electricity spot price at time T, and $R(t,T)$ is the risk premium. According to [23] the forward risk premium represents "the equilibrium compensation for

bearing the price and/or demand risk for the underlying commodity." The literature treating this topic has shown that the risk premium sign does not need to be strictly negative [7, 21, 23]. A motivation for including a risk premium approach in hedging strategy decisions is to benefit from the possible positive risk premiums and hence the excess return such contracts can provide [1]. For a more thorough examination of risk premiums in commodity markets see [17].

Botterud et al. [8] use the risk premium approach to examine the relationship between the spot and futures prices in the Nordic electricity market from 1995 to 2001. They explain the sign of the risk premium by the risk aversion and flexibility of both buyers and sellers. Hydropower producers are able to quickly ramp production, allowing them to take advantage of the market price fluctuations by adjusting their generation. The attractiveness of fixing the price by using futures for hedging all of the expected production is therefore reduced. At the same time, the production flexibility enables producers to profit from price peaks in the spot market. On the other hand, the demand side has limited ability to adjust demand with respect to spot price changes. As a consequence it may be attractive to fix the price for expected future demand in order to reduce the negative effect of large price spikes. Botterud et al. [8] find that futures prices on average have been higher than spot prices in the period of 1995 to 2001, which according to (13.8) gives a positive risk premium and in this way contradicts the classical literature. They pinpoint that the results should be treated with caution due to the limited data available in the electricity market.

Lucia and Torró [25] examine the sign and size of the risk premium in the Nordic electricity market between 1998 and 2007. They find that risk premiums on average are positive and vary throughout the year. Positive risk premiums are observed for contracts in periods where demand is high, such as during autumn and winter. This result is in accordance with the equilibrium model of [7]. They also find significant evidence of a structural break in the prediction power of this model in the Nord Pool market after the winter 2002–2003.

13.2.6 Copula, a Tool to Link Price and Production

Correlation is a key factor in risk management as risk generally is the result of both the variance of individual variables and their covariance. As an example the risk in a portfolio of stocks is dependent on not only the individual variance of the shares but also how they tend to covariate. Analogously, most of the risk in the revenue of a hydropower supplier stems from the individual risk of the price and the production volumes, and how these covariate. Historically the most popular way to describe covariance between two or more variables have been the Pearson correlation coefficient, ρ, explained in [3]. This coefficient is a simple and exact measure for covariance between elliptically distributed variables, but as distributions get more non-normal, skewed, heavy-tailed, and tail-dependent, the correlation coefficient tend to underestimate risk [16].

Copulas represent a new way to describe the dependency structure of the covariance between distributions and were introduced by [31]. He showed that every joint distribution can be written as in (13.9) where C is a copula and $F_1(x_1),\ldots,F_n(x_n)$ are cumulative probabilities of the variables $x_1,\ldots,x_n$. The mostly used copulas are bivariate, and a bivariate function must satisfy four properties to qualify as a two-dimensional copula. These are listed in (13.10) and explained thoroughly in [2].

$$F(x_1,\ldots,x_n) = C(F_1(x_1),\ldots,F_n(x_n)) \tag{13.9}$$

$$\begin{aligned}
&1)\quad C:[0,1]\times[0,1]\rightarrow[0,1]\\
&2)\quad C(u_1,0)=C(0,u_2)=0\\
&3)\quad C(u_1,1)=u_1 \text{ and } C(1,u_2)=u_2\\
&4)\quad C(v_1,v_2)-C(u_1,v_2)\geq C(v_1,u_2)-C(u_1,u_2)\ \forall\ u_1,u_2,v_1,v_2\in[0,1],\\
&\quad \text{with } u_1\leq v_1 \text{ and } u_2\leq v_2
\end{aligned} \tag{13.10}$$

There exist a large number of functions C defined in (13.9), satisfying the properties of a bivariate copula listed in (13.10). These functions have different dependency structure and can therefore be adapted to various problems requiring a more flexible tool than the linear correlation coefficient. Copula functions have parameters that need calibration to provide an optimal fit to the data. The estimation of the copula parameters is usually done by a maximum likelihood estimation of the joint distribution of the dependent variables. Once the likelihood value is obtained, the best copula can be selected based on an information criterion such as the Akaike information criterion (AIC) or Bayesian information criterion (BIC). If the existing families of copulas provide an unsatisfying fit to the data an alternative approach could be to implement an empirical copula. Alexander [2] presents a straightforward way to create the empirical copula following (13.11). In (13.11) $\hat{C}$ is the cumulative copula function, $\hat{c}$ is the density function, T is the number of observations, and x and y are the two dependent variables. For a more thorough study of the copula framework see [33].

$$\begin{aligned}
\hat{C}\left(\frac{i}{T},\frac{j}{T}\right) &= \frac{\text{Number of pairs } (x,y) \text{ such that } x\leq x^{(i)} \text{ and } y\leq y^{(j)}}{T}\\
\hat{c}\left(\frac{i}{T},\frac{j}{T}\right) &= \left\{\begin{array}{cc} T^{-1}, & \text{if } (x^i,y^j) \text{ is an element of the sample,}\\ 0, & \text{otherwise}\end{array}\right\}
\end{aligned} \tag{13.11}$$

Following (13.11) one obtains an empirical copula density function, $\hat{c}$, and cumulative distribution function, $\hat{C}$, for the joint densities as illustrated in Tables 13.1 and 13.2, respectively.

Copulas have not yet been given much attention in the nonfinancial literature, and the use of copulas in risk modeling for electricity suppliers in the Nordic power market is not an exception. So far, copulas have mainly been applied to commodity

Table 13.1 An example of the empirical copula density function, $\hat{c}$, calculated from (13.11)

F(x)/F(y)	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
0.0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.1	0.000	0.009	0.015	0.009	0.009	0.015	0.012	0.006	0.009	0.006	0.009
0.2	0.000	0.006	0.021	0.000	0.003	0.009	0.018	0.009	0.009	0.006	0.015
0.3	0.000	0.012	0.018	0.009	0.006	0.003	0.015	0.012	0.003	0.009	0.006
0.4	0.000	0.009	0.018	0.003	0.003	0.018	0.006	0.003	0.012	0.015	0.021
0.5	0.000	0.003	0.006	0.015	0.006	0.015	0.009	0.021	0.009	0.009	0.009
0.6	0.000	0.003	0.006	0.009	0.006	0.015	0.003	0.012	0.015	0.009	0.012
0.7	0.000	0.006	0.009	0.021	0.015	0.009	0.006	0.012	0.012	0.018	0.009
0.8	0.000	0.012	0.012	0.009	0.006	0.006	0.012	0.009	0.012	0.012	0.009
0.9	0.000	0.009	0.003	0.009	0.009	0.018	0.012	0.009	0.009	0.012	0.012
1	0.000	0.003	0.021	0.012	0.015	0.012	0.009	0.015	0.006	0.015	0.000

The first row and column are cumulative probabilities for the two dependent variables x and y. The table illustrates the joint probability density function, and areas with many high densities represent scenarios that are likely to occur. Conversely, areas with many zeros represent unlikely situations

markets to determine the spark spread [5]. Still, there are several reasons to believe that copulas will have the ability to describe the dependency structure between price and production in a better way than a linear correlation coefficient. Risks faced by hydropower producers have several characteristics in common with risks encountered in traditional financial applications. First, electricity prices are far from normally distributed. Second, one could expect a strong tail dependency between price and production. High prices often occur during cold winters with high production despite low production willingness due to low reservoir levels. Low prices are common during wet periods where producers generate as much as they can to reduce the risk of spillage. Thus, a copula's advantage in modeling non-normal distributions and dependency between extreme values seems like a desirable feature in hydropower risk management.

Finally we note that copulas are not a panacea in risk management. A natural alternative, favored by most electricity companies, is using a fundamental (bottom-up) model to capture the relationship between local production and local price. The advantage of such an approach includes the possibility to consider increased renewable penetration over time.

13.3 Hedge Ratios Obtained from Historical Data

The purpose of this chapter is to examine optimal swap hedging strategies for hydropower producers to reduce risks. It is therefore of interest to investigate the historically optimal hedge ratios. These historical hedging levels can be used as benchmarks for the theoretically obtained hedge ratios from the model later in this chapter. Historical spot and swap prices along with production volumes for a Norwegian hydropower producer are considered from 2006 to 2010 on a weekly basis.

Table 13.2 An example of the empirical cumulative copula function, $\hat{C}$

F(x)/F(y)	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.1	0.00	0.01	0.02	0.03	0.04	0.06	0.07	0.07	0.08	0.09	0.10
0.2	0.00	0.01	0.05	0.06	0.07	0.09	0.12	0.14	0.16	0.17	0.20
0.3	0.00	0.03	0.08	0.10	0.12	0.14	0.19	0.21	0.23	0.25	0.30
0.4	0.00	0.04	0.11	0.13	0.15	0.19	0.24	0.27	0.30	0.34	0.40
0.5	0.00	0.04	0.12	0.15	0.18	0.24	0.30	0.35	0.39	0.43	0.50
0.6	0.00	0.04	0.12	0.17	0.20	0.28	0.34	0.40	0.46	0.51	0.60
0.7	0.00	0.05	0.14	0.20	0.25	0.33	0.40	0.48	0.54	0.62	0.70
0.8	0.00	0.06	0.16	0.24	0.29	0.38	0.46	0.54	0.62	0.70	0.80
0.9	0.00	0.07	0.17	0.26	0.32	0.43	0.52	0.61	0.70	0.79	0.90
1.0	0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00

The first row and column are cumulative probabilities for the two dependent variables x and y. Note that $\hat{C}$ is calculated by (13.11) and follows the conditions in (13.10)

Table 13.3 summarizes what are found to be the optimal static hedge ratios and how to optimally invest in selected swap contracts with one week, one month, one quarter, and one year to delivery, in order to minimize the risk in the 2006 to 2010 period. Variance is minimized; $CFaR_{5\%}$ and $CCFaR_{5\%}$ are maximized and compared with the natural hedge situation. The natural hedge is the same as selling all production in the spot market. For the obtained hedge ratios, it is assumed that a rolling investment in the front contract is taken. A 10% investment in weekly contracts would therefore imply a sale of 10% of next week's expected production in weekly contracts each Friday from 2006 to 2010.

The first row in Table 13.3 presents the expected cash flow of each strategy compared with the natural hedge case. The cash flow of the unhedged scenario is therefore 100%. When the other risk measures are considered a cash flow of 95.9%, 98.3%, and 99.3% of the unhedged return is obtained for minimum variance, maximum $CFaR_{5\%}$, and maximum $CCFaR_{5\%}$, respectively. As all expected cash flow values for the risk measures are below 100% there are costs associated with hedging. Variance in expected revenue is illustrated on the second row. As before, the minimized risk measures' variance is compared with the unhedged variance. The variance is thus reduced by hedging. Hedge effectiveness illustrates the same as the variance and represents the percentage decrease in variance of each strategy compared to the unhedged case. In this way the sum of the variance and hedge effectiveness row is 100% for each column. $CFaR_{5\%}$ and $CCFaR_{5\%}$ are maximized on row four and five, respectively, and the listed numbers illustrate the percentage of the unhedged expected cash flow the $CFaR_{5\%}$ and $CCFaR_{5\%}$ attain. For example, the $CFaR_{5\%}$ value of 44.6% for the natural hedge situation means that in 5 % of the outcomes the cash flow will be less or equal to 44.6% of the unhedged expected cash flow. The higher this value is, the better, since it represent the worst case cash flow. $CCFaR_{5\%}$ measure more extreme values than $CFaR_{5\%}$, so the percentage numbers for $CCFaR$ are lower. As seen in the table hedging reduces downside risk. The hedge ratios, (HR), in Table 13.3 represent the percentage of the expected

Table 13.3 Performance of several hedging strategies based on optimization of spot and swap contract prices from 2006 to 2010

	Natural hedge	Minimum variance	Maximum $CFaR_{5\%}$	Maximum $CCFaR_{5\%}$
Expected cash flow	100	95.9	98.3	99.3
Variance	100	62.2	68.9	78.9
Hedge effectiveness	–	37.8	31.1	21.1
$CFaR_{5\%}$	44.6	42.9	49.7	48.3
$CCFaR_{5\%}$	38.5	33.1	40.3	40.5
Hedge ratio (HR)	–	47.5	28.0	15.9
% of HR in 1WF	–	–	–	–
% of HR in 1MF	–	57.9	48.1	42.3
% of HR in 1QF	–	1.3	37.3	57.4
% of HR in 1YF	–	40.8	14.7	0.2

All numbers are in percent of the natural hedge situation. Hedging reduces the expected cash flow for a hydropower producer, but can provide risk protection observed by lower variance and higher hedge effectiveness, $CFaR_{5\%}$ and $CCFaR_{5\%}$. The optimal hedge ratio drops when measures that consider tail events are considered

production a producer should hedge to minimize the risk measure in question. It is specified how this hedging level should be allocated between weekly, monthly, quarterly, and yearly contracts. In this way the four last rows sum to 100 % for the different risk measures. The total investment in each contract is therefore the suggested hedge ratio multiplied with the percentage of the hedge in the contracts.

Examining Table 13.3 one observes that hedging might reduce risks at the expense of a slightly reduced cash flow. Depending on the considered risk measure, different optimal hedge ratios are obtained. The more a risk measure considers tail risk and extreme values, the lower the optimal hedge ratio is. Finally, it seems undesirable to invest in weekly swaps to eliminate risk. These results are not surprising as short-term swaps are more correlated to spot prices than long-term swaps and will therefore eliminate less risk. Also, it seems reasonable that risk measures that consider extreme events give lower hedge ratios. A producer will as an example incur a great loss if it is highly hedged when a price spike occurs. Although such events are rare, they will affect the $CFaR_{5\%}$ and even more the $CCFaR_{5\%}$ but only have a marginal effect on the variance.

In this historical analysis weekly risk is considered. Natural seasonal cash flow differences, due to price and production differences between winter and summer months, are attempted hedged away. The reason to consider weekly variations despite this obvious drawback is the short, five-year, time horizon of available swap data. For a hydropower producer, annual cash flow fluctuations are of greater interest than weekly variations. However, it is meaningless to investigate risk measures such as $CFaR_{5\%}$ and $CCFaR_{5\%}$ in a data set consisting of five observations.

With the historical optimal hedge ratios in mind it is time to develop a model that can provide data for a theoretical risk analysis.

13.4 Derivation of the Copula-Based Monte Carlo Model

A challenge in financial risk management is how to cope with non-normality of the distributions of risky variables and their interdependency. As discussed in Sect. 13.2.6, a copula framework will be developed to deal with some of the shortcomings of existing linear correlation models. Further, knowledge about the price-production dependency structure can be valuable for hedging decisions in order to define adequate hedge ratios and optimal use of the available derivative contracts.

To investigate and evaluate hedge ratios a copula-based Monte Carlo simulation approach is used to generate possible cash flow outcomes for a hydropower producer. Dependent electricity spot/swap prices, S/F, and production volumes, P, must be simulated to obtain the cash flow outcomes, since these factors are the only dynamic variables in the cash flow expression in (13.4). Figure 13.2 illustrates how a large sample of dependent prices and production volumes can be generated through a copula-based Monte Carlo simulation. The copula-based Monte Carlo simulation will be described in four steps for explanatory reasons. First, the input variables, production and price (model 1), to the empirical copula are treated in step 1. Then, in step 2, the construction of the empirical copula is elaborated. Subsequently, step 3 explains the generation of correlated cumulative probability values for price and production. Finally, the procedure of linking these cumulative probabilities to production values, spot, and swap prices is considered in step 4.

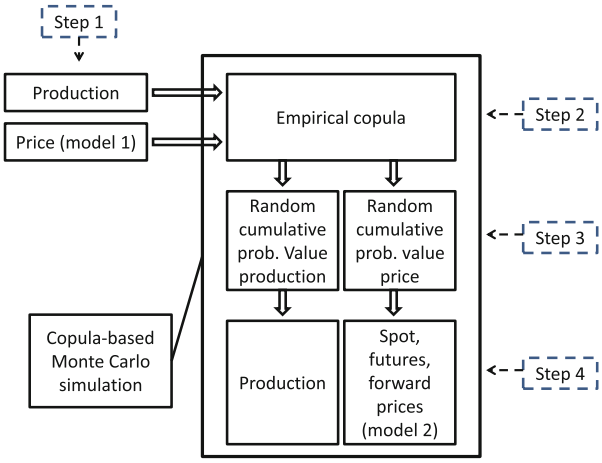

Fig. 13.2 An overview over the copula-based Monte Carlo model. Model 1 represents the spot price model of [4] which is used to generate spot prices. Together with historical production, the spot prices are used to construct an empirical copula. From the copula, a large sample of dependent cumulative probability values for price and production is randomly generated. The cumulative probabilities are then linked to production and spot/swap numbers. To connect spot/swap prices a new model is necessary since model 1, used for the input values, cannot be employed to estimate swaps with the available data. The two-factor model developed in [24] is therefore used, and constitutes model 2

13.4.1 Production and Price Input to the Empirical Copula

This section will treat the input variables, price and production, to the copula and represents step 1 in Fig. 13.2. An empirical copula requires a large sample of correlated data points to capture the existing dependency structure. Hence, long data series of price and production are necessary. Figure 13.3 depicts the historical average weekly system spot prices and average weekly production volumes for a Norwegian hydropower producer from 2000 to 2011.

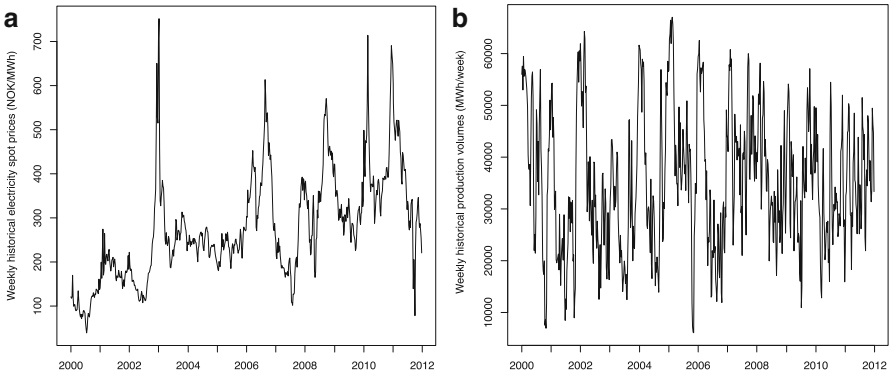

Fig. 13.3 Time series of weekly historical system spot prices and production volumes from 2000 to 2011. Prices are obtained from Nord Pool's ftp server and production volumes are received from a Norwegian hydropower producer. The figures reveal that neither the price nor the production is normally distributed and both functions seem to be extremely volatile and contain spikes. (**a**) Average weekly system spot prices (**b**) Average weekly production volumes

As the electricity price dynamics have changed over the years, a modified and simplified model following the work of [4] has been selected to estimate historical prices conserving the pricing dynamics observed today. Besides, this model enables an estimation of electricity prices going further back than the available market spot prices, satisfying the need of long data series for the empirical copula construction. The spot price model, model 1 in Fig. 13.2, is defined in (13.12) with deviation from normal accumulated reservoir levels (Δ_{H_t}), 12-month accumulated inflow (I_{12M}), and an inflation-adjusted oil product index ($O_{Adj.}$) as inputs. The data are obtained from the Norwegian Water Resources and Energy Directorate (NVE), Nord Pool, and Reuters EcoWin, respectively. Nord Pool also provided the spot prices used to calibrate this price model. To prevent estimated spot prices to fall too low, the oil index is adjusted according to the consumer price index. Seasonal load variations are also accounted for by inclusion of a sine function.

$$\begin{aligned}\ln(S(t)) =& \beta_0 + \beta_1 \sin\left(\frac{2\pi}{52}t + \phi\right) \\ &+ \beta_H \Delta_{H_t} + \beta_I I_{12M} + \beta_O O_{Adj.}\end{aligned} \tag{13.12}$$

All data are collected on weekly basis and span from 1986 to 2011, except the spot prices used for the 2005–2011 calibration period. In (13.12), the adjusted oil index and spot prices are transformed by the natural logarithm. Table 13.4 summarizes the descriptive data of the input variables. The data are observed to be non-normally distributed as the normality Jarque-Bera test is rejected for all factors included in Table 13.4 with a p-value of less than 0.001.

Table 13.4 Descriptive statistics for weekly input variables to the modified spot price model of [4] spanning from 1986 to 2011 (2005–2011 for the dependent variable $S(t)$)

Descriptive statistics	Δ Hydrobalance (GWh)	Average 12-month inflow (GWh)	Adjusted oil index	Spot prices (NOK/MWh)
Min	−26.87	90991	2.24	100.1
Max	25.51	153902	16.53	1403.7
Avg	−0.37	123471	6.21	334.7
Med	1.59	122222	4.98	307.2
St.dev.	9.84	13599	2.96	128.9
Skew	−0.41	0.10	1.10	2.14
Ex. kurt	−0.38	−0.57	0.26	11.25
JB	46.67	20.40	279	2524.7
No. of obs.	1357	1357	1357	418

Estimated coefficient values, obtained from the least sum of squares approach, are presented in Table 13.5. All coefficients are significant. From the coefficient's sign it is apparent that a negative hydrobalance deviation, representing low reservoir levels, leads to higher prices. Surprisingly, high yearly inflow has historically contributed to higher prices, which contradicts common sense. The influence of this variable can hence be questioned. However, as seen in the descriptive statistics, the product of the inflow coefficient and the range of the variable is only a third of the hydrobalance deviation effect, and it might therefore work as a counterweight. Finally, fuel costs represented by an oil index are as expected positively correlated to the spot price. Having the regression coefficients, weekly time series of electricity spot prices can be generated. Electricity prices are generated back to 1986, when the history of the underlying input variables starts.

The empirical copula requires a large number of data points to capture the dependency structure of the input variables. However, for a hydropower producer the annual variations in cash flow and hence the yearly dependency between the underlying variables is most interesting, since seasonal effects are expected and preferably should not affect the dependency structure of the copula. Although a 26-year history of data is estimated, yearly prices do not provide sufficient data points for a robust estimation. By comparing the autocorrelation in price and production it is seen that the autocorrelation is higher for prices than for the production. Therefore the prices must be considered in an autocorrelation analysis. Prices are highly autocorrelated (see Fig. 13.4), which set a lower bound to the frequency of the input

Table 13.5 Estimated coefficient values in (13.12)

Coefficient	Value	Stdev.	t-value
β_0	2.64	0.22	12.5
β_1	0.10	1.55e-2	6.5
ϕ	1.55	–	–
β_H	−3.71e-2	2.43e-3	−15.3
β_I	1.18e-5	1.75e-6	6.7
β_O	0.67	5.09e-2	13.1

The estimates are obtained by regressing Eq (13.12) on historical weekly spot prices from 2005 to 2011. β_1 underlines the presence of seasonal load variations. From β_H and β_O it appears that low reservoir levels and high fuel prices contribute to higher spot prices. The inflow to the reservoirs, β_I, seems to work as a counterweight to the reservoir levels as it results in lower spot prices. All values are significant. This model gives a $\bar{R}^2$ of 0.58

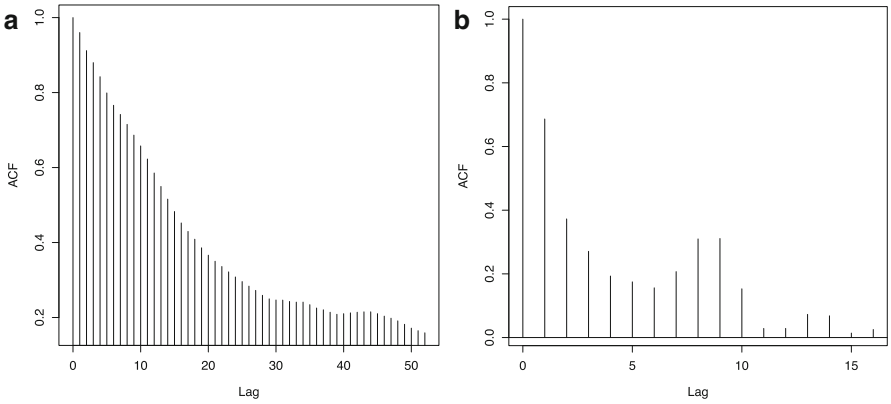

Fig. 13.4 Autocorrelation plots of historical spot prices from 2000 to 2011. Autocorrelation of weekly data is strong and persistent for many weeks. The autocorrelation from one quarter to the next is less prominent than for consecutive weeks, though the quarterly autocorrelation is still existent. Quarterly data are better suited as input to the empirical copula than weekly data. (**a**) Autocorrelation plot of average weekly historical spot prices (**b**) Autocorrelation plot of average quarterly historical spot prices

price data. Autocorrelated input to an empirical copula results in a dependency structure where some outcomes will have a much higher probability than in reality, which is clearly an undesirable feature. Weekly data should therefore be avoided as input to the empirical copula and one should strive to use low-frequency data to limit the negative effect of autocorrelation. If high-frequency data are selected, seasonality and autocorrelation will be problematic. Conversely, long-term average will not permit a well-fitted copula, due to the lack of data. For this reason, quarterly data are selected as input to the copula. In this way, the autocorrelation of the input price is reduced from the weekly resolution and a considerable number of data, 104 points, are used in the empirical copula calibration. Nevertheless, seasonal effects will still be present and result in more extreme variations in the output scenarios than would have been the situation if annual data were used. To exemplify, the range of

the output scenarios is wider for seasonal than for yearly data since high production/prices occurring during the winter can coexist with low production/prices from the summer. This is a shortcoming of the model.

Table 13.6 Descriptive statistics for price output of the modified spot price model developed in [4] together with historical observed production volumes from a Norwegian hydropower producer

	Price 13 weeks (NOK/MWh)	Prod. 13 weeks (GWh/quarter)
Min	77.25	17.18
Max	624.43	57.23
Avg	212.64	34.35
Med	189.18	34.14
St. dev.	102.02	8.88
Skew	1.24	0.37
Ex. kurt.	1.85	−0.23
JB	41.41	2.59
No. of obs.	104	104

The data set consists of 104 observations of quarterly data. These data will be used as input to construct the empirical copula

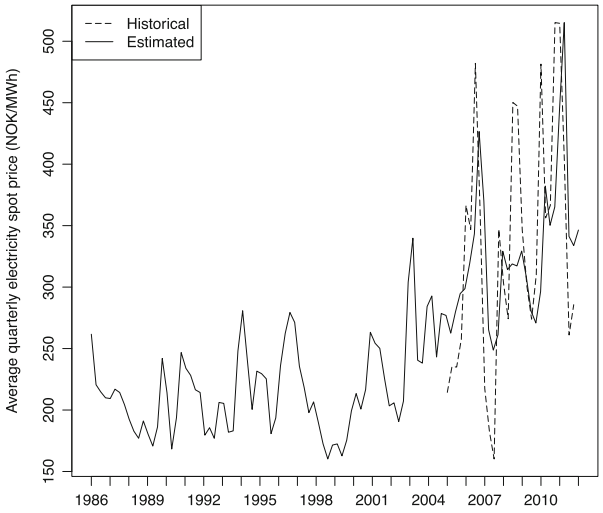

Fig. 13.5 An overview of the estimated spot prices from (13.12) and the actual realized spot prices in the 1986–2011 and 2005–2011 period, respectively. Prior to 2003, the estimated prices were at a significant lower level than in the 2003–2011 period. This is due to the level of the underlying variables to the price model. The price jump in the model bodes well with the detected structural break in [25]

The modeled weekly prices are converted into quarterly data and paired with quarterly historical production volumes for a Norwegian hydropower producer. Historical production volumes are used and not production plans, as [30] confirmed that historical average production is more accurate for predicting production. Descriptive data for the quarterly price and production used as input to the empirical copula is shown in Table 13.6. Comparing the modeled quarterly 1986–2011 price data in Table 13.6 with the weekly observed 2005–2011 price data in Table 13.4, it appears that the average of the quarterly data is lower than the weekly calibration data. The reason for this is the level of the input variables to the model which resulted in lower prices from 1986 to 2005 than from 2005 to 2011, and the difference in price average is therefore not surprising. Figure 13.5 illustrates this trend well, with electricity prices being low until 2003 where they suddenly increased. During the winter 2002–2003 there was a shock in the market and this may have shifted the price level and price behavior [25]. The estimated price model seems to capture this shift quite well. Also, the range of the quarterly data is narrower than that of the weekly data as quarterly average reduces the magnitude of spikes. Still, the minimum quarterly price is lower than the weekly price, and the minimum quarterly price was thus realized prior to 2005. Finally, the Jarque-Bera test, JB, underlines the non-normality of the input data, which further motivate the copula approach.

13.4.2 Construction of the Empirical Copula

With the input variables to the copula explained, the next step will be to create a copula to relate the dependency between price and production and this constitutes step 2 in Fig. 13.2.

There exist numerous predefined copula functions with different dependency structures between the variables of interest, such as the Clayton and Gumbel copula treated in detail in [33]. As explained in Sect. 13.2.6 copulas have mainly been applied to relate risks in stock portfolios, and a literature search for copulas applied to track dependency between price and production for commodities has been without success.

To explain the obtained empirical copula, the joint cumulative distribution and its level curves, depicted in Fig. 13.6, can be investigated. The cumulative probabilities of possible prices $F(u|v=V)$, obtained from the copula $C(u,v)$, given a production corresponding to the cumulative probability $v=V$ are represented in (13.13):

$$F(u|v=V)=\frac{F(u\cap v=V)}{F(v=V)}=\frac{C(u,v=V)}{F(v=V)}=\frac{C(u,V)}{V} \tag{13.13}$$

The numerator in the equation represents the cumulative probability for the (u,V) sample space in Fig. 13.6 where u is variable and V is fixed. For example, with $V=0.1$ corresponding to a production of approximately 290 GWh/quarter (Fig. 13.7b), a plot of the conditional cumulative price probability distribution, $F(u|V)$, can be generated by using (13.13). The resulting conditional cumulative probability distribution is graphed in Fig. 13.8. Note that $u=F(u)$ as u is a cumulative probability.

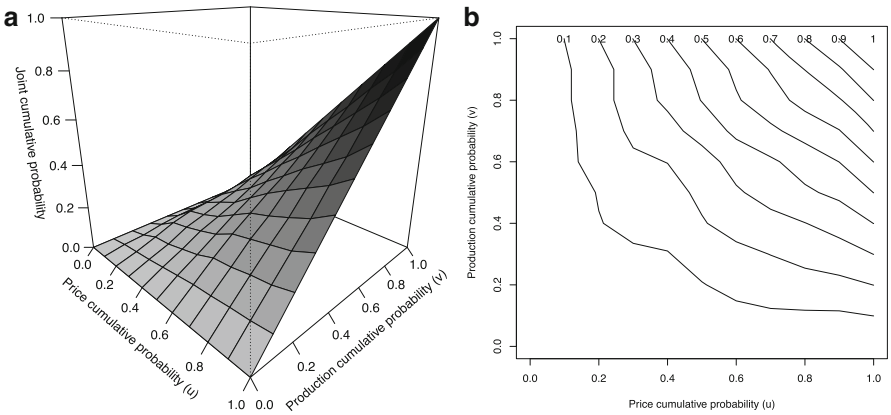

Fig. 13.6 Empirical copula based on average quarterly price and production data from 1986 to 2011. Note that the x- and y-axis represent the cumulative probability values of the input price and production distributions. The level curves could have been smoother if a larger data sample were used to generate the copula. Alternatively, a possibility could be to smooth the data points in the empirical copula. (**a**) Joint cumulative probability of price and production, $C(u, v)$ (**b**) Level curves of the copula

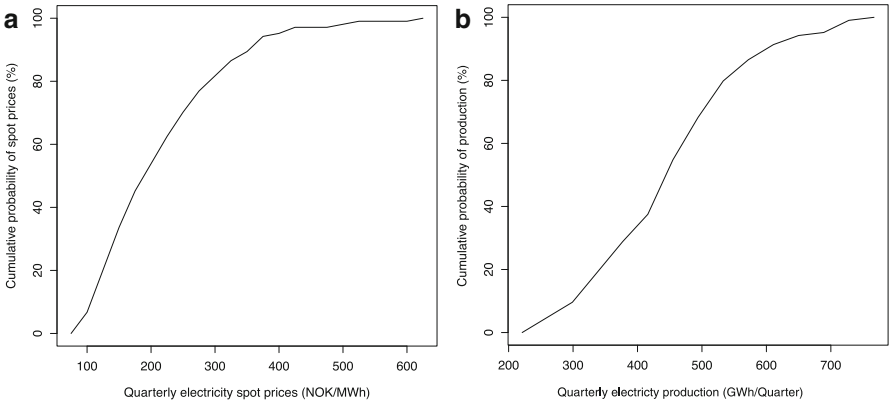

Fig. 13.7 Relationship between estimated quarterly electricity spot prices from (13.12), actual quarterly production volumes for a hydropower producer, and their respective cumulative distributions for the 1986 to 2011 period. From the horizontal flat part of the cumulative price curve it appears that some extreme price spikes have occurred during the sample period (**a**) Cumulative distribution of the electricity price (**b**) Cumulative distribution of the production volume

From Fig. 13.8 it appears that conditional on a low production, $V = 0.1$, the expected prices are generally higher than if prices and production volumes were independent. A similar analysis with production conditional on price can be performed by switching u and v.

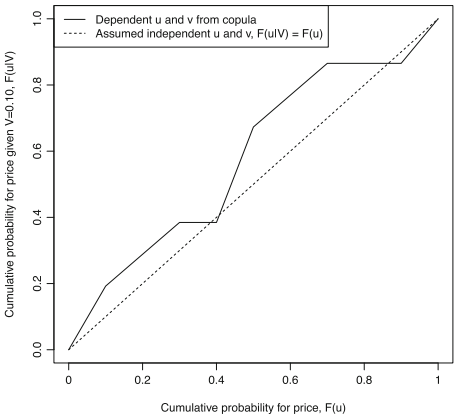

Fig. 13.8 Illustration of the cumulative price distribution conditional on a fixed production corresponding to a cumulative probability $V = 0.1$. The conditional price distribution obtained from the copula is compared to an assumed situation with independent price and production. The flat parts of the curve in the $0.3 \leq u \leq 0.4$ and $0.7 \leq u \leq 0.9$ areas are probably due to lack of data. Note that $u = F(u)$ as u is a cumulative probability. The conditional probability curve lies above the unconditional probability curve; thus based on the copula approach one should expect higher than usual spot prices when the production is low

13.4.3 Scenario Generation of Prices and Production

The next step in the model, step 3 depicted in Fig. 13.2, is to generate dependent cumulative probabilities of price and production volume.

The empirical copula function developed in Sect. 13.4.2 is used to generate numerous scenarios of price and production. These scenarios are simulated by first drawing one random uniformly distributed number between zero and one, representing the cumulated probability for the production, (V). In order to relate the cumulated production probability with a correlated cumulative price probability a new random uniformly distributed number between zero and one, W, is drawn and multiplied with the cumulative price probability, V. This product, VW, represents the conditional copula value $C(u, v = V)$, where V is known and u is yet to be determined. As $VW = C(u, v = V)$, (13.14) can be used to find the unknown u:

$$VW = C(u, v = V) \rightarrow W = \frac{C(u, v = V)}{V} = F(u|v = V) \tag{13.14}$$

From the equation it appears that W is the conditional cumulative probability of u given $v = V$, $F(u|v = V)$, defined in (13.13). The relationship between u and $F(u|v = V)$ was elaborated in Sect. 13.4.2 and exemplified with $V = 0.1$ in Fig. 13.8. To obtain u it is sufficient to find the abscissa of the function $F(u|v = V)$ with ordinate W. The determination of u is illustrated in Fig. 13.9, where random values of V and W are drawn equal to 0.1 and 0.6, respectively. The cumulative price probability, u, is then found to equal 0.47.

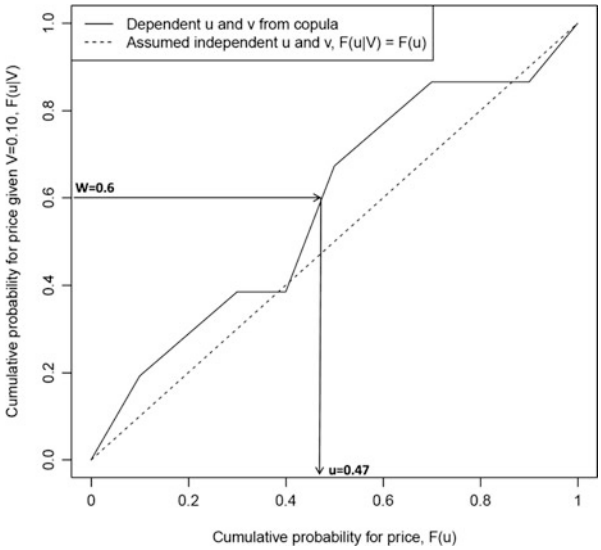

Fig. 13.9 Illustration of how to obtain the cumulative price probability u when a random cumulative production probability of $V = 0.1$ is drawn. A random W-value of 0.6 is also generated to link the production to a correlated random price. A resulting pair of (u,V) with values (0.47,0.1) is obtained from this simulation

The process of sampling correlated pairs of (u,v)-values from the empirical copula distribution can be repeated a large number of times and hence forms a copula-based Monte Carlo simulation.

13.4.4 Connecting the Cumulative Probability Pairs, (u,v), to Production Values and Spot/Swap Prices

The last step in the simulation process, step 4 in Fig. 13.2, is to link the cumulative probabilities from step 3 to production and price numbers.

First, production is considered. A data set of cumulative probabilities for production, v, has previously been generated. These probabilities are linked to the same distribution of quarterly production data used as input to the empirical copula. The relationship between production and its cumulative probabilities is illustrated in Fig. 13.7b. To obtain the production value corresponding to the cumulative probability v one must find the abscissa of the curve in the figure with ordinate v. The production value is found by interpolation of the cumulative distribution.

Second, prices are generated. The process is more cumbersome than for the production, as swap prices must be linked to the electricity spot prices. This is necessary since swap prices are required in later risk analysis, where swaps with different maturities are included in the hedging strategy. The model used for generating

input spot prices cannot be used to simulate historic swap prices, due to some missing input variables for forward price estimation. To generate a data set with related pairs of spot and swap prices, the method of [24] has been selected. Their two-factor model is defined in (13.15) and (13.16). This model will be treated in depth before an explanation of how to link cumulative price probabilities to spot and swap prices is given.

$$\begin{aligned}
&\ln(S_t) = f(t) + \chi_t + \xi_t \\
&f(t) = \gamma_0 + \gamma_1 \sin(\frac{2\pi t}{365} + \phi_k) \\
&d\chi_t = -\kappa\chi_t dt + \sigma_\chi dZ_\chi \\
&d\xi_t = \mu_\xi dt + \sigma_\xi dZ_\xi \\
&dZ_\chi dZ_\xi = \rho dt
\end{aligned} \tag{13.15}$$

In this model spot-forward prices can be approximated with two Brownian motions, a mean-reverting short-term factor, χ_t, and a long-term trend factor, ξ_t. These factors are driven by correlated normal error terms, dZ_χ and dZ_ξ, with a correlation coefficient ρ. The spot and swap prices are internally consistent, stochastic, and time dependent. Seasonality in prices is accounted for by adding a sine function with period one year, $f(t)$. The spot price S_t and the forward price $F_{T,t}$ with time to maturity T, at time t, are defined to follow (13.16):

$$\ln(S_t) = f(t) + \chi_t + \xi_t$$

$$\begin{aligned}
\ln(F_{T,t}) = &E_t(S_{T+t}) = f(T+t) + e^{-\kappa T}\chi_t + \xi_t + \mu_\xi T \\
&+ (1 - 2e^{-2\kappa T})\frac{\sigma_\chi^2}{4\kappa} + \frac{1}{2}\sigma_\xi^2 T \\
&+ (1 - 2e^{-\kappa T})\frac{\rho\sigma_\chi\sigma_\xi}{\kappa}
\end{aligned} \tag{13.16}$$

The forward price model is estimated using historical daily input data for spot, weekly, monthly, quarterly, and yearly contracts from 02.01.2006 to 30.04.2010. This period is chosen as some of the forward contracts had a different structure prior to 2006 and the available data stopped in 2010. 26,064 observations are considered, consisting of 23 different swap contracts and the system spot price. The number of days to delivery for the contracts is also used as input to the Kalman filter estimation. Descriptive data for historical spot and some selected forward contracts used as input to the Kalman filter are presented in Table 13.7.

Coefficients and the two factors, χ_t and ξ_t, are estimated by running a Kalman filter on (13.15). For an introduction to Kalman filtering see [13]. The results are summarized in Table 13.8 and Fig. 13.10.

These coefficients and factors can be used to generate spot and swap prices, and such simulation yields spot and contract prices with descriptive statistics summarized in Table 13.9. The output data are observed to be non-normal.

Table 13.7 Descriptive statistics for historical observed spot and selected forward contracts used as input to the Kalman filter with 1086 observations of each contract from 02.01.2006 to 30.04.2010

	Spot	1WF	1MF	1QF	1YF
Min	80.94	114.65	155.93	185.36	249.17
Max	1090.02	723.15	675.00	667.98	558.28
Avg	343.81	339.45	348.78	362.08	371.37
Med	334.15	332.24	335.72	338.05	357.01
St.dev	111.93	108.97	105.35	102.89	59.86
Skew	0.62	0.41	0.40	0.62	0.76
Ex. kurt	1.82	0.055	-0.22	-0.16	0.02
JB	220.75	31.23	31.83	70.48	105.06

Table 13.8 Estimated values of the coefficients in the two-factor model of [24]

Coefficient	Value
μ_ξ	-0.043
σ_ξ	0.810
κ	1.793
σ_χ	0.264
ρ	-0.268
γ_0	1.000
γ_1	0.100
ϕ_k	-0.743

The values are obtained by running a Kalman filter on (13.15). The long-term drift factor μ_ξ is slightly negative and the mean-reversion coefficient κ is relatively high which gives a half-life, $\ln(2)/\kappa$, of fluctuations of less than a half week. The correlation coefficient ρ is closer to 0 than to -1, so the two processes are quite independent. The constant γ_0 equals 1 and could have been omitted with a resulting upward shift of 1 unit in the long-term drift factor μ_ξ

Comparing the statistics of the input data to the Kalman filter (Table 13.7), with the output in Table 13.9, it appears that the range of the output data is narrower than in the input data and the standard deviation slightly lower. The difference is most prominent for short-term contracts. These observations should not come as a surprise since a well-known shortcoming of two-factor models, such as the one derived by [24], is the volatility structure they assume. Although such models fit observed prices quite well, the volatility term structure is not captured accurately. Cortazar , Naranjo [11] show how such models tend to underestimate the volatility structure of oil and copper forwards. As the erroneous volatility estimation is particularly strong not only for short-term contracts, but also for long-term contracts, the estimated volatility is consistently below that of their observed data. As electricity shares many of the same properties as other commodities it is likely that the same problem arises for electricity swaps, just as observed in Tables 13.7 and 13.9. The tendency to underestimate volatility in swap contracts is an observation one needs to bear in mind during the later risk analysis.

Having described the pricing relationship between spot and swap prices thoroughly, it is possible to create one single distribution including these two variables.

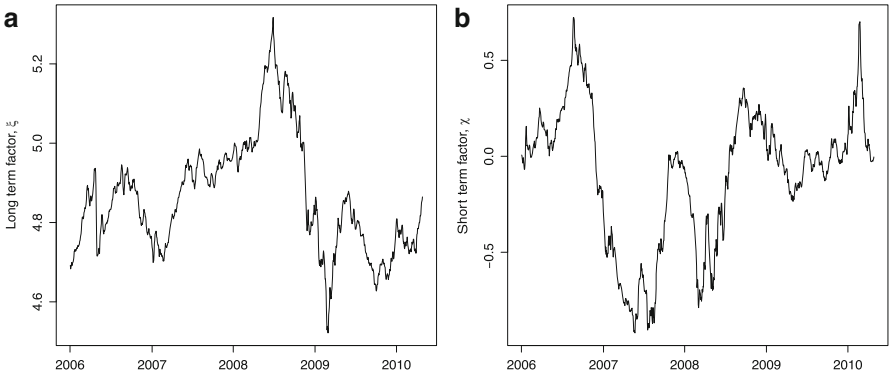

Fig. 13.10 Estimated time series for the long-term factor ξ_t and short-term factor χ_t from the [24] two-factor model. The factors have a relatively low correlation coefficient of -0.27 and two factors provide better fit than a one-factor model. No clear trend can be seen for the long-term factor ξ_t which should not come as a surprise since electricity is not a commodity from which an investor expect any return. The short-term factor fluctuates around a zero mean which could be expected for a mean-reverting process. (**a**) Long term factor ξ_t (**b**)Short term factor χ_t

Table 13.9 Descriptive statistics for estimated spot and selected forward contracts calculated from the forward equation, (13.16), with the coefficients,Table 13.8, and factors, Fig. 13.10, from the Kalman filter estimation as input

	Spot	1WF	1MF	1QF	1YF
Min	146.41	150.02	158.86	196.26	266.76
Max	664.56	662.77	630.06	630.11	540.86
Avg	314.51	318.04	327.87	344.79	377.52
Med	309.35	311.03	308.71	321.03	367.41
St.dev	103.86	103.79	105.22	100.12	59.72
Skew	0.50	0.48	0.50	0.90	0.69
Ex. kurt	−0.02	−0.10	−0.33	0.37	−0.06
JB	7.25	6.89	8.08	24.69	13.75

This distribution can then be linked to the cumulated probabilities u. First, a table with possible spot and swap prices with different maturities are generated, as represented in Table 13.10.

In Tab. 13.10, the first column represents the date with daily frequency. The last date in the table, $t = 0$, can be considered as today, whereas the negative times above represent the number of days prior to today. The spot price S_t and swap prices $F_{T,t}$, where T describes the different swaps, are then generated for each date t with (13.16).

A time analysis of the realized prices obtained by a producer in the derivative market is then conducted, as depicted in Table 13.11. The motivation is to relate

Table 13.10 Swap contracts available from (13.15)

Time (t)	Spot	Week 1	Week 2	Week 3	Week 4	Month 1	Month 2	Month 3	Quarter 1	Quarter 2	Quarter 3	Quarter 4	Year 1
-20	S_{-20}	$F_{1W,-20}$	$F_{2W,-20}$	$F_{3W,-20}$	$F_{4W,-20}$	$F_{1M,-20}$	$F_{2M,-20}$	$F_{3M,-20}$	$F_{1Q,-20}$	$F_{2Q,-20}$	$F_{3Q,-20}$	$F_{4Q,-20}$	$F_{1Y,-20}$
-19	S_{-19}	$F_{1W,-19}$	$F_{2W,-19}$	$F_{3W,-19}$	$F_{4W,-19}$	$F_{1M,-19}$	$F_{2M,-19}$	$F_{3M,-19}$	$F_{1Q,-19}$	$F_{2Q,-19}$	$F_{3Q,-19}$	$F_{4Q,-19}$	$F_{1Y,-19}$
							...						
-3	S_{-3}	$F_{1W,-3}$	$F_{2W,-3}$	$F_{3W,-3}$	$F_{4W,-3}$	$F_{1M,-3}$	$F_{2M,-3}$	$F_{3M,-3}$	$F_{1Q,-3}$	$F_{2Q,-3}$	$F_{3Q,-3}$	$F_{4Q,-3}$	$F_{1Y,-3}$
-2	S_{-2}	$F_{1W,-2}$	$F_{2W,-2}$	$F_{3W,-2}$	$F_{4W,-2}$	$F_{1M,-2}$	$F_{2M,-2}$	$F_{3M,-2}$	$F_{1Q,-2}$	$F_{2Q,-2}$	$F_{3Q,-2}$	$F_{4Q,-2}$	$F_{1Y,-2}$
-1	S_{-1}	$F_{1W,-1}$	$F_{2W,-1}$	$F_{3W,-1}$	$F_{4W,-1}$	$F_{1M,-1}$	$F_{2M,-1}$	$F_{3M,-1}$	$F_{1Q,-1}$	$F_{2Q,-1}$	$F_{3Q,-1}$	$F_{4Q,-1}$	$F_{1Y,-1}$
0	S_0	$F_{1W,0}$	$F_{2W,0}$	$F_{3W,0}$	$F_{4W,0}$	$F_{1M,0}$	$F_{2M,0}$	$F_{3M,0}$	$F_{1Q,0}$	$F_{2Q,0}$	$F_{3Q,0}$	$F_{4Q,0}$	$F_{1Y,0}$

The first column of the table gives the time at which the spot price and swap contracts are traded. The upper two rows illustrate the term structure of the swaps. The table can be used to understand how contracts traded on different days are denoted and hence be used as a reference for Table 13.11 where the maturity date of these contracts is shown

Table 13.11 Rearranged swap contracts available from (13.15) illustrate how the realized price and swap prices are linked

Time day	Spot	Week 1	Week 2	Week 3	Week 4	Month 1	Month 2	Month 3	Quarter 1	Quarter 2	Quarter 3	Quarter 4	Year 1
−20	S_{-20}	$F_{1W,-27}$	$F_{2W,-34}$	$F_{3W,-41}$	$F_{4W,-48}$	$F_{1M,-48}$	$F_{2M,-81}$	$F_{3M,-112}$	$F_{1Q,-112}$	$F_{2Q,-202}$	$F_{3Q,-293}$	$F_{4Q,-385}$	$F_{1Y,-385}$
−19	S_{-19}	$F_{1W,-26}$	$F_{2W,-33}$	$F_{3W,-40}$	$F_{4W,-47}$	$F_{1M,-47}$	$F_{2M,-80}$	$F_{3M,-111}$	$F_{1Q,-111}$	$F_{2Q,-201}$	$F_{3Q,-292}$	$F_{4Q,-384}$	$F_{1Y,-384}$
							...						
−3	S_{-3}	$F_{1W,-10}$	$F_{2W,-17}$	$F_{3W,-24}$	$F_{4W,-31}$	$F_{1M,-31}$	$F_{2M,-64}$	$F_{3M,-95}$	$F_{1Q,-95}$	$F_{2Q,-185}$	$F_{3Q,-276}$	$F_{4Q,-368}$	$F_{1Y,-368}$
−2	S_{-2}	$F_{1W,-9}$	$F_{2W,-16}$	$F_{3W,-23}$	$F_{4W,-30}$	$F_{1M,-30}$	$F_{2M,-63}$	$F_{3M,-94}$	$F_{1Q,-94}$	$F_{2Q,-184}$	$F_{3Q,-275}$	$F_{4Q,-367}$	$F_{1Y,-367}$
−1	S_{-1}	$F_{1W,-8}$	$F_{2W,-15}$	$F_{3W,-22}$	$F_{4W,-29}$	$F_{1M,-29}$	$F_{2M,-62}$	$F_{3M,-93}$	$F_{1Q,-93}$	$F_{2Q,-183}$	$F_{3Q,-274}$	$F_{4Q,-366}$	$F_{1Y,-366}$
0	S_0	$F_{1W,-7}$	$F_{2W,-14}$	$F_{3W,-21}$	$F_{4W,-28}$	$F_{1M,-28}$	$F_{2M,-61}$	$F_{3M,-92}$	$F_{1Q,-92}$	$F_{2Q,-182}$	$F_{3Q,-273}$	$F_{4Q,-365}$	$F_{1Y,-365}$

The contracts are sorted so that their maturity date corresponds to the date in the first row

the prices of swap contracts and thereby the realized price for the electricity sold, with spot prices. This new way to illustrate spot and swap prices might be useful to investigate the effect of swaps in hedging decisions. The producer achieves a realized price F_t for the electricity it sells in the derivative market at time t given by (13.17), where $F_{T,t}$ correspond to the different swaps traded at time t. W_{F_T} is the weight of a producer's total derivative investment positioned in each contract. $F_{T,t=(t-T)}$ represents the swap price T days ahead of time t. Note that $\sum_T W_{F_T} = 1$.

$$F_t = \sum_T W_{F_T} F_{T,t=(t-T)}, T \in \{1W, 2W, 3W, \ldots, 1Y, 2Y, 3Y\} \tag{13.17}$$

To exemplify how to interpret Table 13.11, a two-week swap is considered at time $t = 0$, the last row in the table. The $F_{2W,-14}$ entry illustrates that the price of a two-week swap at time $t = 0 - 14 = -14$ thus two weeks before $t = 0$ can be considered as the realized price of the electricity if production is hedged using this contract at time $t = -14$. This hedged price can therefore be compared with the spot price at $t = 0$. A similar approach can be made for all other dates t and for all other maturities T. Thus, the volatility of the realized cash flow over time can be examined by using (13.4) with S_t, F_t, and H as input variables.

To create an empirical distribution for spot/swap prices, the rows in Table 13.11 are sorted with increasing spot price, but retaining the same swap prices to the spot prices as in the table. The rows in the table are thus shuffled.

Having created an empirical distribution for spot/swap prices it is now possible to link the cumulative price probabilities, u from step 3 in Fig. 13.2, to simulated spot and swap prices. This is simply done by finding the two successive rows in the sorted table corresponding to the nearest lower and higher u and interpolating between these two rows for each spot and swap contract, as illustrated in Table 13.12. Hence, daily spot and swap prices for all u can be obtained. The price output of the copula will therefore be based on daily and not quarterly data, even though the production has quarterly resolution. The minimum, average, and maximum values of the output price from the Kalman filter, Table 13.9, are to some extent higher than the quarterly data used as input to the empirical copula, Table 13.6. Nonetheless, the standard deviations of the two data sets are almost identical, and since the risk measures in

this chapter will be based on relative measures, the choice of working with two different pricing models will not disturb the risk analysis a lot.

Table 13.12 The sorted table of spot and swap prices with their cumulative probability u

Cumul.prob.	Spot	Week	Week	Week	Week	Month	Month	Month	Quarter	Quarter	Quarter	Quarter	Year
(u)		1	2	3	4	1	2	3	1	2	3	4	1
0	…	…	…	…	…	…	…	…	…	…	…	…	…
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.46	S_{t_X}	F_{1W,t_X}	F_{2W,t_X}	F_{3W,t_X}	F_{4W,t_X}	F_{1M,t_X}	F_{2M,t_X}	F_{3M,t_X}	F_{1Q,t_X}	F_{2Q,t_X}	F_{3Q,t_X}	F_{4Q,t_X}	F_{1Y,t_X}
0.47	Interp.	Interp.	Interp.	Interp.	Interp.	Interp.	Interp.	Interp.	Interp.	Interp.	Interp.	Interp.	Interp.
0.49	S_{t_Y}	F_{1W,t_Y}	F_{2W,t_Y}	F_{3W,t_Y}	F_{4W,t_Y}	F_{1M,t_Y}	F_{2M,t_Y}	F_{3M,t_Y}	F_{1Q,t_Y}	F_{2Q,t_Y}	F_{3Q,t_Y}	F_{4Q,t_Y}	F_{1Y,t_Y}
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1	…	…	…	…	…	…	…	…	…	…	…	…	…

Interpolation is used to connect the cumulative probability u obtained in Fig. 13.9 to the empirical price distribution from the Kalman filter

As pairs of related spot, swap, and production values now are available, sums and products of these variables can easily be calculated. From the variance of these sums and products it is possible to obtain covariance between price and production that were previously unavailable. This copula-based Monte Carlo simulation can hence be applied to evaluate the price and production uncertainty on cash flow with measures such as CFaR, CCFaR, and hedge effectiveness. The large number of different swap contracts available from the two-factor model also renders possible an analysis of how the term structures of such contracts influence the hedging performance and the hedge ratios of a hydropower producer.

13.5 Results and Discussion

With the copula-based Monte Carlo model developed in Sect. 13.4, 10,000 scenarios of dependent electricity spot, swap, and production values are generated. These values and different hedge ratios are then used as input to the expression of the hydropower producer's cash flow in (13.4). Thus, for each hedge ratio the resulting 10,000 cash flow scenarios can be used to examine the cash flow uncertainty expressed by different risk measures.

13.5.1 Risk Premium

Before the risk of the cash flow is assessed, an analysis of the risk premium of several swaps traded in the 2006 to 2010 period at Nord Pool is conducted. The risk premium is studied to judge the attractiveness of these derivatives. As mentioned

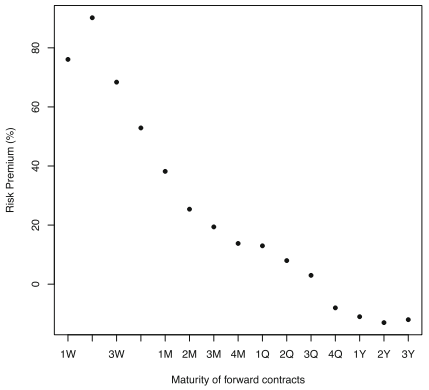

Fig. 13.11 Annualized risk premiums for different forward contracts. Calculated from (13.18). There is a clear downward trend in the annualized risk premium with respect to the time to maturity. Short-term swaps are therefore economically more attractive to hydropower producers than long-term contracts

in Sect. 13.2.5 the risk premium of the traded swaps may be connected to the term structure of the contracts. Hence, the risk premium is examined to enable an analysis of the trade-off between risk and return. The risk premium is defined according to (13.18):

$$\begin{aligned} R(t,T) &= \frac{F_{T,t} - E_t[S_T]}{F_{T,t}} = 1 - \frac{\sum_{t=T-P}^{T} S_t}{PF_{T,t}}, \\ \text{Annualized } R(t,T) &= (1 + R(t,T))^{\frac{365}{T}} - 1 \end{aligned} \tag{13.18}$$

where t is a date, T is the time to expiration of a contract, and P is the delivery length of the contract. Thus, $(\sum_{t=T-P}^{T} S_t)/P$ is the average spot price during the delivery period and $F_{T,t}$ is the forward price of a swap contract with time to maturity, T, at time t.

A summary of the annualized risk premiums is depicted in Fig. 13.11. The figure reveals the tendency of a decreasing risk premium when the time to maturity of these contracts increases. This is consistent with the findings of [8]. Also, the slightly negative drift term, μ_ξ, in (13.16) for the-long term evolution in forward prices bodes well with the decreasing risk premium since the price of the contract then decreases with the time to maturity. The decreasing risk premium with longer time to maturity is also consistent with the hedging pressure in the market, explained in Sect. 13.2.5. Consumers tend to hedge themselves in the short term whereas producers often prefer long-term contracts in their hedging strategies. This creates an unbalanced demand-supply situation for swap contracts which affects the pricing of the contracts in the direction of higher risk premiums for short-term contracts and low or even negative risk premiums for forwards with long time to maturity. The risk premium present in swap agreements argues for the use of short-term contracts by

producers to obtain an advantageous realized price for the electricity secured in the derivative market. However, the risk premium has little to do with the elimination of risk as yearly variations and extreme prices will still affect the cash flow greatly. Thus, both the risk premiums and the contracts' abilities to reduce risk should be considered in the hedging decisions.

13.5.2 Minimum Variance Analysis

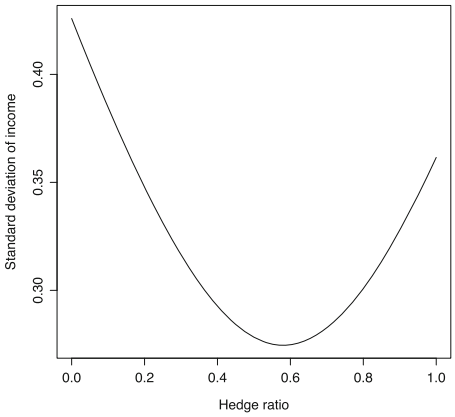

Fig. 13.12 Standard deviation of cash flow measured as a fraction of expected cash flow, varying the hedge ratio. The minimum standard deviation is obtained at a hedging level of 57.0% of expected production. In the plot the term structure of the hedged swaps is neglected

A minimum variance analysis can be carried out to measure and reduce risk. With dependent price and production data series from the copula-based Monte Carlo simulation, variance in cash flow can be minimized by choosing a hedge ratio according to (13.6) in Sect. 13.2.3. The hedge ratio represents the percentage of the expected production that should be sold in the forward market. Still, this analysis does not take into account which swaps to include in a power portfolio since (13.6) ignores the term structure of these derivatives, thus neglecting that weekly and yearly contracts affect risk reduction differently. However, this approach gives a benchmark for the optimal hedge ratio.

Figure 13.12 depicts the standard deviation of the electricity producer's income as a function of the hedge ratio. Minimum variance is obtained for a hedge ratio of about 57.0%. This hedge ratio is consistent and almost equal to the tax-neutral hedge of 58.3% elaborated in Sect. 13.2.3. Hence, the copula framework used to generate price and production pairs has only marginal effect on the variance of the cash flow and barely changes the optimal hedging level. The figure still underlines the significant variance reduction effect of hedging. For a non-hedged producer the volatility

of the cash flow is about 42% and drops to approximately 28% when the optimal hedging level is chosen. A question yet to be answered is how the time horizon of different derivative contracts affects risk reduction and how a power portfolio should be composed. Possibly, the optimal hedge ratio can be affected by this choice.

13.5.3 Restricted Minimum Variance Analysis

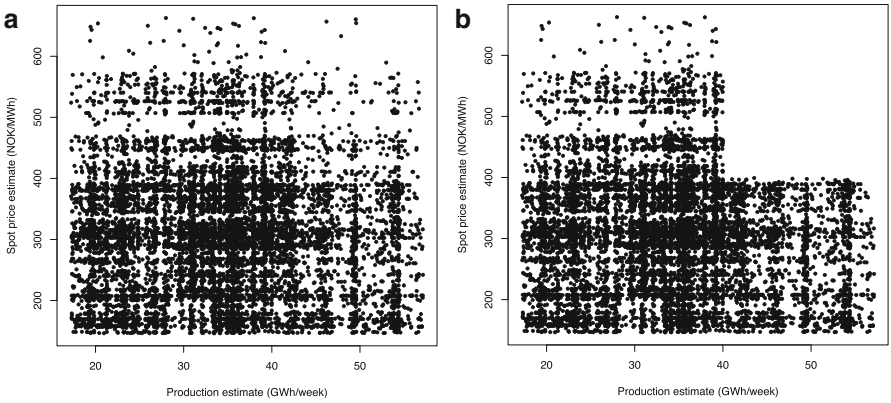

Fig. 13.13 Unrestricted and restricted scenarios for price and production for a hydropower producer. The density of the points underlines the probability of the outcomes. Some small discontinuities are the result of lack of input data to the empirical copula. In the restricted copula analysis, some scenarios that greatly exceed historic outcomes have been eliminated. For the unrestricted situation the optimal hedge ratio is 57.0% and it drops to 51.0% for the restricted scenario. The unrestricted scenarios will be used in later risk analyses (**a**) Unrestricted scenarios (**b**) Restricted scenario

The generated series for price and production applied to evaluate volatility in the previous section permit scenarios where both very high price and production are connected. Persistent high prices are only viable in the hydro-dominated Nord Pool area during cold and dry periods, which drain the reservoirs to a critical level. During these periods very high production is not desirable, and the very high price and production scenario is therefore unlikely. For this reason it is interesting to investigate the consequence of excluding the assumed improbable scenarios from the data set. An illustration of the effect of the data set when the high price, high production scenarios are deleted is shown in Fig. 13.13. The consequence on the optimal hedge ratio is a minimum variance obtained for a hedging level of 51.0% of the expected production. The reduction from the unrestricted simulation emphasis that a producer should be careful to hedge as much as the tax-neutral hedge of 58.3% and should probably have a ceiling of the hedge ratio closer to 50% when risk reduction is measured by the variance framework. A lower optimal hedging level is the result

of a more prominent natural hedge for the restricted copula case, reflecting a more negative correlation between price and production, than observed in the unrestricted minimum variance analysis. In the next analyses, the unrestricted copula scenarios are considered.

13.5.4 Cash Flow at Risk Analysis

Cash flow at risk is used as a tool to measure downside risk which is relevant for a hydropower producer that operates in a sector where prices are subjected to extreme fluctuations. CFaR and CCFaR are treated more closely in Sect. 13.2.1. The chosen threshold value of these risk measures is set equal to $\alpha = 5\%$. This risk level reflects the secure environment in which hydropower producers operate with stable earnings and low probability of facing financial distress. These criteria should be determining when a company chooses risk measures, according to [32], and $CFaR_{5\%}$ and $CCFaR_{5\%}$ seem suitable. An even higher risk threshold can also be argued for; Fleten et al. [18] use as an example a $VaR_{10\%}$ to monitor risk for a hydropower producer.

The cash flow at risk analysis conducted in this chapter considers the time horizon of the hedging, which was a shortcoming of the minimum variance approach in the previous sections. For contracts with long time to maturity, the spot price has time to deviate a lot from the expected level if price estimates were wrong. Long-term contracts are therefore less correlated with the spot price in their maturity period than contracts with shorter time to maturity. For these short-term contracts, estimates are rarely far out of range. Stated differently, since one knows less about what will happen far into the future than the possible outcome of the next days or weeks, long-term contracts are less correlated with the spot price in their delivery period than short-term contracts. This feature can be the reason behind some of the characteristics of the calculated $CFaR_{5\%}$ and $CCFaR_{5\%}$ in Fig. 13.14 that are discussed below. The $CFaR_{5\%}$ and $CCFaR_{5\%}$ as percentage of expected cash flow in the figures are defined in (13.19). High $CFaR_{5\%}$ and $CCFaR_{5\%}$ values are favorable, since the threshold values then are closer to the expected cash flow.

$$\text{(C)CFaR as a \% of expected cash flow} = \frac{(C)CFaR}{E[CF]}100\% \tag{13.19}$$

First, as the time to maturity of the contracts contained in the hedged portfolio increases, the downside risk measured by $CFaR_{5\%}$ and $CCFaR_{5\%}$ is reduced when the optimal hedge ratio is chosen. For a portfolio with one-week contracts, Fig. 13.14a, the $CFaR_{5\%}$ is only 40% of the mean for all hedge ratios. With yearly contracts, Fig. 13.14d, the same number is about 70% for an optimal hedge ratio. Second, when contracts with longer time to maturity are used, the optimal hedge ratio drops. For short-term contracts there is no clear optimal hedge ratio; Fig. 13.14a reveals an

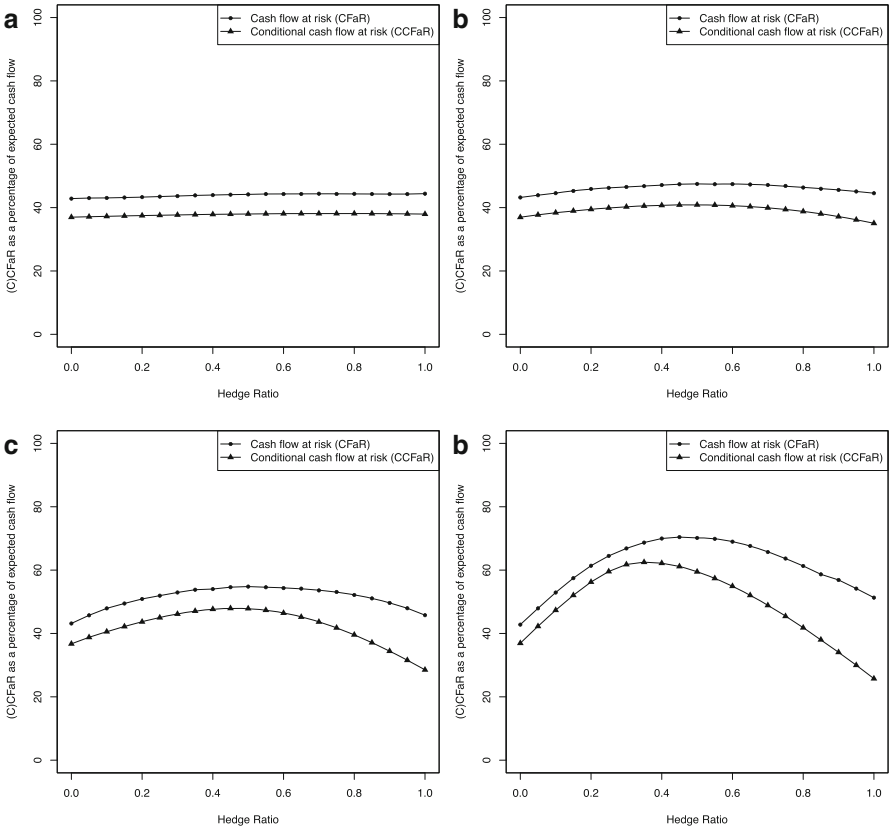

Fig. 13.14 Downside risk in cash flow for a producer hedging only 1-week futures, 1-month forward, 1-quarter forward, and 1-year forward contracts as a function of the hedge ratio. CFaR and CCFaR are represented as a percentage of expected cash flow. For long-term swaps the CFaR and CCFaR curves are more parabolic and can eliminate more downside risk than short-term contracts, as observed by the higher obtained values. The hedge ratio that reduces most downside risk is the abscissa of the maximum of the curves, and the optimal hedging level drops with the length of the hedged contracts. (**a**) Power portfolio containing spot and 1-week contracts (**b**) Power portfolio containing spot and 1-month contracts (**c**) Power portfolio containing spot and 1-quarter contracts (**d**) Power portfolio containing spot and 1-year contracts

almost flat behavior. A relatively high hedge ratio would therefore not imply less risk than a lower one. The optimal hedge ratio then drops successively for monthly and quarterly contracts, Figs. 13.14b and 13.14c, and attains a minimum level of approximately 35% when $CCFaR_{5\%}$ is assessed for one-year contracts in Fig. 13.14d. The hedge ratio that minimizes downside risk is always lower for $CCFaR_{5\%}$ than for $CFaR_{5\%}$, as $CCFaR_{5\%}$ punishes extreme events more severely than $CFaR_{5\%}$. Thirdly, when the time to maturity of the contracts increases, it is more important to choose the correct hedge ratio. Short-time horizons yield relatively flat $CFaR_{5\%}$ and $CCFaR_{5\%}$ curves whereas longer-time horizons yield a more parabolic-shaped

$CFaR_{5\%}$ and $CCFaR_{5\%}$ curve. Thus, an overhedged producer using long-term swaps may experience higher risk than an unhedged producer if its hedge ratio greatly exceeds the optimal level.

13.5.5 Hedge Effectiveness

Hedging effectiveness, defined in (13.1), has also been assessed to evaluate how swap contracts with different maturities affect the variance reduction in cash flow. Hedge effectiveness is treated more thoroughly in Sect. 13.2.1. The hedge effectiveness analysis conducted herein includes the time perspective of the hedge as opposed to the minimum variance analysis in Sect. 13.5.2. The results of the hedge effectiveness analysis are presented in Figs. 13.15 and 13.16.

The figure underlines that any contract with a time to maturity of less than two months is not likely to eliminate more than 10% of the variance in cash flow at any hedging level. Conversely, contracts with longer time to maturity may eliminate almost 50% of the producer's revenue variance. This result emphasizes that it is pointless to use short-term swaps if the aim is to reduce variance in cash flow. The finding can possibly explain the surprising result in an empirical analysis of hedging policies among Norwegian hydropower producers by [30]. In their study the majority of producers did not obtain a significant reduction in their cash flow volatility. However, they achieved a substantial part of their profit from their hedging program. It seems therefore likely that many hydropower producers focus on increased profitability rather than risk reduction. If the aim of the hedge is to reduce risk, the hedge effectiveness analysis underlines that most risk is eliminated for hedge ratios in the 40–60% area (Fig. 13.16a, b, and c). As explained in Sect. 13.5.4, overhedging can be very risky, and Fig. 13.16d stresses how the variance reduction collapses when the hedging level increases to 90% of the expected production. Overhedging may hence result in increased volatility and all risk protection can be lost. As hedging generally leads to reduced revenue, overhedging implies higher risk and lower return.

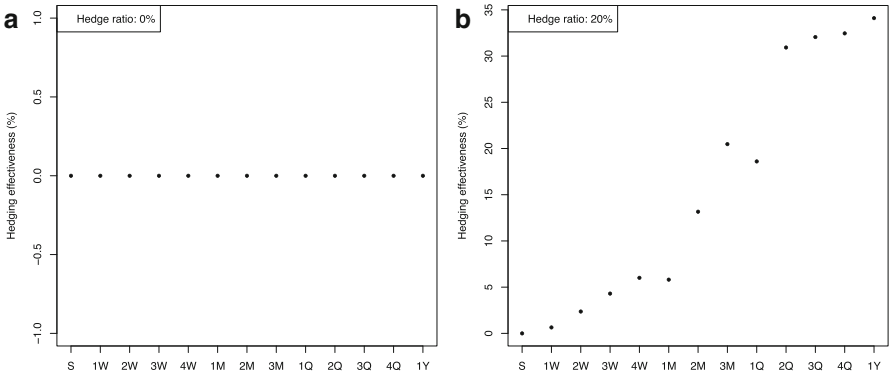

Fig. 13.15 Hedge effectiveness for swaps with different maturities for various hedge ratios. For the unhedged case the hedge effectiveness is zero. (**a**) 0 % Hedge (**b**) 20 % Hedge

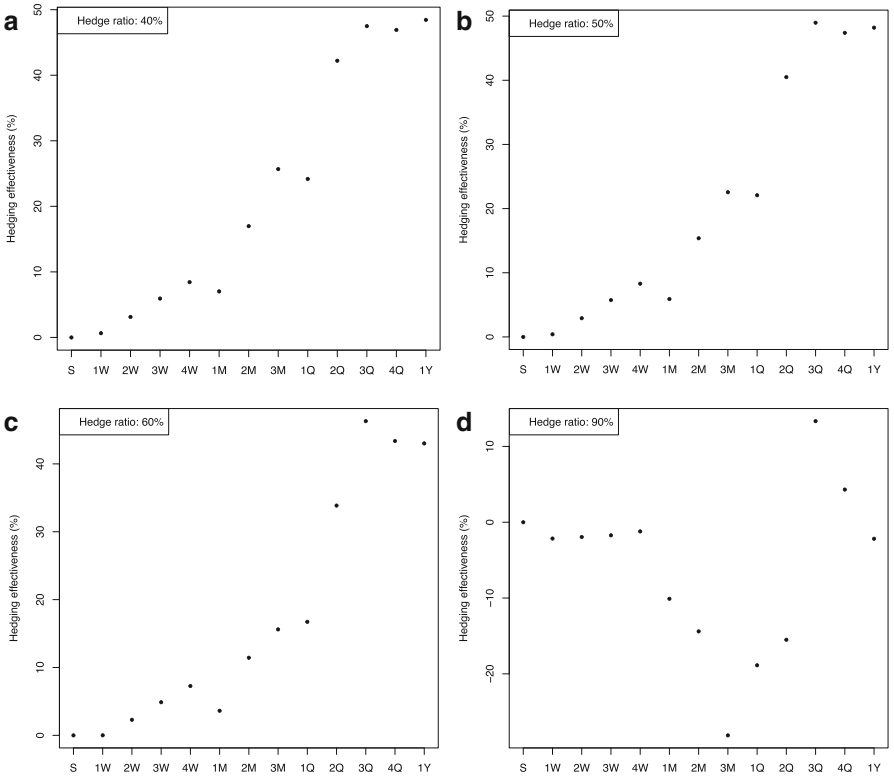

Fig. 13.16 Hedge effectiveness for swaps with different maturities for various hedge ratios. The discontinuity in the increasing hedge effectiveness trend observed for 1-month and 1-quarter forwards might be due to the different contract structure than for the preceding points on the abscissa. The increasing hedge effectiveness with time to maturity illustrates that long-term contracts eliminate more risk than short-term contracts at an adequate hedging level. The optimal hedging level is between 40 and 60%. When overhedged, as in Fig. 13.16d, the hedge effect collapses and leads to increased cash flow volatility. (**a**) 40 % Hedge (**b**) 50 % Hedge (**c**) 60 % Hedge (**d**) 90 % Hedge

13.5.6 Model Results Compared with Historical Hedge Ratios

In Sect. 13.3, Table 13.3, the optimal hedge ratios obtained from the historical data are 47.5%, 28.0 %, and 15.9% for minimum variance, $CFaR_{5\%}$, and $CCFaR_{5\%}$, respectively. In the analyses following the copula-based Monte Carlo simulation, Sects. 13.5.4 and 13.5.5, the optimal hedging levels are 40–60% for the hedge effectiveness approach and about 45% and 35% for $CFaR_{5\%}$ and $CCFaR_{5\%}$. The empirical variance is compared to hedge effectiveness since both measures minimize variance and include the time perspective. Thus, it appears that the empirical results are in line with the outcome of simulations conducted in this chapter. However, the empirical results tend to recommend slightly lower hedge ratios than the

copula-based values. As discussed previously this could be founded in the estimation of the spot-swap relationship with a two-factor model which normally underestimates the volatility of the swap contracts. The estimated swap contracts may therefore be more risky than supposed in the analyses. If a more complete and complex model for the spot-swap price relationship had been selected, the obtained optimal hedge ratios would probably have been lower.

It is also interesting to observe that none of the historical optimal hedging strategies involve investment in weekly contracts. This same observation is discussed in the $CFaR_{5\%}$, $CCFaR_{5\%}$, and hedge effectiveness analyses with a conclusion that weekly contracts are too correlated with the spot price to provide risk elimination, and at best yield a positive risk premium for the producer.

Finally, it seems like the empirical analysis obtains less risk elimination, measured by hedge effectiveness, $CFaR_{5\%}$, and $CCFaR_{5\%}$, than the copula framework claims possible. This problem questions the adequacy and robustness of the copula-based Monte Carlo simulation.

13.5.7 Implications

The implications of the previous analyses are that a producer should adjust its hedging strategy according to the purpose of the hedge. The minimum variance analysis provides an easy and comprehensive picture of the optimal hedging level, with a target hedge ratio in the 51–57.0% range. However, this analysis seems too simplistic as it ignores the term structure effects of the swap contracts. The analysis shows that it is possible to reduce the standard deviation in cash flow from about 42% in the unhedged case to approximately 28% when an optimal hedge ratio is chosen; see Fig. 13.12.

Extension of the variance approach by observing hedge effectiveness of different hedging strategies, consisting of investing a variable part of the expected production in one swap contract at a time, is shown in Figs. 13.15 and 13.16. The hedging effectiveness measure supports the minimum variance approach, but specifies that the maximum risk reduction is only possible with long-term contracts. Besides, it is shown that hedging by use of short-term contracts is almost pointless if the aim is to reduce risk. The $CFaR_{5\%}$ and $CCFaR_{5\%}$ analyses present similar results. Short-term contracts have only a marginal risk-reducing effect, shown by the flat curves in Fig. 13.14a, and the investment in these derivatives therefore provides negligible risk protection for a hydropower producer. On the other hand, the long-term contracts may reduce risk significantly, depicted by the parabolic $CFaR_{5\%}$ and $CCFaR_{5\%}$ curves in Fig. 13.14c and d. The hedge effectiveness, $CFaR_{5\%}$, and $CCFaR_{5\%}$ approaches show that hedging by means of swaps with longer time to maturity can almost halve the volatility and the downside risk in cash flow if appropriate hedge ratios are chosen. Note that a detailed analysis of the appropriate hedge ratios should be undertaken to prevent risky overhedging.

Nevertheless, the attractiveness of the long-term contracts lies only in their risk-reducing nature as they are priced with a marginally positive or even a negative risk premium as depicted in Fig. 13.11. Conversely, short-term contracts are generally priced with a positive risk premium and the premium decreases as the maturity, and hence the risk-eliminating ability of the swaps increases. Fleten et al. [18] also find that hedging costs are higher when producers use contracts with long time to maturity. Thus, the usual risk-reward relationship, faithful to the findings of [26], also applies to the hedging strategy of hydropower producers.

Swaps can therefore be used for two main purposes by a hydropower producer: as speculation in short-term contracts with the aim to obtain attractive prices, but without eliminating much risk, and alternatively as risk-reduction strategies investing in long-term, risk-reducing swaps, achieving a less attractive premium for this risk protection. These double possible uses of these derivatives can probably be the source of the troubling findings of [30], discussed briefly in Sect. 13.5.5. The tendency of hydropower companies to profit from their hedging transactions rather than reducing cash flow volatility can therefore be founded in hedging biased towards short-term instead of long-term contracts. Translated, this means that hydropower companies engage in value adding rather than risk-reducing hedging strategies.

13.6 Conclusion

For renewables producers, price and feed-in uncertainty are the two most important operational risk factors. An empirical copula is suggested to link the price and production volume in a new way. The copula offers an improved relationship between variables, including flexibility in tail dependency and normality assumptions, which a linear correlation coefficient does not allow. This chapter develops a copula-based Monte Carlo model to investigate hedge ratios for Norwegian hydropower producers taking into account price and production volume uncertainties. The variances in revenue, hedge effectiveness, CFaR, and CCFaR are used as risk measures to examine how swaps with different maturities affect a hydropower producer's hedging strategy and hedge ratios.

Swaps with short time to maturity are shown to have little effect on risk reduction measured by hedge effectiveness, CFaR, and CCFaR. Conversely, long-term contracts should be preferred in order to obtain the highest level of risk reduction measured by the proposed risk measures. Also, the optimal hedge ratio shifts towards lower levels when the time to maturity of the hedged swaps included in the power portfolio increases. This is due to the long-term contracts' lower correlation with the spot price which offers a better risk reduction than short-term swaps. Overhedging, meaning hedging too much of the expected production, in long-term derivatives may result in a risk increase in cash flow instead of risk reduction. The assessed risk measures give different results when it comes to the optimal hedge ratio. Thus, it may be problematic to recommend one specific risk measure and one single hedge ratio. The choice of risk measure must therefore be based on the

hydropower producer's approach to risk. Anyway, for all risk measures a hedge ratio of 35–60 % of expected production invested in long-term contracts is observed to give the highest risk reduction.

The hedging performance of swap contracts is seen in light of the expected risk premium for these derivatives. The risk premium is a decreasing function of the time to maturity of the swaps, and the low or even negative premium achieved for long-term contracts can be considered as a cost of the provided risk reduction. Hence, swap agreements can be used for two main purposes by a hydropower producer: as speculation in short-term contracts with the aim to obtain attractive prices but without removing much risk and alternatively as a risk-reduction strategy taking positions in long-term swaps with a negative or less attractive risk premium.

The copula-based model developed in this chapter has some shortcomings. First, the issue with two sets of prices is problematic, with one set used to construct the copula and another set of spot and swap prices used as an output distribution from the copula. The swap price model also underestimates the volatility structure and contributes to higher optimal hedge ratios. Preferably one single pricing model able to simulate a long history of spot and swap prices consistent with today's pricing level and independent of the production should have been used. Second, price hedging has only been assessed in this chapter and not production risk. This is due to the nonexistent market for weather derivatives in the Nord Pool area which can allow producers to hedge their inflow risk and thereby the production uncertainty. Finally, prices and production volumes are seasonally dependent and the natural revenue variations based on the seasonal fluctuation are to some extent attempted hedged away. The optimal hedge ratios for a hydropower producer might therefore be lower than those recommended in this chapter, since yearly variations are more interesting for a hydropower producer than seasonal fluctuations. Quarterly data are considered to provide a sufficient sample size for the empirical copula estimation. Another effect of using quarterly and not annual data is that the autocorrelation of the input data to the empirical copula is higher. This results in some scenarios with a higher probability than what is actually the case. Consequently, the copula-based Monte Carlo simulation generates more of these scenarios, affecting the analysis of the output data sample. This may in worst case give misleading results. A purely empirical copula approach for price and production modeling can therefore be problematic. In further research it might be interesting to go beyond the empirical framework and make more assumptions to deal with seasonality, autocorrelation, and lack of data.

13.7 Acknowledgements

We greatly appreciate the contribution from Siri Line Hove Ås and Henning Nymann at TrønderEnergi for providing data and answering all our questions. Further, we would like to thank the staff at Nord Pool and the Norwegian Water Resources and Energy Directorate (NVE) for providing us with necessary data. We are

also grateful for comments at the ElCarbonRisk workshop in Molde May 21–22, 2012. We recognize the Norwegian research center CenSES, Centre for Sustainable Energy Studies, and acknowledge financial support from the Research Council of Norway through project 199904.

References

1. Adam TR, Fernando CS (2006) Hedging speculation and shareholder value. J Financ Econ 81(2), 283–309
2. Alexander C (2008) Practical Financial Econometrics, Market Risk Analysis, vol 2. Wiley, Chichester
3. Alexander C (2008) Quantitative Methods in Finance, Market Risk Analysis, vol 1. Wiley, Chichester
4. Andresen A, Sollie JM Hybrid modelling of the electricity spot price in the Nordic electricity market (2011). Working paper, Norwegian University of Science and Technology
5. Benth FE, Kettler PC (2011) Dynamic copula models for the spark spread. Quant Finance 11(3):407–421
6. Benth FE, Koekebakker S (2008) Stochastic modeling of financial electricity contracts. Energ Econ 30(3):1116–1157
7. Bessembinder H, Lemmon ML (2002) Equilibrium pricing and optimal hedging in electricity forward markets. The J Finance 57(3):1347–1382
8. Botterud A, Bhattacharyya AK, Ilic M (2002) Futures and spot prices an analysis of the Scandinavian electricity market. In: Proceedings of the 34th Annual North American Power Symposium (NAPS) 2002
9. Botterud A, Kristiansen T, Ilic MD (2010) The relationship between spot and futures prices in the Nord Pool electricity market. Energ Econ 32:967–978
10. Cartea A, Villaplana P (2008) Spot price modeling and the valuation of electricity forward contracts: the role of demand and capacity. J Bank Finance 32, 2502–2519
11. Cortazar G, Naranjo L (2006) An N-factor Gaussian model of oil futures prices. The J Futures Markets 26(3):243–268
12. Deng SJ, Oren SS (2006) Electricity derivatives and risk management. Energy 31:940–953
13. Durbin J, Koopman SJ (2001) Time series analysis by state space methods. Oxford University Press
14. Ederington LH (1979) The hedging performance of the new futures markets. The J Finance 34(1):157–170
15. Eichhorn A, Römisch W (2005) Wegner I Mean-risk optimization of electricity portfolios using multiperiod polyhedral risk measures. In: IEEE Power Tech, 2005. IEEE

16. Embrechts P, McNeil A, Straumann D (2002) Correlation and dependence in risk management: properties and pitfalls. In: Dempster MAH (ed) Risk management: value at risk and beyond, Cambridge University Press, Cambridge
17. Fama EF, French KR (1987) Commodity futures prices: Some evidence on forecast power, premiums, and the theory of storage. The J Bus 60(1):55–73
18. Fleten SE, Bråthen E, Nissen-Meyer SE (2010) Evaluation of static hedging strategies for hydropower producers in the Nordic market. J Energ Markets 4(3):1–28
19. Fleten SE, Wallace SW, Ziemba WT Hedging electricity portfolios using stochastic programming. In: Greengard C, Ruszczyński A (eds) Decision making under Uncertainty: energy and power, IMA volumes on mathematics and its applications, vol 128, pp. 71–93. Springer, New York (2002)
20. Genest C, Gendron M (2009) Bourdeau-Brien M The advent of copulas in finance. The Eur J Finance 15:609–618
21. Hirshleifer D(1990) Hedging pressure and futures price movements in a general equilibrium model. Econometrica 58(2):411–428
22. Kettunen J, Salo A, Bunn DW(2010) Optimization of electricity retailer's contract portfolio subject to risk preferences. IEEE Trans Power Syst 25(1): 117–128
23. Longstaff FA, Wang AW (2004) Electricity forward prices: a high-frequency empirical analysis. J Finance 59:1877–1900
24. Lucia JJ, Schwartz ES (2002) Electricity prices and power derivatives. Evidence from the Nordic Power Exchange. Rev Derivatives Res 5(1):5–50
25. Lucia JJ, Torró H (2011) On the risk premium in Nordic electricity futures prices. Int Rev Econ Finance 20:750–763
26. Markowitz H (1952) Portfolio selection. J Finance 7, 77–91
27. Nord Pool: Trade at Nord Pool ASA's financial market. Technical report Nord Pool ASA (2010)
28. Paravan D, Sheble GB, Golob R Price risk and volume risk management for power producers. In: International conference on probabilistic methods applied to power systems (2004)
29. Rockafellar RT, Uryasev S (2000) Optimization of conditional value-at-risk. J Risk 2:21–42
30. Sanda GE, Olsen ET, Fleten SE (2013) Selective hedging in hydro-based electricity companies. Energ Econ 40:326–338
31. Sklar A (1959) Fonctions de répartition à n dimensions et leurs marges, vol 8. Publications de l'Institut de Statistique de l'Université de Paris
32. Stulz RM (1996) Rethinking risk management. J Appl Corp Finance 9:8–25
33. Trivedi PK, Zimmer DM (2005) Copula modeling: An introduction for practitioners. Found Trends in Econometrics 1(1):1–111
34. Unger G (2002) Hedging strategy and electricity contract engineering. Ph.D. thesis, Swiss Federal Institute of Technology
35. Vehviläinen I, Keppo J (2003) Managing electricity market price risk. Eur J Oper Res 145(1):136–147

Chapter 14
Investment in Stochastic Electricity-Production Facilities

Luis Baringo and Antonio J. Conejo

Abstract This chapter considers a profit-oriented private investor interested in building stochastic electricity-production facilities, such as solar and wind power plants. This investor sells its production in a competitive pool-based electricity market and faces uncertainties related to demand growth, its production level, and its investment cost. Adopting a multistage approach, a stochastic complementarity model is formulated to determine the optimal capacity to be built by the investor to maximize its expected profit while minimizing its profit volatility. An example considering a wind power investor is presented to illustrate the working of the proposed model.

14.1 Investment in Stochastic Electricity-Production Facilities

14.1.1 Generation Capacity Investment

Generation capacity investment constitutes a relevant problem in electricity markets. To tackle this problem two different approaches are generally considered: a centralized framework [1] and a market framework [2].

The first of these approaches, i.e., a centralized framework [1], usually determines the generation capacity expansion plan based on a worst scenario case and considering the whole electric energy system. This approach, although relevant for the efficient operation of the system as a whole, is not of interest for a particular investor aiming at maximizing its own expected profit.

The market approach represents the operation of the market in which producers and consumers participate. This approach allows representing the perspective

L. Baringo • A.J. Conejo (✉)
University of Castilla-La Mancha, Ciudad Real, Spain
e-mail: luis.baringo@gmail.com; antonio.conejo@uclm.es

R.M. Kovacevic et al. (eds.), *Handbook of Risk Management in Energy Production and Trading*, International Series in Operations Research & Management Science 199,
DOI 10.1007/978-1-4614-9035-7_14,

of a profit-oriented investor. Within this framework, there are several approaches for the generation capacity expansion (or investment) problem, e.g., static [3] or dynamic [4]; strategic [2] or competitive [5]; and with [6] or without network representation [7].

In this chapter, a market framework and a multistage approach are considered to address a particular but important case of the generation capacity investment problem: the investment in stochastic electricity-production facilities. Stochastic units are those units whose production is variable and uncertain, e.g., solar and wind power units. Note that the production of stochastic units depends on the availability of a natural resource such as solar intensity or wind speed. Thus, conventional models have to be modified to incorporate the uncertain production characteristics of this type of generation units [8, 9].

14.1.2 Uncertainty

There are several parameters that influence investment decisions, e.g., the future demand growth, equipment outages, investment costs, generation costs, and the production of stochastic units. Most of these parameters are subject to uncertainty and thus, an adequate modeling of such uncertainty is a must to achieve effective investment decisions.

Among the parameters subject to uncertainty, those with the greatest influence in the investment in stochastic electricity-production facilities are the demand growth, the production of stochastic units, and the investment cost. This is so because:

1. Under a market framework, the demand in the system and its growth directly influence the market clearing prices that in turn influence the investment decisions.
2. The uncertainty in the production of stochastic units influences the required capacity to be installed.
3. As the technology related to stochastic units matures, their future investment costs are expected to decrease. However, the decrease rate is uncertain, which has a significant impact on investment decisions.

The above sources of uncertainty are represented through a set of scenarios within a stochastic programming model [10, 11], an appropriate model to address this type of investment problems. Section 14.2 of this chapter provides the methodologies used for modeling of different sources of uncertainty.

14.1.3 Planning Horizon

A multistage approach is adopted by considering a planning horizon comprising a specific number of time periods. Each time period spans a known number of years.

This multistage approach allows making investment decisions at different points in time and provides flexibility to adapt to changes in the conditions that influence the investment decisions.

In order to characterize the different sources of uncertainty, each time period is represented by a single year within the corresponding period (e.g., the last year), which is considered the reference year of the whole time period.

The investment decisions concerning the capacities of the stochastic production units to be built are made at the beginning of each time period.

14.1.4 Risk Management

The key objective of a private investor is maximizing its expected profit. However, maximizing expected profit may lead to extreme cases in which the investor achieves a very high profit in some scenarios but incurs very high losses in others. The investor may not be able to assume such losses and might prefer to reduce its profit volatility despite having to reduce its expected profit as well. Thus, it is important to introduce a metric to control the risk of profit volatility. To do so, we use the conditional value-at-risk (CVaR) metric [12, 13] that can be easily implemented through linear constraints.

In a profit maximization problem, the CVaR is defined as the expected value of the profit smaller than the $(1-\alpha)$-quantile of the profit distribution, being α a given confidence level.

There are two manners of implementing the CVaR in a multistage model such as the one considered in this chapter. The first one seeks to reduce the risk of profit volatility per period. The second one seeks to reduce this risk throughout the whole planning horizon. This second approach is used in this chapter.

14.1.5 Complementarity Model

The considered investor aims at making investment decisions to maximize its expected profit from selling its production in the market while minimizing its profit volatility. However, these investment decisions are related to the market in which the investor sells its production once the newly built capacity is ready to operate. Note that the production of the newly built units influences the clearing of the market and that the outputs of the market influence in turn the investment decisions. Thus, the investment model must incorporate the clearing of the market as an additional constraint.

However, note that the market clearing is itself an optimization problem that seeks to maximize social welfare or to minimize generation cost. Thus, the investment model becomes an optimization problem constrained by other optimization problem, i.e., a complementarity model.

Figure 14.1 illustrates the structure of this complementarity model which comprises an upper-level problem and a collection of lower-level problems, i.e., a bi-level model. On one hand, the upper-level problem is an optimization problem that aims at maximizing the investor expected profit while minimizing its profit volatility. On the other hand, the lower-level problems represent the market clearing under different operating conditions, scenarios, and time periods. Note that investment decisions influence the market clearing (through the production of the newly built units), and that the investor obtains the clearing prices from the market clearing problems, which in turn influence its investment decisions.

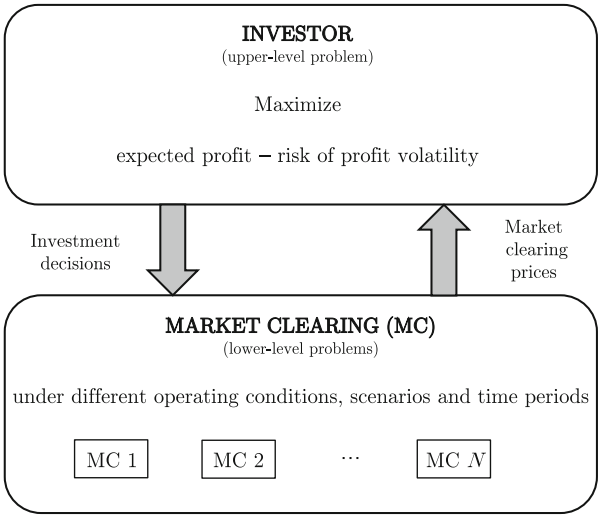

Fig. 14.1 Complementarity model structure

14.1.6 Chapter Organization

The remaining of this chapter is organized as follows. Section 14.2 describes the modeling of the sources of uncertainty. Section 14.3 provides a risk-constrained multistage bi-level model to decide the optimal investment in stochastic electricity-production facilities, as well as a procedure to transform such bi-level model into a mathematical program with equilibrium constraints (MPEC) and then into a mixed-integer linear programming (MILP) model. Both Sects. 14.2 and 14.3 include clarifying examples to illustrate the uncertainty modeling and the bi-level model, respectively. Finally, Sect. 14.4 summarizes the chapter and provides some relevant remarks and conclusions.

14.2 Uncertainty Modeling

This section describes the modeling of the sources of uncertainty that affect the investment decisions in stochastic electricity-production facilities. The explanations below are given for the particular case of investment in wind power facilities. However, extending the analysis to consider alternative stochastic-production units is straightforward.

14.2.1 Sources of Uncertainty

As explained in Sect. 14.1.2 of this chapter, there are three main sources of uncertainty that influence wind power investment decisions, namely, the future demand growth, the wind power production, and the future investment cost.

For the sake of simplicity and due to their high impact on investment decisions, only these three sources of uncertainty are considered below. However, additional sources of uncertainty (e.g., generation costs and equipment outages) can be considered in a similar manner.

14.2.2 Demand and Wind Power Production Uncertainty

The demand and the wind power production are usually anticorrelated since low demands (during the night) generally correspond to high wind power productions and high demands (during the day) generally correspond to low wind power productions. Thus, the uncertain character of both parameters has to be addressed jointly in order to account for this negative correlation.

The aim of the wind power investor is to decide the wind power units to be built throughout an existing electric energy system at the beginning of each of the time periods comprising the planning horizon. As explained in Sect. 14.1.3, each of these time periods is modeled using a representative year. The demand and wind power production uncertainties throughout the representative years are modeled as described below.

We consider hourly historical values (throughout one or several past years) of demand and wind speed in different locations of the electric energy system under study. First, wind speed values are transformed into wind power capacity factors (defined as the wind power production per MW) through appropriate wind speed/wind power production curves. Second, demand values in each location are divided by the peak demand in each particular location, rendering a set of values of demand factors. We obtain hourly values of the demand factor and of the wind power capacity factor in each location of the electric energy system under study which represent different operating conditions.

This set of values represents the historical demand and wind power production profile in the electric energy system under study. This profile also accounts for the correlation among demand and wind power capacity factors. For the sake of simplicity, this demand and wind power production profile is considered fixed throughout the whole planning horizon. Thus, these historical demand and wind power capacity factors are considered to represent the demand and wind power production uncertainty in the representative year of each time period of the planning horizon. It is important to emphasize that wind power capacity factors do not change throughout the years but the demand grows. However, for simplicity we assume that all demands grow in the same proportion, i.e., the *geometry* remains unaltered.

Considering all historical operating conditions as input data of the investment model may result in intractability, particularly for realistic electric energy systems. Thus, the K-means clustering method [14, 15] is used to transform the historical data into a reduced data set that maintains the information of and the correlation among the demand and the wind power capacity factors of the historical data.

The working of the K-means technique to reduce the historical operating condition is summarized below.

We define a cluster as a group of observations (e.g., demand and wind power capacity factors in different locations) that are similar among them but different from the observations in other clusters. Additionally, we define the centroid of each cluster as the mean value of the observations allocated to the cluster. Given these definitions, the K-means follows the iterative algorithm below:

1. Select an appropriate number of clusters.
2. Define initial clusters and the initial centroid of each cluster, e.g., randomly allocating the observations to different clusters.
3. Compute the distances (e.g., quadratic distances) between each historical observation and each cluster centroid.
4. Allocate each historical observation to the closest cluster according to the calculated distances.
5. Recalculate the centroid of each cluster.

Steps 3–5 above are repeated iteratively until there is no change in the compositions of the clusters in two consecutive iterations.

Note that the output of the K-means technique is a set of clusters, each one defined by its centroid and the number of historical observations allocated to it. The centroids comprise the values of the demand and the wind power capacity factors in each location of the system under study, which represent the system operating condition. On the other hand, the number of historical observations allocated to each cluster gives the weight of the cluster. As we represent each period using a reference year, we define this weight as the number of hours in the representative year of each period that are represented by each cluster, i.e., the weight of each operating condition in each representative year. Additional details of the K-means technique can be found in [14].

Demand and wind power capacity factors as per unit values are considered fixed throughout the planning horizon. However, this is not the case for the demand and

the wind power production which generally increase. Such variations have to be properly represented. On one hand, the wind power production depends on the installed wind power capacity which is a decision variable of the problem and, thus, its increase is determined by the solution of the problem. On the other hand, the uncertainty in the demand growth throughout the planning horizon is represented using a set of scenarios as depicted in the scenario tree of Fig. 14.2. In this particular example of two time periods, the scenario tree comprises two demand growth scenario realizations in the first time period and two demand growth scenario realizations in the second period for each scenario realization in the first one, which results in four demand growth scenarios (D1, D2, D3, and D4) for the whole planning horizon. In this example, there is only one possible investment decision at the beginning of the planning horizon (i.e., at the beginning of the first period) and two alternative investment decisions at the beginning of the second period depending on the scenario realization in the first one: one investment decision for scenarios D1 and D2; and one investment decision for scenarios D3 and D4.

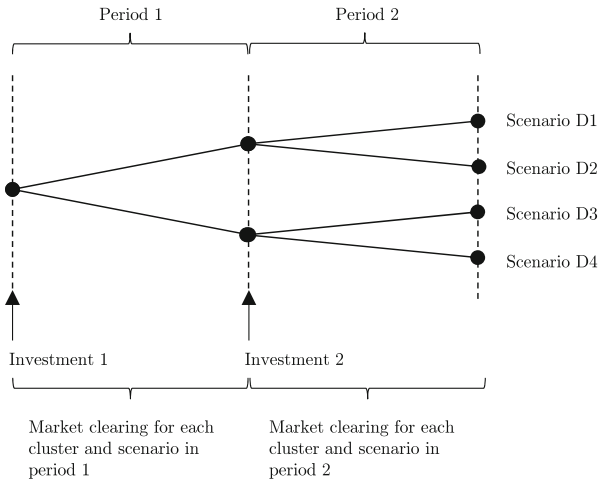

Fig. 14.2 Scenario tree considering uncertainty in demand growth

As explained above, demand and wind power capacity factors are considered fixed throughout the planning horizon. This assumption implies that all demands grow in the same proportion. Note that this may not be the case, specially if the system under study has (or may have in the future) a smart grid technology enabling the implementation of demand response programs that may play a significant role in changing the consumer behavior and thus, the *geometry* of the demand. However, note that if this is the case, different demand and wind power production profiles may be considered through additional scenarios.

14.2.3 Investment Cost Uncertainty

As the wind power technology matures, the future wind power investment cost is expected to decrease. However, this potential decrease is uncertain and thus, this uncertainty has to be properly modeled. To do so, we use a set of scenarios to represent the investment cost uncertainty.

Figure 14.3 depicts an example of a scenario tree representing the investment cost uncertainty. In this example, there are three possible investment cost scenario realizations for the second period (IC1, IC2, and IC3). In this case, there is a single investment decision at the beginning of the first period and three alternatives investment decisions at the beginning of the second one, depending on the investment cost scenario realization.

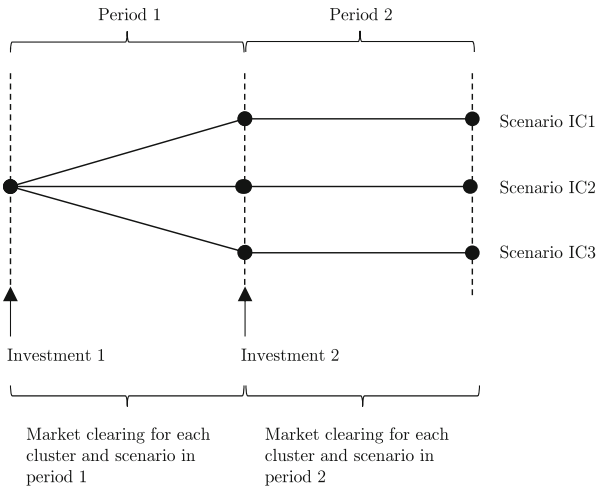

Fig. 14.3 Scenario tree considering uncertainty in investment cost

The main difference in the uncertainty modeling of the investment cost and in the demand growth is that in the first case, the wind power investor knows the actual investment cost at the time it makes its investment decisions. The investment cost in the first period is known at the point in time that the investment decisions for this period are made. However, at this point in time, the investor does not know the investment cost in the second period. Nevertheles, this investment cost is known at the time the investor makes its investment decisions. On the other hand, the demand growth in the first period is not known at the time the wind power investor makes its investment decisions for this period. Thus, the uncertainty in demand growth usually entails a higher risk than the uncertainty associated with investment costs.

The demand growth and the investment cost scenarios are generated using appropriate forecasting tools [16, 17].

14.2.4 Decision Sequence: Scenario Tree

The sources of uncertainty described in Sects. 14.2.2 and 14.2.3 above are independent and thus, the final scenario tree includes all possible scenario combinations. For the considered examples in Figs. 14.2 and 14.3, the final scenario tree comprises twelve scenario realizations (four demand growth scenarios and three investment cost scenarios) as depicted in Fig. 14.4. In this case, there is a single investment decision at the beginning of the planning horizon which does not depend on the future scenario realizations and six possible investment decisions at the beginning of the second period depending on the demand growth and the investment cost scenario realization in the first period (six alternatives).

It is important to note that each of these scenarios comprises a specific number of clusters (those obtained using the K-means technique and representing different system operating conditions) for each time period, which account for the demand and wind power capacity factors variability throughout the reference year of each period.

Given this framework, the decision sequence is as follows:

1. At the beginning of the planning horizon, i.e., at the beginning of the first time period, the wind power investor decides the wind power capacity to be built at this point in time. These investment decisions are *here-and-now* decisions since they do not depend on any scenario realization.
2. Once the investment decisions for the first period are made, the market is cleared for each cluster within each scenario in the first period. From the market clearing we obtain power productions, power flows, market clearing prices, etc.
3. The first period concludes and the wind power investor knows the actual scenario realization in this period (i.e., it knows the demand growth in the first period and the investment cost for the second period). Depending on the scenario realization, the investor makes its investment decisions for the second period. These are *wait-and-see* decisions with respect to the first period since they do depend on the scenario realization in this period and *here-and-now* decisions with respect to the second period (and the following ones if there are more than two) since they do not depend on the scenario realization in the future periods.
4. Once the investment decisions for the second period are made, the market is cleared for each cluster within each scenario in the second period.

Finally, steps 3 and 4 are repeated for each time period of the planning horizon if there were more than two.

14.2.5 Illustrative Example

In order to illustrate the modeling of the different sources of uncertainty, we consider the three-node electric energy system depicted in Fig. 14.5. This system comprises three nodes, three demands, three generation units, and three transmission lines.

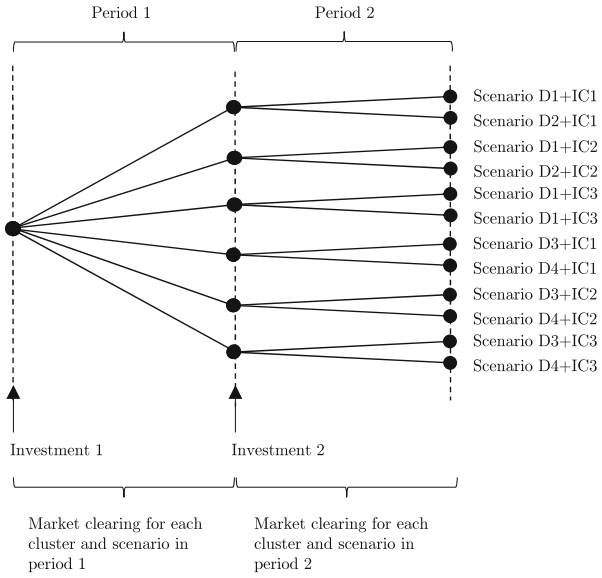

Fig. 14.4 Scenario tree considering uncertainty in both demand growth and investment cost

There are two wind zones: North wind zone (comprising nodes 1 and 2) and south wind zone (node 3).

Data of demand is obtained from aggregated historical data throughout year 2008 in the Iberian Peninsula [18]. The demand values are then divided by the peak demand to obtain demand factors. On the other hand, historical wind speed data of the Spanish towns of Tortosa (Northeast Spain) and Tarifa (Southwest Spain) are considered to characterize the wind speed in the north and south zones, respectively. These wind speed data are obtained using the databases developed by the University of Cantabria [19, 20]. To obtain the corresponding wind power capacity factors, we consider the wind speed/wind power production curve of a Nordex N80/2500 turbine [21]. Finally, we consider a two-period planning horizon, each one comprising 6 years.

14.2.5.1 Clusters

Historical data of demand and wind power capacity factors comprise 8,760 sets of values (demand and wind power capacity factors in each zone and for each hour of year 2008). Each set represents a system operating condition. If we use these historical data as input data of an investment model, we may face intractability. Thus, we apply the K-means technique explained in Sect. 14.2.2 to reduce these historical data into a set of clusters.

Table 14.1 provides the demand and wind power capacity factor data of the ten clusters in which we allocate the historical data. The second column of this table

gives the demand factors, which for the sake of simplicity are assumed to be the same at all the nodes of the system. The third and fourth columns provide the wind power capacity factors in the north and south wind zones, respectively. Both the demand and wind power capacity factors within each cluster represent a system operating condition. Finally, the fifth column gives the number of historical observations that are allocated to each cluster, which represent the weight of each cluster in the reference year of each period.

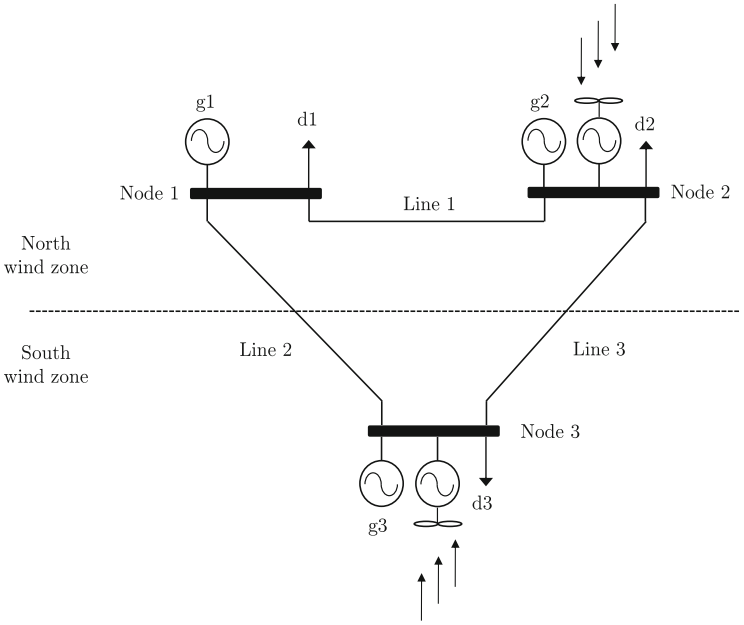

Fig. 14.5 Three-node electric energy system

This set of ten clusters covers the information of the historical observations as well as the correlation among demand and wind power capacity factors in different zones.

14.2.5.2 Demand Growth Scenarios

First, we assume that the future wind power investment cost is not subject to uncertainty. For this purpose, we consider the scenario depicted in Fig. 14.2 comprising two demand growth scenario realizations in the first period and two demand growth scenario realizations in the second period depending on the scenario realization in the first one.

The probability that the demand in the first period is 15 % higher than the demand prior to the beginning of the planning horizon is 0.7, while the probability of being

Table 14.1 Cluster results: demand and wind power capacity factor data

Cluster	Demand factors [p.u.]	Capacity factors [p.u.]		Number of hours [h]
		Wind zone		
		North	South	
1	0.8091	0.0331	0.1211	1712
2	0.7362	0.9191	0.1230	715
3	0.7158	0.7389	0.8100	216
4	0.6674	0.6152	0.0962	477
5	0.6885	0.5308	0.5028	300
6	0.7358	0.0598	0.8771	1312
7	0.7138	0.0406	0.4798	1365
8	0.6214	0.0268	0.0773	1810
9	0.8353	0.2746	0.1266	586
10	0.8286	0.4517	0.0933	267

10 % lower than this demand is 0.3. On the other hand, in the second period, the demand may be 10 % higher and 5 % lower than the demand in the first period with probabilities of 0.6 and 0.4, respectively. Table 14.2 summarizes the demand growth scenario data. The total weight of each scenario is computed as the product of the probability of occurrence of the demand variations in the first and second periods.

Table 14.2 Uncertainty in demand growth: scenario data

Scenario	Period 1		Period 2		Total weight
	Demand growth (%)	Weight	Demand growth (%)	Weight	
D1	+15	0.7	+10	0.6	0.42
D2			−5	0.4	0.28
D3	−10	0.3	+10	0.6	0.18
D4			−5	0.4	0.12

14.2.5.3 Investment Cost Scenarios

In this case we assume that the uncertainty does not affect the demand growth and we consider the scenario tree depicted in Fig. 14.3, which considers three different

investment costs in the second period. The probability that the investment cost in the second period is 15 % higher than that in the first one is 0.2; the probability of being equal to the investment cost in the first period is 0.4, while the probability that the investment cost in the second period is 25 % lower than in the first one is 0.4. Table 14.3 summarizes the investment cost scenario data.

Table 14.3 Uncertainty in investment cost: scenario data

Scenario	Investment cost variation in period 2 (%)	Weight
IC1	+15	0.2
IC2	0	0.4
IC3	−25	0.4

14.2.5.4 Demand Growth and Investment Cost Scenarios

Finally, we consider that the uncertainty affects both the demand growth and the future investment cost. We consider all the scenario combinations of Figs. 14.2 and 14.3 which result in the final scenario tree depicted in Fig. 14.4. Table 14.4 summarizes the scenario data for this case.

There are 12 scenarios for the whole planning horizon as a result of four demand growth scenarios and three investment cost scenarios. The weight of each scenario is computed as the weight of the demand growth scenario times the weight of the investment cost scenario. For example, the weight of scenario D1+IC1 (0.084) is obtained as the weight of scenario D1 (0.42) times the weight of scenario IC1 (0.2).

14.3 Risk-Constrained Multistage Wind Power Investment

In this section, a risk-constrained multistage model is formulated to determine the wind power capacity to be built by a wind power investor to maximize its expected profit while minimizing its profit volatility. This model is based on a bi-level model which can be recast as a MILP problem.

14.3.1 Notation

The notation of the proposed model is provided below for quick reference.

Table 14.4 Uncertainty in both demand growth and investment cost: scenario data

Scenario	Period 1	Period 2		Total weight
	Demand growth (%)	Investment cost variation (%)	Demand growth (%)	
D1+IC1	+15	+15	+10	0.084
D2+IC1			−5	0.056
D1+IC2		0	+10	0.168
D2+IC2			−5	0.112
D1+IC3		−25	+10	0.168
D2+IC3			−5	0.112
D3+IC1	−10	+15	+10	0.036
D4+IC1			−5	0.024
D3+IC2		0	+10	0.072
D4+IC2			−5	0.048
D3+IC3		−25	+10	0.072
D4+IC3			−5	0.048

14.3.1.1 Indices

c	Index for clusters
d	Index for demands
g	Index for generation units (other than candidate wind power units)
l	Index for transmission lines
n	Index for nodes
$r(l)$	Index of the receiving-end node of line l
$s(l)$	Index of the sending-end node of line l
t	Index for time periods
ω	Index for scenarios

14.3.1.2 Constants

a_t	Amortization factor in the tth time period
c_g^{G}	Marginal cost of the gth generation unit
$c_t^{\mathrm{inv}}(\omega)$	Wind power investment cost in the tth time period and scenario ω
$c_t^{\max}$	Investment budget in the tth time period
$k_{d,c}^{\mathrm{D}}$	Demand factor of the dth demand in the cth cluster
$k_{n,c}^{\mathrm{W}}$	Wind power capacity factor at node n in the cth cluster
N_c^{H}	Number of hours comprising the cth cluster
$P_{d,t}^{\mathrm{D}}(\omega)$	Peak load of the dth demand in the tth time period and scenario ω
$P_g^{\mathrm{G,max}}$	Capacity of the gth generation unit

$P_l^{\mathrm{L,max}}$	Transmission capacity of the lth transmission line
x_l	Reactance of line l
$X_n^{\mathrm{W,max}}$	Maximum wind power capacity that can be built at node n throughout the planning horizon
α	Confidence level used to compute the CVaR
β	Weighting parameter used to model the trade-off between expected profit and CVaR
$\gamma(\omega)$	Weight of scenario ω

14.3.1.3 Variables

$P_{g,c,t}^{\mathrm{G}}(\omega)$	Power produced by the gth generation unit in the cth cluster, the tth time period, and scenario ω
$P_{l,c,t}^{\mathrm{L}}(\omega)$	Power flow through the lth transmission line in the cth cluster, the tth time period, and scenario ω
$P_{n,c,t}^{\mathrm{W}}(\omega)$	Wind power production at node n in the cth cluster, the tth time period, and scenario ω
$X_{n,t}^{\mathrm{W}}(\omega)$	Wind power capacity to be built at node n at the beginning of the tth time period, and scenario ω
$\delta_{n,c,t}(\omega)$	Voltage angle at node n in the cth cluster, the tth time period, and scenario ω
$\vartheta_{n,c,t}(\omega)$	LMP at node n in the cth cluster, the tth time period, and scenario ω
$\eta(\omega), \zeta$	Auxiliary variables used to compute the CVaR

14.3.1.4 Dual Variables

The dual variables below are associated with the following constraints:

$\chi_{n,c,t}(\omega)$	Zero voltage angle at the reference node in the cth cluster, the tth time period, and scenario ω
$\lambda_{n,c,t}(\omega)$	Generation-demand balance at node n in the cth cluster, the tth time period, and scenario ω
$\phi_{l,c,t}(\omega)$	Power flow through transmission line l in the cth cluster, the tth time period, and scenario ω
$\phi_{l,c,t}^{\max}(\omega)$	Capacity of transmission line l in direction $s(l)$-$r(l)$ in the cth cluster, the tth time period, and scenario ω
$\phi_{l,c,t}^{\min}(\omega)$	Capacity of transmission line l in direction $r(l)$-$s(l)$ in the cth cluster, the tth time period, and scenario ω
$\varphi_{g,c,t}^{\max}(\omega)$	Capacity of generation unit g in the cth cluster, the tth time period, and scenario ω
$\varphi_{g,c,t}^{\min}(\omega)$	Nonnegativity of the production of generation unit g in the cth cluster, the tth time period, and scenario ω
$\xi_{n,c,t}^{\max}(\omega)$	Upper limit of voltage angle at node n in the cth cluster, the tth time period, and scenario ω
$\xi_{n,c,t}^{\min}(\omega)$	Lower limit of voltage angle at node n in the cth cluster, the tth time period, and scenario ω

14.3.1.5 Sets

Δ^{C}	Set of indices of clusters
Δ^{G}	Set of indices of generation units (other than candidate wind power units)
Δ^{N}	Set of indices of nodes
Δ^{T}	Set of indices of time periods
Δ^{ω}	Set of indices of scenarios
Λ_n^{D}	Set of indices of demands located at node n
Λ_n^{G}	Set of indices of generation units (other than candidate wind power units) located at node n
$\Psi_t(\omega)$	Set of parameters defining scenario ω in the tth time period

14.3.2 Bi-level Formulation

The considered risk-constrained multistage wind power investment model is formulated below [9]:

$$\underset{\Omega^{\mathrm{U}} \cup \Omega_{c,t}^{\mathrm{L}}(\omega)}{\text{Maximize}}$$

$$\sum_{\omega \in \Delta^{\omega}} \gamma(\omega) \left\{ \sum_{t \in \Delta^{\mathrm{T}}} \left[\sum_{c \in \Delta^{\mathrm{C}}} N_c^{\mathrm{H}} \sum_{n \in \Delta^{\mathrm{N}}} \vartheta_{n,c,t}(\omega) P_{n,c,t}^{\mathrm{W}}(\omega) \; - a_t \sum_{n \in \Delta^{\mathrm{N}}} c_t^{\mathrm{inv}}(\omega) X_{n,t}^{\mathrm{W}}(\omega) \right] \right\} + \beta \left(\zeta - \frac{1}{1-\alpha} \sum_{\omega \in \Delta^{\omega}} \gamma(\omega) \eta(\omega) \right) \tag{14.1a}$$

subject to

$$0 \le P_{n,c,t}^{\mathrm{W}}(\omega) \le k_{n,c}^{\mathrm{W}} \sum_{b \le t} X_{n,t}^{\mathrm{W}}(\omega), \quad \forall n, \forall c, \forall t, \forall \omega \tag{14.1b}$$

$$\sum_{n \in \Delta^{\mathrm{N}}} c_t^{\mathrm{inv}}(\omega) X_{n,t}^{\mathrm{W}}(\omega) \le c_t^{\max}, \quad \forall t, \forall \omega \tag{14.1c}$$

$$0 \le \sum_{t \in \Delta^{\mathrm{T}}} X_{n,t}^{\mathrm{W}}(\omega) \le X_n^{\mathrm{W,max}}, \quad \forall n, \forall \omega \tag{14.1d}$$

$$\vartheta_{n,c,t}(\omega) = \lambda_{n,c,t}(\omega), \quad \forall n, \forall c, \forall t, \forall \omega \tag{14.1e}$$

$$X_{n,t}^{\mathrm{W}}(\omega_i) = X_{n,t}^{\mathrm{W}}(\omega_{\tilde{i}}), \quad \forall n, \forall t, \forall \omega_i, \forall \omega_{\tilde{i}} : \Psi_b(\omega_i) = \Psi_b(\omega_{\tilde{i}}), \forall b < t \tag{14.1f}$$

$$\zeta - \sum_{t \in \Delta^{\mathrm{T}}} \left[\sum_{c \in \Delta^{\mathrm{C}}} N_c^{\mathrm{H}} \sum_{n \in \Delta^{\mathrm{N}}} \vartheta_{n,c,t}(\omega) P_{n,c,t}^{\mathrm{W}}(\omega) - a_t \sum_{n \in \Delta^{\mathrm{N}}} c_t^{\mathrm{inv}}(\omega) X_{n,t}^{\mathrm{W}}(\omega) \right] \le \eta(\omega), \quad \forall \omega \tag{14.1g}$$

$$\eta(\omega) \ge 0, \quad \forall \omega, \tag{14.1h}$$

where $\lambda_{n,c,t}(\omega), \forall n, \in \arg\left\{ \underset{\Omega^{\mathrm{L}}_{c,t}(\omega)}{\text{Minimize}} \right.$

$$\sum_{g \in \Delta^{\mathrm{G}}} c^{\mathrm{G}}_{g} P^{\mathrm{G}}_{g,c,t}(\omega) \tag{14.2a}$$

subject to

$$\sum_{g \in \Lambda^{\mathrm{G}}_{n}} P^{\mathrm{G}}_{g,c,t}(\omega) + P^{\mathrm{W}}_{n,c,t}(\omega) - \sum_{l|s(l)=n} P^{\mathrm{L}}_{l,c,t}(\omega) + \sum_{l|r(l)=n} P^{\mathrm{L}}_{l,c,t}(\omega) = \sum_{d \in \Lambda^{\mathrm{D}}_{n}} P^{\mathrm{D}}_{d,t}(\omega) k^{\mathrm{D}}_{d,c} : \lambda_{n,c,t}(\omega), \quad \forall n \tag{14.2b}$$

$$P^{\mathrm{L}}_{l,c,t}(\omega) = \frac{1}{x_l}\left(\delta_{s(l),c,t}(\omega) - \delta_{r(l),c,t}(\omega)\right) : \phi_{l,c,t}(\omega), \quad \forall l \tag{14.2c}$$

$$-P^{\mathrm{L,max}}_{l} \le P^{\mathrm{L}}_{l,c,t}(\omega) \le P^{\mathrm{L,max}}_{l} : \phi^{\min}_{l,c,t}(\omega), \phi^{\max}_{l,c,t}(\omega), \quad \forall l \tag{14.2d}$$

$$0 \le P^{\mathrm{G}}_{g,c,t}(\omega) \le P^{\mathrm{G,max}}_{g} : \varphi^{\min}_{g,c,t}(\omega), \varphi^{\max}_{g,c,t}(\omega), \quad \forall g \tag{14.2e}$$

$$-\pi \le \delta_{n,c,t}(\omega) \le \pi : \xi^{\min}_{n,c,t}(\omega), \xi^{\max}_{n,c,t}(\omega), \quad \forall n \setminus n\text{: ref.} \tag{14.2f}$$

$$\delta_{n,c,t}(\omega) = 0 : \chi_{n,c,t}(\omega), \quad n\text{: ref.} \tag{14.2g}$$

$$\left.\right\}, \forall c, \forall t, \forall \omega,$$

where

$$\Omega^{\mathrm{U}} = \left\{ X^{\mathrm{W}}_{n,t}(\omega), \forall n, \forall t, \forall \omega; P^{\mathrm{W}}_{n,c,t}(\omega), \vartheta_{n,c,t}(\omega), \forall n, \forall c, \forall t, \forall \omega; \eta(\omega), \forall \omega; \zeta \right\}, \tag{14.3}$$

and

$$\Omega^{\mathrm{L}}_{c,t}(\omega) = \left\{ P^{\mathrm{G}}_{g,c,t}(\omega), \forall g; P^{\mathrm{L}}_{l,c,t}(\omega), \forall l; \delta_{n,c,t}(\omega), \forall n; \lambda_{n,c,t}(\omega), \forall n; \phi_{l,c,t}(\omega), \right.$$
$$\phi^{\min}_{l,c,t}(\omega), \phi^{\max}_{l,c,t}(\omega), \forall l; \varphi^{\min}_{g,c,t}(\omega), \varphi^{\max}_{g,c,t}(\omega), \forall g; \xi^{\min}_{n,c,t}(\omega), \xi^{\max}_{n,c,t}(\omega),$$
$$\left. \forall n \setminus n\text{: ref.}; \chi_{n,c,t}(\omega), n\text{: ref.} \right\}, \forall c, \forall t, \forall \omega. \tag{14.4}$$

The bi-level model (14.1) and (14.2) comprises an upper-level problem (14.1) and a set of lower-level problems (14.2), one for each cluster c, time period t, and scenario ω. The dual variable associated with each constraint of the lower-level problems is indicated after a colon.

The optimization variables of the lower-level problems are the variables in sets $\Omega^{\mathrm{L}}_{c,t}(\omega)$, $\forall c$, $\forall t$, $\forall \omega$. The lower-level problems constraint the upper-level one and thus, the optimization variables of the upper-level problem include variables in sets $\Omega^{\mathrm{L}}_{c,t}(\omega)$, $\forall c$, $\forall t$, $\forall \omega$ and additional variables in set Ω^{U}.

The upper-level problem (14.1) aims at maximizing the expected profit achieved by the wind power investor plus a coefficient times the CVaR. The CVaR is the metric used to control the risk of profit volatility [12, 13].

The objective function (14.1a) of the upper-level problem comprises three terms:

1. Each term $\vartheta_{n,c,t}(\omega) P^{\mathrm{W}}_{n,c,t}(\omega)$ is the revenue achieved by the wind power investor for selling its wind power production in the pool per cluster, time period, and scenario. The wind power investor is paid the locational marginal price (LMP) of the node at which wind power is produced times the wind power production. Since each time period is represented by a reference year, these revenues are multiplied by the number of hours within each cluster to obtain annual revenues in each time period and for each scenario.
2. Each term $c^{\mathrm{inv}}_{t}(\omega) X^{\mathrm{W}}_{n,t}(\omega)$ is the investment cost incurred by the wind power investor per time period and scenario for building new wind power capacity. These terms are multiplied in each period by an amortization rate a_t to make investment costs and revenues comparable across time periods.

   Terms in items 1 and 2 are multiplied in each scenario by the weight of the corresponding scenario to obtain expected profits.
3. Term $\beta \left(\zeta - \frac{1}{1-\alpha} \sum_{\omega \in \Delta^{\omega}} \gamma(\omega) \eta(\omega) \right)$ is a coefficient β times the CVaR. Coefficient β is a weighting factor that models the trade-off between expected profit and CVaR.

For the sake of simplicity, we consider that all the monetary values are referred to a single point in time and thus, it is not necessary to multiply them by discount rates.

Constraints (14.1b)–(14.1e) represent the wind power operation and investment limits and conditions for all clusters, periods, and scenarios. Constraints (14.1b) limit the wind power production to the wind power availability at each node and for each cluster, time period, and scenario. Note that the wind power capacity built at the beginning of any time period is available during this and the remaining periods of the planning horizon. Constraints (14.1c) impose a cap on investment budget for each time period and scenario. Equations (14.1d) enforce the nonnegativity of the wind power capacity to be built and limit it to a maximum throughout the planning horizon. Finally, constraints (14.1e) state that the price paid to the wind power investor for its production is the LMP of the node at which wind power is produced. These LMPs are obtained as the dual variables associated with the balance constraints in the lower-level problems for each node, cluster, time period, and scenario.

Constraints (14.1f) are non-anticipativity constraints that avoid anticipating information. They impose that the wind power capacity to be built at the beginning of a time period depends on the scenario realization of the previous periods, but it is unique for all the possible scenario realizations in the current and future periods [22].

Finally, constraints (14.1g) and (14.1h) allow computing the CVaR metric through linear expressions. Further information on the CVaR is available in [12, 13].

Additionally, the upper-level problem is also constrained by a set of lower-level problems (14.2) which represent the clearing of the pool for each cluster, time period, and scenario.

The objective function (14.2a) to be maximized is the social welfare, which is equivalent to minimizing generation costs since demands are considered fixed within each cluster, time period, and scenario; wind power investor offers its production at zero price; and producers other than the wind power investor offer their productions at their marginal costs.

Constraints of these lower-level problems include equalities (14.2b) that represent the generation-demand balance at each node of the system; equalities (14.2c) that define the power flows through transmission lines, limited to the transmission capacity of the corresponding lines by constraints (14.2d); constraints (14.2e) that impose upper and lower bounds on the power produced by generation units other than the wind power units; and finally constraints (14.2f) and (14.2g) that limit the voltage angles and fix equal to zero the voltage angle at the reference node, respectively. Note that the network topology is explicitly modeled by (14.2b)–(14.2d) using a direct current (dc) representation and disregarding losses [23].

14.3.3 MPEC Formulation

Bi-level model (14.1) and (14.2) is transformed into an MPEC following the procedure explained below.

Lower-level problems (14.2) represent the market clearing under different clusters, time periods, and scenarios and constraint the upper-level problem (14.1). Since each of these lower-level problems is linear and thus convex, the Karush–Kuhn–Tucker (KKT) conditions are necessary and sufficient optimality conditions [24]. Thus, each lower-level problem can be replaced by its KKT optimality conditions, which are in turn included as additional constraints of the upper-level problem rendering an MPEC, whose formulation is provided below:

$$\underset{\Omega^{\mathrm{U}} \cup \Omega^{\mathrm{L}}_{c,t}(\omega)}{\text{Maximize}} \quad (14.1a) \tag{14.5a}$$

subject to

$$\text{Constraints (14.1b)–(14.1h)} \tag{14.5b}$$

$$\left\{ \text{Constraints (14.2b)–(14.2g)} \right. \tag{14.5c}$$

$$c^{\mathrm{G}}_{g,t} - \lambda_{n(g),c,t}(\omega) + \varphi^{\max}_{g,c,t}(\omega) - \varphi^{\min}_{g,c,t}(\omega) = 0, \quad \forall g \tag{14.5d}$$

$$\lambda_{s(l),c,t}(\omega) - \lambda_{r(l),c,t}(\omega) - \phi_{l,c,t}(\omega) + \phi^{\max}_{l,c,t}(\omega) - \phi^{\min}_{l,c,t}(\omega) = 0, \quad \forall l \tag{14.5e}$$

$$\sum_{l|s(l)=n} \frac{1}{x_l} \phi_{l,c,t}(\omega) - \sum_{l|r(l)=n} \frac{1}{x_l} \phi_{l,c,t}(\omega) + \xi^{\max}_{n,c,t}(\omega) - \xi^{\min}_{n,c,t}(\omega) = 0,$$

$$\forall n \setminus n\text{: ref.} \tag{14.5f}$$

$$\sum_{l|s(l)=n} \frac{1}{x_l}\phi_{l,c,t}(\omega) - \sum_{l|r(l)=n} \frac{1}{x_l}\phi_{l,c,t}(\omega) - \chi_{n,c,t}(\omega) = 0, \quad n\text{: ref.} \tag{14.5g}$$

$$0 \le \phi^{\max}_{l,c,t}(\omega) \perp P^{\mathrm{L,max}}_{l} - f_{l,c,t}(\omega) \ge 0, \quad \forall l \tag{14.5h}$$

$$0 \le \phi^{\min}_{l,c,t}(\omega) \perp f_{l,c,t}(\omega) + P^{\mathrm{L,max}}_{l} \ge 0, \quad \forall l \tag{14.5i}$$

$$0 \le \varphi^{\max}_{g,c,t}(\omega) \perp P^{\mathrm{G,max}}_{g} - P^{\mathrm{G}}_{g,c,t}(\omega) \ge 0, \quad \forall g \tag{14.5j}$$

$$0 \le \varphi^{\min}_{g,c,t}(\omega) \perp P^{\mathrm{G}}_{g,c,t}(\omega) \ge 0, \quad \forall g \tag{14.5k}$$

$$0 \le \xi^{\max}_{n,c,t}(\omega) \perp \pi - \delta_{n,c,t}(\omega) \ge 0, \quad \forall n \setminus n\text{: ref.} \tag{14.5l}$$

$$0 \le \xi^{\min}_{n,c,t}(\omega) \perp \delta_{n,c,t}(\omega) + \pi \ge 0, \quad \forall n \setminus n\text{: ref.} \tag{14.5m}$$

$$\Big\}, \forall c, \forall t, \forall \omega.$$

MPEC (14.5) above is a single-level problem but includes different sources of nonlinearities, namely, the objective function (14.5a), constraints (14.1g), and the complementarity constraints (14.5h)–(14.5m). The following subsection explains how to transform this MPEC into an MILP problem that can be solved using traditional branch-and-cut techniques.

14.3.4 MILP Formulation

The MPEC (14.5) above has the following nonlinearities:

1. Each term $\sum_{n\in\Omega^{\mathrm{N}}} \vartheta_{n,c,t}(\omega) P^{\mathrm{W}}_{n,c,t}(\omega)$ in the objective function (14.5a) and in constraints (14.1g).
2. The complementarity constraints of the KKT conditions (14.5h)–(14.5m).

Each term $\sum_{n\in\Omega^{\mathrm{N}}} \vartheta_{n,c,t}(\omega) P^{\mathrm{W}}_{n,c,t}(\omega)$ can be replaced by an exact equivalent linear expression using the strong duality equality (SDE) as explained below [25]. The SDE states that if a problem is convex (as it is the case of each of the lower-level problems), the objective functions of the primal and dual problems have the same value at the optimum:

$$\begin{aligned}\sum_{g\in\Omega^{\mathrm{G}}} c^{\mathrm{G}}_{g,t} P^{\mathrm{G}}_{g,c,t}(\omega) = {} & \sum_{n\in\Omega^{\mathrm{N}}} \lambda_{n,c,t}(\omega) \left[\sum_{d\in\Lambda^{\mathrm{D}}_n} P^{\mathrm{D}}_{d,t}(\omega) k^{\mathrm{D}}_{d,c} - P^{\mathrm{W}}_{n,c,t}(\omega) \right] \\ & - \sum_{l\in\Omega^{\mathrm{L}}} \left[\phi^{\max}_{l,c,t}(\omega) + \phi^{\min}_{l,c,t}(\omega) \right] P^{\mathrm{L,max}}_{l} - \sum_{g\in\Omega^{\mathrm{G}}} \varphi^{\max}_{g,c,t} P^{\mathrm{G,max}}_{g} \\ & - \sum_{n\in\Omega^{\mathrm{N}}\setminus n\text{: ref.}} \left[\xi^{\max}_{n,c,t}(\omega) + \xi^{\min}_{n,c,t}(\omega) \right] \pi, \quad \forall c, \forall t, \forall \omega. \end{aligned} \tag{14.6}$$

Since $\vartheta_{n,c,t}(\omega) = \lambda_{n,c,t}(\omega), \forall n, \forall c, \forall t, \forall \omega$, as stated in constraints (14.1e), (14.6) allows reformulating each term $\sum_{n\in\Omega^{\mathrm{N}}} \vartheta_{n,c,t}(\omega) P^{\mathrm{W}}_{n,c,t}(\omega)$ as a function of exclusively linear terms:

$$\begin{aligned}\sum_{n\in\Omega^{\mathrm{N}}} \vartheta_{n,c,t}(\omega) P^{\mathrm{W}}_{n,c,t}(\omega) = \sum_{n\in\Omega^{\mathrm{N}}} \lambda_{n,c,t}(\omega) \sum_{d\in\Lambda^{\mathrm{D}}_n} P^{\mathrm{D}}_{d,t}(\omega) k^{\mathrm{D}}_{d,c} \\ - \sum_{l\in\Omega^{\mathrm{L}}} \left[\phi^{\max}_{l,c,t}(\omega) + \phi^{\min}_{l,c,t}(\omega)\right] P^{\mathrm{L,max}}_{l} - \sum_{g\in\Omega^{\mathrm{G}}} \varphi^{\max}_{g,c,t} P^{\mathrm{G,max}}_{g} \\ - \sum_{n\in\Omega^{\mathrm{N}}\setminus n:\,\mathrm{ref.}} \left[\xi^{\max}_{n,c,t}(\omega) + \xi^{\min}_{n,c,t}(\omega)\right] \pi - \sum_{g\in\Omega^{\mathrm{G}}} c^{\mathrm{G}}_{g,t} P^{\mathrm{G}}_{g,c,t}(\omega), \quad \forall c, \forall t, \forall \omega.\end{aligned} \tag{14.7}$$

On the other hand, the complementarity constraints of the KKT optimality conditions (14.5h)–(14.5m) have the form $0 \le e \perp h \ge 0$. These terms can be replaced by the following exact equivalent linear expressions based on the Fortuny-Amat transformation [26]:

$$e \le Mu \tag{14.8a}$$
$$h \le M(1-u) \tag{14.8b}$$
$$e, h \ge 0 \tag{14.8c}$$
$$u \in \{0,1\}, \tag{14.8d}$$

where M is a sufficiently large positive constant [26].

Using (14.7) and (14.8), the risk-constrained multistage wind power investment problem can be finally formulated as the MILP problem below:

$$\underset{\Omega^{\mathrm{U}}\cup\Omega^{\mathrm{L}}_{c,t}(\omega)\cup\Omega^{\mathrm{A}}_{c,t}(\omega)}{\text{Maximize}}$$

$$\begin{aligned}\sum_{\omega\in\Delta^{\omega}} \gamma(\omega) \Bigg\{ \sum_{t\in\Delta^{\mathrm{T}}} \Bigg[\sum_{c\in\Delta^{\mathrm{C}}} N^{\mathrm{H}}_{c} \Bigg(\sum_{n\in\Omega^{\mathrm{N}}} \lambda_{n,c,t}(\omega) \sum_{d\in\Lambda^{\mathrm{D}}_n} P^{\mathrm{D}}_{d,t}(\omega) k^{\mathrm{D}}_{d,c} \\ - \sum_{l\in\Omega^{\mathrm{L}}} \left(\phi^{\max}_{l,c,t}(\omega) + \phi^{\min}_{l,c,t}(\omega)\right) P^{\mathrm{L,max}}_{l} - \sum_{g\in\Omega^{\mathrm{G}}} \varphi^{\max}_{g,c,t} P^{\mathrm{G,max}}_{g} \\ - \sum_{n\in\Omega^{\mathrm{N}}\setminus n:\,\mathrm{ref.}} \left(\xi^{\max}_{n,c,t}(\omega) + \xi^{\min}_{n,c,t}(\omega)\right) \pi - \sum_{g\in\Omega^{\mathrm{G}}} c^{\mathrm{G}}_{g,t} P^{\mathrm{G}}_{g,c,t}(\omega) \Bigg) \\ - a_t \sum_{n\in\Delta^{\mathrm{N}}} c^{\mathrm{inv}}_{t}(\omega) X^{\mathrm{W}}_{n,t}(\omega) \Bigg] \Bigg\} + \beta \left(\zeta - \frac{1}{1-\alpha} \sum_{\omega\in\Delta^{\omega}} \gamma(\omega)\eta(\omega)\right)\end{aligned} \tag{14.9a}$$

subject to

$$\text{Constraints (14.1b)–(14.1f) and (14.1h)} \tag{14.9b}$$

$$\zeta - \sum_{t\in\Delta^{\mathrm{T}}}\Bigg[\sum_{c\in\Delta^{\mathrm{C}}} N_c^{\mathrm{H}}\Bigg(\sum_{n\in\Omega^{\mathrm{N}}}\lambda_{n,c,t}(\omega)\sum_{d\in\Lambda_n^{\mathrm{D}}} P_{d,t}^{\mathrm{D}}(\omega)k_{d,c}^{\mathrm{D}} - \sum_{l\in\Omega^{\mathrm{L}}}\left(\phi_{l,c,t}^{\max}(\omega)+\phi_{l,c,t}^{\min}(\omega)\right)P_l^{\mathrm{L,max}} - \sum_{g\in\Omega^{\mathrm{G}}}\varphi_{g,c,t}^{\max}P_g^{\mathrm{G,max}} - \sum_{n\in\Omega^{\mathrm{N}}\setminus n\text{: ref.}}\left(\xi_{n,c,t}^{\max}(\omega)+\xi_{n,c,t}^{\min}(\omega)\right)\pi - \sum_{g\in\Omega^{\mathrm{G}}} c_{g,t}^{\mathrm{G}}P_{g,c,t}^{\mathrm{G}}(\omega)\Bigg) - a_t\sum_{n\in\Delta^{\mathrm{N}}} c_t^{\mathrm{inv}}(\omega)X_{n,t}^{\mathrm{W}}(\omega)\Bigg] \le \eta(\omega), \quad \forall\omega \tag{14.9c}$$

$$\left\{\text{Constraints (14.2b)–(14.2g) and (14.5d)–(14.5g)}\right. \tag{14.9d}$$

$$\phi_{l,c,t}^{\max}(\omega) \le M^{\phi,\max}u_{l,c,t}^{\phi,\max}(\omega), \quad \forall l \tag{14.9e}$$

$$P_l^{\mathrm{L,max}} - f_{l,c,t}(\omega) \le M^{\phi,\max}\left(1-u_{l,c,t}^{\phi,\max}(\omega)\right), \quad \forall l \tag{14.9f}$$

$$\phi_{l,c,t}^{\min}(\omega) \le M^{\phi,\min}u_{l,c,t}^{\phi,\min}(\omega), \quad \forall l \tag{14.9g}$$

$$f_{l,c,t}(\omega) + P_l^{\mathrm{L,max}} \le M^{\phi,\min}\left(1-u_{l,c,t}^{\phi,\min}(\omega)\right), \quad \forall l \tag{14.9h}$$

$$\varphi_{g,c,t}^{\max}(\omega) \le M^{\varphi,\max}u_{g,c,t}^{\varphi,\max}(\omega), \quad \forall g \tag{14.9i}$$

$$P_g^{\mathrm{G,max}} - P_{g,c,t}^{\mathrm{G}}(\omega) \le M^{\varphi,\max}\left(1-u_{g,c,t}^{\varphi,\max}(\omega)\right), \quad \forall g \tag{14.9j}$$

$$\varphi_{g,c,t}^{\min}(\omega) \le M^{\varphi,\max}u_{g,c,t}^{\varphi,\min}(\omega), \quad \forall g \tag{14.9k}$$

$$P_{g,c,t}^{\mathrm{G}}(\omega) \le M^{\varphi,\min}\left(1-u_{g,c,t}^{\varphi,\min}(\omega)\right), \quad \forall g \tag{14.9l}$$

$$\xi_{n,c,t}^{\max}(\omega) \le M^{\xi,\max}u_{n,c,t}^{\xi,\max}(\omega), \quad \forall n\setminus n\text{: ref.} \tag{14.9m}$$

$$\pi - \delta_{n,c,t}(\omega) \le M^{\xi,\max}\left(1-u_{n,c,t}^{\xi,\max}(\omega)\right), \quad \forall n\setminus n\text{: ref.} \tag{14.9n}$$

$$\xi_{n,c,t}^{\min}(\omega) \le M^{\xi,\min}u_{n,c,t}^{\xi,\min}(\omega), \quad \forall n\setminus n\text{: ref.} \tag{14.9o}$$

$$\delta_{n,c,t}(\omega) + \pi \le M^{\xi,\min}\left(1-u_{n,c,t}^{\xi,\min}(\omega)\right), \quad \forall n\setminus n\text{: ref.} \tag{14.9p}$$

$$\phi_{l,c,t}^{\max}(\omega), \phi_{l,c,t}^{\min}(\omega) \ge 0, \quad \forall l \tag{14.9q}$$

$$\varphi_{g,c,t}^{\max}(\omega), \varphi_{g,c,t}^{\min}(\omega) \ge 0, \quad \forall g \tag{14.9r}$$

$$\xi_{n,c,t}^{\max}(\omega), \xi_{n,c,t}^{\min}(\omega) \ge 0, \quad \forall n\setminus n\text{: ref.} \tag{14.9s}$$

$$u_{l,c,t}^{\phi,\max}(\omega), u_{l,c,t}^{\phi,\min}(\omega) \in \{0,1\}, \quad \forall l \tag{14.9t}$$

$$u_{g,c,t}^{\varphi,\max}(\omega), u_{g,c,t}^{\varphi,\min}(\omega) \in \{0,1\}, \quad \forall g \tag{14.9u}$$

$$u_{n,c,t}^{\xi,\max}(\omega), u_{n,c,t}^{\xi,\min}(\omega) \in \{0,1\}, \quad \forall n \setminus n\text{: ref.} \tag{14.9v}$$

$$\Big\}, \forall c, \forall t, \forall \omega,$$

where $M^{\phi,\max}$, $M^{\phi,\min}$, $M^{\varphi,\max}$, $M^{\varphi,\min}$, $M^{\xi,\max}$, and $M^{\xi,\min}$ are large enough constants [26], and

$$\Omega_{c,t}^{\mathrm{A}}(\omega_s) = \Big\{ u_{l,c,t}^{\phi,\max}(\omega), u_{l,c,t}^{\phi,\min}(\omega), \forall l; u_{g,c,t}^{\varphi,\max}(\omega), u_{g,c,t}^{\varphi,\min}(\omega), \forall g; u_{n,c,t}^{\xi,\max}(\omega), u_{n,c,t}^{\xi,\min}(\omega), \forall n \setminus n\text{: ref.} \Big\}, \forall c, \forall t, \forall \omega, \tag{14.10}$$

are sets of auxiliary binary variables.

14.3.5 Illustrative Example

14.3.5.1 Data

MILP problem (14.9) is applied to the three-node system depicted in Fig. 14.5. This system comprises three nodes, one generation unit per node, one demand per node, and three transmission lines. Node 1 is the reference node.

Table 14.5 provides the generation unit and demand data. The first column gives the nodes at which generation units and demands are located. The second and third columns provide the capacity of each generation unit and the corresponding marginal cost, respectively. For the sake of simplicity, both the capacities and marginal costs are considered fixed throughout the planning horizon. Finally, the fourth column provides the peak demand at each node of the system prior to the beginning of the planning horizon. These peak demands multiplied by the corresponding demand factors give the demands for each operating condition, time period, and scenario.

Table 14.5 Generation unit and demand data

Node	Generation units		Peak demand [MW]
	$P_g^{\mathrm{G,max}}$ [MW]	c_g^{G} [$/MWh]	
1	150	76	150
2	150	58	120
3	120	65	120

All transmission lines are considered to have identical parameters with a reactance equal to 0.2 p.u. and a transmission capacity of 100 MW.

Wind power capacity can be built at nodes 2 and 3 up to 300 MW at each node throughout the planning horizon. Investment costs at the beginning of the planning horizon are $1,000,000 per MW. The investment budget is considered unlimited.

The planning horizon comprises two time periods of 6 years each. The demand and wind power capacity factors throughout the reference year of each period are represented using the ten clusters whose data are provided in Table 14.1. The amortization rates are considered equal to 0.26 and 0.13 in the first and second periods, respectively.

Regarding the uncertainty in the demand growth and investment costs, we consider the three cases below:

1. Uncertainty only affects demand growth. This case corresponds to the scenario data provided in Fig. 14.2 and Table 14.2.
2. Uncertainty only affects investment cost. This case corresponds to the scenario data provided in Fig. 14.3 and Table 14.3.
3. Uncertainty affects both demand growth and investment cost. This case corresponds to the scenario data provided in Fig. 14.4 and Table 14.4.

The subsections below provide the results for the three above cases. MILP problem (14.9) is solved for two different values of parameter β that realizes the trade-off between expected profit and risk of profit volatility:

1. $\beta = 0$: this case corresponds to a risk-neutral investor. This investor aims at maximizing its expected profit regardless of its profit volatility.
2. $\beta = 10$: this case corresponds to a risk-averse investor. This investor prefers to reduce its profit volatility despite the subsequent decrease in its expected profit.

In all cases a confidence level $\alpha = 0.95$ is considered.

Problem (14.9) is solved using CPLEX 12.2.0.1 [27] under GAMS [28] on a Linux-based server with four processors clocking at 2.9 GHz and 250 GB of RAM.

14.3.5.2 Results: Uncertainty in Demand Growth

Results corresponding to the scenario data of Fig. 14.2 and Table 14.2 (uncertainty in demands growth) and for risk-neutral and risk-averse investors are provided in Table 14.6. The first column provides the value of weighting parameter β. The second column indicates the scenarios. The third/fourth and fifth/sixth columns give the wind power capacity to be built at the beginning of the first/second period at nodes 2 and 3, respectively.

Note that in this case there are two demand growth scenario realizations in the first period and two scenario realizations in the second period depending on the scenario realization in the first one. This results in a single investment decision at the beginning of the planning horizon, which does not depend on any scenario realization (i.e., it is a *here-and-know* investment decision) and two alternative investment

Table 14.6 Results: uncertainty in demand growth

Case	Scenario	Wind power capacity to be installed [MW]			
		Node 2		Node 3	
		Period 1	Period 2	Period 1	Period 2
Risk-neutral ($\beta = 0$)	D1	0	0	300	0
	D2				
	D3		0		0
	D4				
Risk-averse ($\beta = 10$)	D1	0	0	279.7	20.3
	D2				
	D3		0		0
	D4				

decisions for the second period, one for scenarios D1 and D2 and other one for scenarios D3 and D4 (i.e., they are *wait-and-see* investment decisions with respect to the first period and *here-and-know* with respect to the second one).

Node 3 located in the south zone has better wind power conditions than node 2, located in the north zone. Thus, the wind power investor prefers to build wind power capacity at node 3 and does not build any wind power capacity at node 2 in any scenario and time period.

Regarding the differences between the investment decisions of a risk-neutral and a risk-averse investor, note that the first one builds 300 MW at the beginning of the planning horizon (which is the maximum wind power capacity that can be installed at each node for the whole planning horizon) while the risk-averse investor prefers to build a smaller wind power capacity at the beginning of the first period (279.7 MW) and wait until the beginning of the second period to decide on further investment: if the demand in the first period has increased (i.e., scenarios D1 and D2), it builds 20.3 MW that complete the 300 MW capacity; however, if the demand in the first period has decreased (i.e., scenarios D3 and D4), the wind power investor builds no additional capacity. This way, the wind power investor reduces its profit volatility.

14.3.5.3 Results: Uncertainty in Investment Cost

Results corresponding to the scenario data of Fig. 14.3 and Table 14.3 (uncertainty in investment cost) and for risk-neutral and risk-averse investors are provided in Table 14.7. The first column provides the value of weighting parameter β. The second column indicates the different scenarios. The third/fourth and fifth/sixth columns give the wind power capacity to be built at the beginning of the first/second period at nodes 2 and 3, respectively.

Table 14.7 Results: uncertainty in investment cost

Case	Scenario	Wind power capacity to be installed [MW]			
		Node 2		Node 3	
		Period 1	Period 2	Period 1	Period 2
Risk-neutral ($\beta = 0$)	IC1	0	0	267.6	0
	IC2		0		0
	IC3		109		32.4
Risk-averse ($\beta = 10$)	IC1	0	0	267.6	0
	IC2		0		0
	IC3		109		32.4

In this case, there is a single investment decision at the beginning of the first period and three different investment decisions at the beginning of the second period depending on the investment cost in the second period with respect to that in the first one: higher than (scenario IC1), equal to (scenario IC2), or lower than (scenario IC3).

The optimal solution consists of installing 267.6 MW at node 3 at the beginning of the planning horizon, and then, if the investment cost in the second period has decreased (i.e., scenario IC3), the wind power investor decides to install 109 and 32.4 MW of additional wind power capacity at nodes 2 and 3, respectively. As in the case of uncertainty in demand growth, the wind power investor prefers to install wind power capacity at node 3, which has better wind power conditions than node 2.

In this case, there are no differences between the optimal solutions of the risk-neutral and the risk-averse investors. This is so because at the time the wind power investor makes its investment decisions, it knows the actual investment costs.

14.3.5.4 Results: Uncertainty in Both Demand Growth and Investment Cost

Results corresponding to the scenario data of Fig. 14.4 and Table 14.4 (uncertainty in both demand growth and investment cost) and for risk-neutral and risk-averse investors are provided in Table 14.8. The first column provides the value of the weighting parameter β. The second column indicates the different scenarios. The third/fourth and fifth/sixth columns give the wind power capacity to be built at the beginning of the first/second period at nodes 2 and 3, respectively.

This case is a combination of the two previous cases since it considers all possible scenario combinations. The results show that, for the risk-neutral case, the wind power investor decides to build the whole 300 MW at node 3 at the beginning of the planning horizon and to build 132.9 and 63.7 MW at node 2 at the beginning of the second period in scenarios D1+IC3/D2+IC3 and D3+IC3/D4+IC3, respectively. That is, it only builds wind power capacity at node 2 if the investment cost in the

Table 14.8 Results: uncertainty in both demand growth and investment cost

Case	Scenario	Wind power capacity to be installed [MW]			
		Node 2		Node 3	
		Period 1	Period 2	Period 1	Period 2
Risk-neutral ($\beta = 0$)	D1+IC1	0	0	300	0
	D2+IC1				
	D1+IC2		0		0
	D2+IC2				
	D1+IC3		132.9		0
	D2+IC3				
	D3+IC1		0		0
	D4+IC1				
	D3+IC2		0		0
	D4+IC2				
	D3+IC3		63.7		0
	D4+IC3				
Risk-averse ($\beta = 10$)	D1+IC1	0	0	279.7	20.3
	D2+IC1				
	D1+IC2		0		20.3
	D2+IC2				
	D1+IC3		132.9		20.3
	D2+IC3				
	D3+IC1		0		0
	D4+IC1				
	D3+IC2		0		0
	D4+IC2				
	D3+IC3		63.7		20.3
	D4+IC3				

second period has decreased with respect to that in period 1. Moreover, the investor builds higher wind power capacity if the demand in the first period has increased (i.e., scenarios D1 and D2) than if the demand in the first period has decreased (i.e., scenarios D3 and D4).

Regarding risk-averse results, the wind power investor prefers in this case to build lower wind power capacity at node 3 at the beginning of the planning horizon than in the risk-neutral case and to wait until it knows the scenario realization in period 1.

Depending on this scenario realization it decides whether or not to build additional capacity.

Finally, Fig. 14.6 depicts the efficient frontier in this case. The efficient frontier shows how the expected profit decreases as the CVaR increases, as a consequence of changes in the weighting parameter β. Note that as parameter β increases, the wind power investor reduces its profit. However, it also reduces its profit volatility, i.e., it increases its CVaR.

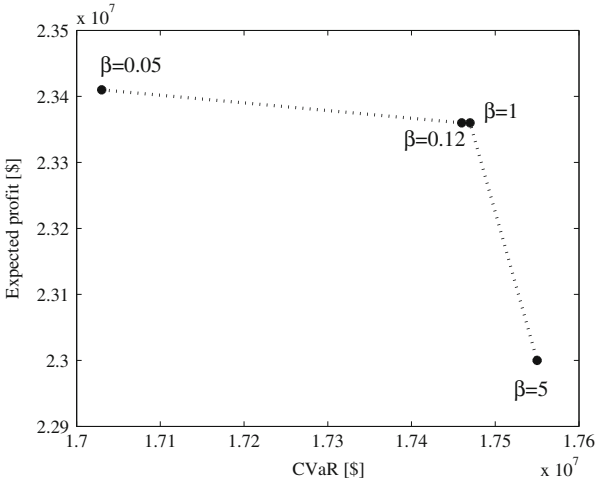

Fig. 14.6 Efficient frontier for the case of uncertainty in both demand growth and investment cost

14.3.5.5 Computational Issues

The computational time required for solving MILP problem (14.9) for the case studies analyzed in the previous subsections is less than 5 s in all cases. However, if problem (14.9) is considered for systems of realistic size and a large number of scenarios are considered as well, it is expected that the computational time drastically increases due mainly to the number of binary variables required to linearize the complementarity constraints.

Nevertheless, note that if investment decisions $X_{n,t}(\omega)$, $\forall n$, $\forall t$, $\forall \omega$, and auxiliary variable ζ are fixed to given values, MILP problem (14.9) decomposes by scenario ω, and Benders' decomposition can be applied [29] to reduce the computational burden.

14.4 Summary and Conclusions

In this chapter we provide a risk-constrained multistage decision-making model to determine the optimal generation investment in stochastic electricity-production facilities. This model is a stochastic bi-level problem that incorporates the clearing of a pool-based electricity market in which the investor sells the production of the newly built capacity. This investor aims at maximizing its expected profit while minimizing its profit volatility.

The multistage approach allows making investment decisions in different points in time to adapt to eventual changes in market and investment cost conditions. In this sense, we consider a planning horizon comprising a number of time periods, each one spanning a specific number of years. Investment decisions, involving stochastic capacity to be built in different locations, can be made at the beginning of each of these time periods.

The risk-constrained approach allows controlling the risk of profit volatility. To do so, the CVaR metric is used to quantify the risk. A weighting parameter in the objective function of the problem allows modeling the trade-off between expected profit and profit volatility and allows obtaining different investment strategies for different risk levels.

The stochastic approach allows incorporating in the model different sources of uncertainty though appropriate scenario trees that represent different realizations of the uncertain parameters.

A study pertaining to wind power investment is presented. In this study, three sources of uncertainty affect the investment decisions: the demand growth, the wind power production, and the investment cost.

Finally, from the theoretical modeling and the study carried out, the conclusions below are in order:

1. A risk-constrained framework is an appropriate approach for investment decision-making since it allows controlling the risk of profit volatility and generally prevents the investor to incur losses.
2. A multistage approach is necessary since demand growth and investment cost are subject to variations in the future and having the possibility of investing in different points in time is advantageous.
3. The resulting bi-level model can be formulated as an MILP problem, which can be solved using available branch-and-cut techniques.
4. The model is computationally tractable provided that the size of the system under study and the number of scenarios are moderate. For large systems and a large number of scenarios, decomposition techniques can be applied to reduce the computational burden.

References

1. Murphy FH, Smeers Y (2005) Generation capacity expansion in imperfectly competitive restructure electricity markets. Oper Res 53(4):646–661
2. Kazempour J, Conejo AJ, Ruiz C (2011) Strategic generation investment using a complementarity approach. IEEE Trans Power Syst 26(2):940–948
3. Wang J, Shahidehpour M, Li Z, Botterud A (2009) Strategic generation capacity expansion planning with incomplete information. IEEE Trans Power Syst 24(2):1002–1010
4. Botterud A, Ilic MD, Wangensteen I (2005) Optimal investment in power generation under centralized and decentralized decision making. IEEE Trans Power Syst 20(1):254–263
5. Oliveira F (2008) The value of information in electricity investment games. Energy Pol 36(7):2364–2375
6. Kamyaz P, Valenzuela J, Park CS (2007) Transmission congestion and competition on power generation expansion. IEEE Trans Power Syst 28(1):401–411
7. Chuang AS, Wu F, Varaiya P (2001) A game-theoretic model for generation expansion planning: problem formulation and numerical comparisons. IEEE Trans Power Syst 16(4):885–891
8. Baringo L, Conejo AJ (2011) Wind power investment within a market environment. App Energy 88(9):3239–3247
9. Baringo L, Conejo AJ (2013) Risk-constrained multi-stage wind power investment. IEEE Trans Power Syst 28(1):401–411
10. Birge JR, Louveaux F (1997) Introduction to stochastic programming. Springer, New York
11. Conejo AJ, Carrión M, Morales JM (2010) Decision making under uncertainty in electricity markets. Springer, New York
12. Rockafellar RT, Uryasev S (2000) Optimization of conditional value-at-risk. J Risk 2:21–41
13. Rockafellar RT, Uryasev S (2002) Conditional value-at-risk for general loss distributions. J. Bank Financ 26:1443–1471
14. Hoppner F, Klawonn F, Kruse R, Rumkler T. (1999) Fuzzy cluster analysis. Wiley, Chichester
15. Baringo L, Conejo AJ (2013) Correlated wind-power production and electric load scenarios for investment decisions. App Energy 101:475–482
16. Hyndman RJ, Shu Fan (2010) Density forecasting for long-term peak electricity demand. IEEE Trans Power Syst 25(2):1142–1153
17. Wiesenthal T, Mercier A, Schade B, Petric H, Szabó (2010) Quantitative assessment of the impact of the strategic energy technology plan on the European power sector. JCR Scientific and Technical Reports
18. Iberian Electricity Pool, OMEL, Spain and Portugal (2013) http://www.omel.es/. Accessed 12 Feb 2013
19. Espejo A, Minguez R, Tomas A, Menendez M, Mendez FJ, Losada IJ (2011) Directional calibrated wind and wave reanalysis databases using instrumental data for optimal design of off-shore wind farms. Paper presented in OCEANS, 2011 IEEE - Spain, pp 1–9, 6–9 June 2011

20. Menendez M, Tomas A, Camus P, Garcia-Diez M, Fita L, Fernandez J, Mendez FJ, Losada IJ (2011) A methodology to evaluate regional-scale offshore wind energy resources. Paper presented in OCEANS, 2011 IEEE - Spain, pp 1–8, 6–9 June 2011
21. Nordex (2013) http://www.nordex-online.com/. Accessed 12 Feb 2013
22. Eichhorn A, Heitsch H, Romisch (2010) Stochastic optimization of electricity portfolios: scenario tree modeling and risk management. In: Handbook of power systems I. Springer, New York, pp 405–432
23. Gómez-Expósito A, Conejo AJ, Cañizares C (2008) Electric energy systems: analysis and operation. Taylor and Francis, Boca Ratón
24. Luenberger DG (1973) Introduction to linear and nonlinear programming. Addison-Wesley publishing company, New York
25. Ruiz C, Conejo AJ (2009) Pool strategy of a producer with endogenous formation of locational marginal prices. IEEE Trans Power Syst 24(4):1855–1866
26. Fortuny-Amat J, McCarl B (1981) A representation and economic interpretation of a two-level programming problem. J Oper Res Soc 32(9):783–792
27. The ILOG CPLEX (2013) http://www.ilog.com/products/cplex/. Accessed 12 Feb 2013
28. Rosenthal RE (2012) GAMS, A user's guide. GAMS Development Corporation, Washington
29. Baringo L, Conejo AJ (2012) Wind power investment: a benders decomposition approach. IEEE Trans Power Syst 27(1):433–441

Part III
Pricing

Chapter 15
Pricing of Energy Contracts: From Replication Pricing to Swing Options

Raimund M. Kovacevic and Georg Ch. Pflug

Abstract The principle of replication or superhedging is widely used for valuating financial contracts, in particular, derivatives. In the special situation of energy markets, this principle is not quite appropriate and might lead to unrealistic high prices, when complete hedging is not possible, or to unrealistic low prices, when own production is involved. Therefore we compare it to further valuation strategies: acceptability pricing weakens the requirement of almost sure replication and indifference pricing accounts for the opportunity costs of producing for a considered contract. Finally, we describe a game-theoretic approach for valuating flexible contracts (swing options), which is based on bi-level optimization.

15.1 Introduction

This chapter deals with energy delivery contracts and their fair prices both from the seller's and the buyer's points of view. A contract between two parties determines the respective obligations of the two contracting sides to deliver or receive energy and to pay or receive money. Typically a contract is valid for a certain period of time and both the energy deliveries and the financial compensations are made at several moments in time. Some energy-related contracts even do not imply delivery of energy, but only financial transfers, which are however related to observable prices in the energy markets.

Suppose for simplicity that a contract states that payments and energy deliveries are due at times $t = 1, 2, \ldots, T$. The payments (cash flows) are denoted by C_t (in currency units). If both parties transfer money to the other one at the same

R.M. Kovacevic
Department of Statistics and Operations Research (ISOR), University of Vienna, Vienna, Austria

G.C. Pflug (✉)
ISOR and IIASA, Laxenburg, Austria
e-mail: georg.pflug@univie.ac.at

R.M. Kovacevic et al. (eds.), *Handbook of Risk Management in Energy Production and Trading*, International Series in Operations Research & Management Science 199,
DOI 10.1007/978-1-4614-9035-7_15, © Springer Science+Business Media New York 2013

time, only the net amount is recorded. By convention, a positive value C_t indicates a money inflow to the contract seller. One contract may involve several forms of energy $j \in \mathscr{J} = \{0,\ldots,J\}$, such as electricity, oil, gas, and coal. To avoid unit conversions as much as possible amounts of energy sources are measured by their energy content (in MWh). Energy of type j delivered in period t (in the time span $(t,t+1]$) is denoted by $D_{t,j}$. We agree that positive amounts $D_{t,j}$ refer to energy inflows, while negative amounts refer to outflows. If the contract involves only one type of energy we will just write D_t.

Both quantities C_t and D_t may be unknown at the time of contracting and may depend on information which is only available at the respective time of settlement (e.g., actual market prices). However, the amounts must be *determinable* by this information. A clause like "The buyer pays 1,000, if there is no extraterrestrial life" is void, because the validity of the condition it is not determinable. Conditions which are determinable but not known at the time of contracting are modeled as random variables.

Pricing principles determine a reasonable price to be offered to the buyer. The basic pricing principle is known as *replication pricing*: the buyer will not accept the price for the contract, if the market offers an alternative possibility, for which the upfront payment is lower and the cash flows or commodity flows are not smaller than the ones contracted. Thus the maximal offered price can be determined by an optimization problem, which is called the replication problem.

The alternative trading/hedging/production strategy belonging to the replication price is called *replication strategy*, if it produces exactly the same cash or commodity flows as the contract under consideration. If it produces larger flows, it is called a *superreplication strategy*. These types of strategies are riskless for the seller: following this strategy, the seller can under no circumstances make losses since the (super)replication must hold with probability 1. This is however a very strong requirement. Quite often, replication strategies do not exist and the superreplication leads to unrealistically high prices, which no buyer would accept. If there are contracts for which replication strategies do not exist in the market, the market is called *incomplete*. Electricity markets are typically incomplete, since replication strategies must use contracts offered on the wholesale market and these few types of contracts are quite simple compared to the possible variety of demand patterns.

As an example, consider a contract for energy delivery of amounts given by Fig. 15.1, upper graph. The lower graph shows possible hedging contracts. Any nonnegative linear combination of them qualifies as a replication strategy, but none of them replicates the demand shown in the upper graph.

In incomplete markets, replication pricing is not appropriate: if the seller wants to conclude a contract, he/she has to accept a certain risk. The *acceptance pricing rule* accounts for that the acceptance price is the minimal upfront payment, which makes the risk of this contract acceptable (but typically not riskless) for the seller. To quantify the notion of acceptable risk, measures of risk are introduced. Again this rule leads to an optimization problem: the minimal upfront payment has to be found under the constraint that the risk value lies below a certain prespecified value.

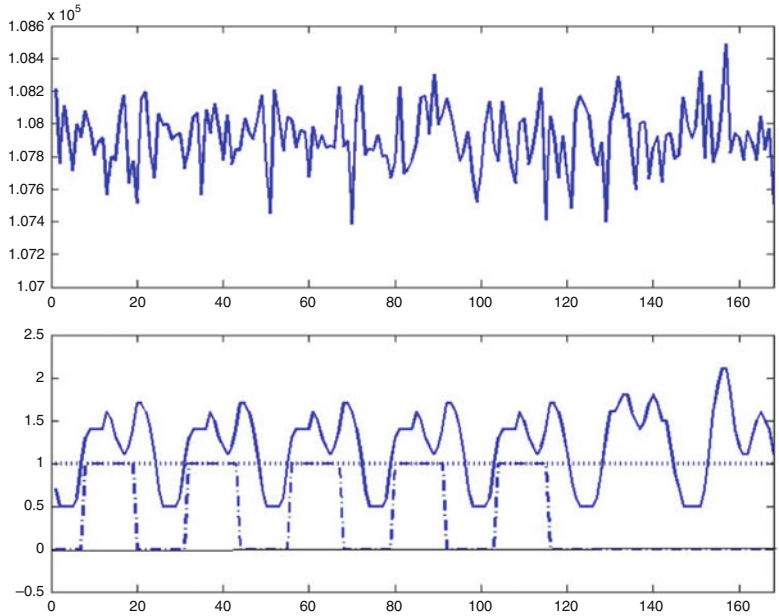

Fig. 15.1 *Above*: a demand profile of an energy buyer for the 168 h of a week. *Below*: the exchange market does only offer specific profiles, the base profile (*dotted*), the peak profile (*dash-dotted*), and the Vattenfall GH0 profile (*solid*)

The acceptance price takes the risk aversion of the seller into account, however does not depend on his/her actual risk exposure. A fine-tailored pricing instrument would take the risk portfolio of the seller into account and would make the contract acceptable only if the total risk exposure of the existing portfolio of contracts augmented by the new contract is acceptable. Notice that this pricing principle depends on the full knowledge of the existing portfolio of resources and contracts, which is not always available. Notice that the same contract may be acceptable for seller A but not acceptable for seller B. Consider for instance the situation when seller A has a lot of baseline energy available, but his contract portfolio is much biased versus peak demand. He would accept a contract which requires delivery in the night hours. On the other hand, if seller B has mostly solar energy to offer, then a contract which delivers at night time risks to require expensive purchases from the spot market and is not advantageous for seller B. The *indifference pricing rule* compares the risk of the existing resource and contract portfolio with the portfolio augmented with the new contract. The indifference price is the lowest price such that the risk of the augmented portfolio is not larger than the risk of the actual portfolio.

The three pricing principles (replication, acceptance, indifference) are applicable for *rigid contracts*, for which all conditions are fixed at contracting time 0. While amounts and prices may depend on future parameters and are considered as random variables at time 0, their way of calculation cannot be changed later by the contract parties. In contrast, *flexible contracts* allow the specification of demands by the

buyer at later times. Whenever a contract party has the right to exercise the contract in his discretion, this fact has to be taken into account in the price modeling process. While the ask price of rigid contracts are determined by a regular optimization problem (as described above), the pricing of flexible contracts requires the solution of a bi-level problem. A bi-level problem consists of two coupled optimization problems: an upper-level and a lower-level problem. The lower level describes the optimal reaction of the contract buyer to the price asked by the seller. The seller has to anticipate the reaction of the buyer when calculating the best ask price. Given this reaction, the seller may find the price according to one of the three principles: replication, acceptance, and indifference.

We summarize the mentioned approaches in the following overview:

(1) *(Super)replication* is based on the nonexistence of a better investment strategy for all scenarios.
(2) *Acceptance pricing* is based on the nonexistence of a better investment strategy with an acceptable risk. Superreplication is the special case if only zero risk is acceptable.
(3) *Indifference pricing* considers the actual risk exposure of the seller and accepts only if the additional contract does not increase the risk exposure. It requires to consider and model the full portfolio of all existing contracts and goes far beyond case (2) as there only the contract under consideration has to be considered.
(4) For the pricing of *flexible contracts* the anticipated behavior of the counterparty is taken into account when the price is calculated.

In principle, also the pricing for flexible contracts may be based on (1) replication, (2) acceptance, or (3) indifference. The superreplication principle is quite unrealistic since it is practically impossible to hedge the risk away simultaneously for all reactions of the buyer. We will concentrate on the acceptance principle and the related model structure for flexible contracts in this chapter. However, indifference pricing for flexible contracts can be easily introduced along the lines of the general indifference pricing approach, analyzed in Sect. 15.4.3.

The chapter is organized as follows: in Sect. 15.2 the concept of replication for financial contracts is presented. This concept is adapted to energy contracts in Sect. 15.3. Section 15.4 deals with the more general notion of acceptance pricing and the next Sect. 15.4.3 with the even more powerful notion of indifference pricing. Finally, bi-level acceptance pricing for electricity swing options is presented in Sect. 15.5.

15.2 Replication Pricing of Financial Contracts

As a starting point, we consider financial contracts, which generate for the contract holder a discrete-time sequence of random cash flows $(c_1, \ldots, c_T)$. Unlike for energy contracts, these cash flows are payments from the contract seller to the contract holder (think of the purchase of a share, which requires the initial payment by the buyer but gives him later the benefits of cash flows c_t as dividend payments).

Later, in the context of energy contracts, we set $C_t = -c_t$. We assume that the cash flows are—if necessary—already discounted to the present day by an appropriate discounting scheme [e.g., using a deterministic or stochastic interest rate process (R_t)]. We further assume that the cash-flow process (c_t) is adapted to a filtration $\mathfrak{F} = (\mathscr{F}_0, \mathscr{F}_1, \ldots, \mathscr{F}_{T-1})$, which models the information available at the respective time t.

At time zero money flows only from the buyer to the seller, as the price of the contract is payable at the beginning. From the contract buyer's side, the cash-flow structure is the same, but the signs are opposite. The main questions in contract pricing are what is the maximal price, which is acceptable for the buyer, and what is the minimal price which is acceptable for the seller and are these prices the same?

To answer these questions, alternative investments have to be taken into account: suppose that $m+1$ investment possibilities are given by a stochastic column price vector $S_t = (S_{t,0}, \ldots, S_{t,m})^\top$ (where $S_{t,0}$ relates to the riskless investment), adapted to the filtration $\mathfrak{F}$. Typically $\mathfrak{F}$ will be modeled as the filtration generated by the price process S_t.

Within our setup a hedging strategy is a nonanticipative row vector process $x = (x_0, \ldots, x_{T-1})$ on $\mathbb{R}^{m+1}$, where $x_t = (x_{t,0}, \ldots, x_{t,m})$ denotes the holdings of the $m+1$ investment possibilities during the time interval $[t, t+1]$. Nonanticipativity means that also the decision process x is adapted to the filtration $\mathfrak{F}$. While we prefer intervals with length 1 for notational simplicity, it is easily possible to extend the notation to include periods with different lengths.

With initial capital w and a trading strategy x, let $Y_t^{w,x}$ be the wealth at time t resulting from this strategy. To be more precise, let $Y_{t-}^{w,x}$ be the wealth just before time t. At time t the portfolio may be restructured and $Y_t^{w,x}$ denotes the wealth just after these transactions are made.

15.2.1 The Upper Price

Given investment opportunities as above, a market price π for a contract is acceptable for the buyer only if there is no better investment for the same or a lower price, i.e., there is no initial payment w and trading strategy x, such that $w < \pi$ and $Y_{t-}^{w,x} - Y_t^{w,x} \geq c_t$ (the cash flows can be paid) for all t and $Y_T^{w,x} \geq 0$ (the terminal wealth is nonnegative).

The upper price π_u is the highest price a potential buyer is willing to pay for the contract with the given cash-flow structure c_t. It is given as the minimal value of the following optimization problem:

$$\left\| \begin{array}{l} \text{Minimize (in } x \text{ and } w\text{): } w \\ \text{subject to} \\ Y_0^{w,x} = w, \\ Y_{t-}^{w,x} - Y_t^{w,x} \geq c_t \qquad t = 1, \ldots, T, \\ Y_T^{w,x} \geq 0, \\ x_t \text{ is nonanticipative.} \end{array} \right. \tag{15.1}$$

Notice that in case that the minimal value in (15.1) is attained by a trading strategy x, then the contract seller may execute this strategy to completely hedge the risk away. He/she would take the price w to invest in such a way that the cash-flows c_t are covered by the earnings of the invested portfolio and the final financial position is nonnegative, i.e., at the end the seller is free of debts out of this contract. The optimal $x = (x_0, \ldots, x_{T-1})$ in (15.1) describes the superreplication strategy.

15.2.2 The Lower Price

The seller receives an initial amount w and has to pay the cash flows c_t at later times. The price π of the contract is acceptable for the seller only if there is no alternative strategy, which receives more at the beginning and has lower liabilities later, i.e., there is no initial liability w and a strategy x, such that $w > \pi$ and $Y_{t-}^{w,x} - Y_t^{w,x} \leq c_t$ a.s., where $Y_t^{w,x}$ denotes now the liability process. At the end of the trading period, the liabilities $Y_T^{w,x}$ must be nonpositive.[1] That is, the lower price π_ℓ of this contract is the maximal value of the following optimization problem for liabilities $Y_t^{w,x}$:

$$\left\| \begin{array}{l} \text{Maximize (in } x \text{ and } w\text{): } w \\ \text{subject to} \\ Y_0^{w,x} \geq w, \\ Y_{t-}^{w,x} - Y_t^{w,x} \leq c_t \qquad t = 1, \ldots, T-1, \\ Y_T^{w,x} \leq 0, \\ x_t \text{ is nonanticipative.} \end{array} \right. \tag{15.2}$$

We call a strategy x which is feasible for this problem a *subreplication strategy*. All prices greater than π_ℓ are in principle acceptable for the seller, although he/she would prefer to get the upper price π_u.

15.2.3 The Linear Setup

In the simplest case with proportional transaction costs and volume-independent prices, the determination of the upper and lower prices amounts to solving a linear (stochastic) program.

If the transaction costs are neglected, the upper price π_u can be calculated by the following linear program[2]:

[1] Negative liabilities are profits.

[2] We denote by $x \cdot S$ the inner product of the vectors x and S.

$$\left\| \begin{array}{l} \text{Minimize (in } x \text{ and } w\text{): } w \\ \text{subject to} \\ x_0 S_0 - w \le 0, \\ x_{t-1} S_t \ge x_t S_t + c_t \qquad t = 1, \ldots, T, \\ x_{T-1} S_T \ge 0, \\ x_t \text{ is nonanticipative.} \end{array} \right. \tag{15.3}$$

In similar manner it is possible to formulate the lower price problem as a linear program:

$$\left\| \begin{array}{l} \text{Maximize (in } x \text{ and } w\text{): } w \\ \text{subject to} \\ x_0 S_0 - w \ge 0, \\ x_{t-1} S_t \le x_t S_t + c_t \qquad t = 1, \ldots, T, \\ x_T S_T \le 0, \\ x_t \text{ is nonanticipative.} \end{array} \right. \tag{15.4}$$

Dualization of these linear programs then entails the following well-known result: let $\tilde{c}_t = c_t / S_{t,0}$ and $\tilde{S}_t = S_t / S_{t,0}$, where $S_{t,0}$ denotes the price of the riskless investment. Then

$$\pi_u = \max\{ \sum_{t=1}^{T} \mathbb{E}_Q(\tilde{c}_t) : \ (\tilde{S}_t) \text{ is a martingale under } Q \ \}, \tag{15.5}$$

$$\pi_\ell = \min\{ \sum_{t=1}^{T} \mathbb{E}_Q(\tilde{c}_t) : \ (\tilde{S}_t) \text{ is a martingale under } Q \ \}. \tag{15.6}$$

Here $\mathbb{E}_Q$ is the expectation w.r.t. the probability measure Q. The upper and the lower price are equal if there is a unique martingale measure Q. In this (rather exceptional) case Q is called *the risk neutral measure* and $\pi_u = \pi_\ell$ for all contracts. If there are several martingale measures, then the model is called an *incomplete market model*. In incomplete markets, typically $\pi_\ell < \pi_u$ and we speak of an *ask-bid interval*. If the price is within the interval $[\pi_\ell, \pi_u]$, then it is acceptable for both parties and none of them may have an arbitrage opportunity.[3] However, even in incomplete markets, there may be contracts for which the inequalities in (15.1) and (15.2) are satisfied as equalities, meaning that the cash-flow c_t can be completely replicated and the optimal superreplication and subreplication strategies coincide. In this case $\pi_\ell = \pi_u$ even in incomplete markets.

On the other hand, for unbounded processes S_t, it may happen that $\pi_u = \infty$ [infeasibility in (15.1)] and/or that $\pi_\ell = -\infty$ [infeasibility in (15.2)]. It is unrealistic to assume that the buyer will pay any arbitrary price if $\pi_u = \infty$. To the contrary, he/she will not conclude the contract at all. That is why we consider acceptance pricing as a realistic alternative.

[3] The ask-bid interval has to be clearly distinguished to the bid-ask spread (bid-price $<$ ask-price) appearing in stock exchanges, when no deal can be made.

Transaction costs lead to an increased ask-bid interval. Furthermore, non-proportional transaction costs result in nonlinear and/or integer programs to be solved for pricing. For a comprehensive overview of convex models see, e.g., [16]. If the hedging process is nonconvex, a duality gap may occur and the notion of martingale measure makes no sense.

15.3 Replication for Energy Contracts

So far we have considered the valuation of pure financial contracts, which leads to classical results of derivative pricing. In the context of energy risk management we have to extend the analysis to deal with energy-related commodities in the following. There is a fundamental difference between pricing in financial markets and pricing in energy markets: while in financial markets the set of feasible trading strategies is considered to be the same for the seller and for the buyer, it is different in energy markets. The seller has a much larger spectrum of possible actions; he/she may use his/her own energy production, buy energy futures, and trade on the wholesale markets; and the buyer typically has no access to these possibilities, maybe for the exception of access to the spot markets. For this reason, the seller may determine the upper price by including all his/her assets in a replication model and may offer this price to the buyer.

15.3.1 Scope and Basic Model Setup

A key difference between financial and commodity derivatives results from the critical role of physical quantities and physical restrictions for the latter. The number of basic financial securities on which derivatives are written (i.e., the market capitalization, number of shares, etc.) is fixed. It is not necessary to produce securities and the market participants can hold arbitrary amounts without physical restrictions at negligible costs. It is also possible to go short in securities to a huge extent. Furthermore constraints on traded amounts and on the speed of trading are almost not existent.

Some energy-related contracts (usually futures) are settled financially and can therefore be viewed as financial contracts. However the picture is completely different for the physical commodities: they are produced and transported, and storage is costly and restricted. In particular negative storage is not possible. Commodities like electricity cannot even be stored. Therefore the usual relation between direct and future financial contracts is in general not valid for commodities and valuation concepts from finance cannot be applied to over-the-counter (OTC) contracts, which are settled physically.

Nevertheless, also for energy contracts we may seek for the smallest amount of cash necessary, at time zero to finance all feasible actions, in particular both physical and financial contractual cash flows. However, but the model has to be extended to capture all peculiarities of energy markets.

In the extended model it is possible to invest in different forms of energies: let $\mathscr{J} = \{0,\dots,J\}$ denote the set of available energy commodities, e.g., gas, heating oil, and electricity, measured by their energy content (MWh). The related spot prices are $S^e_{t,j} \in \mathbb{R}^{J+1}$, and $x^e_{t,j}$ denotes the stored amount of the jth commodities at time t. Throughout this paper the index $j = 0$ is reserved for (physical) electricity delivered at time t.

With minimum storage level zero and maximum levels $\bar{x^e} = (\bar{x}^e_0,\dots,\bar{x}^e_J) \geq 0$ we have to consider constraints

$$0 \leq x^e_t \leq \bar{x}^e. \tag{15.7}$$

For electricity no storage is available and we have $\bar{x}^e_{t,0} = 0$. We assume proportional storage costs ζ^e_j for each storage j.

Moreover, $y^e_t = (y^e_{t,0},\dots,y_{t,J})$ denotes the amounts of energy bought ($y^e_{t,j} \geq 0$) or sold ($y^e_{t,j} \leq 0$) at prices $S^e_{t,j}$ at time t, and matrices Z^e_t with elements $z^e_{t,ij}$ model the amount of energy i used to produce energy j during period $(t,t+1]$. Energy conversion leads to variable operating costs, which we assume to be proportional to input energy. The cost factors can be time dependent and are denoted by $\gamma_{t,ij}$ (currency unit per MWh).

Related to the conversion from energy i to energy j are efficiencies η_{ij}. In this way it is, e.g., possible to model electricity production from different fuels. Conversion between different forms of energy is restricted by lower and upper bounds $\underline{z}_t, \bar{z}_t$, i.e.,

$$0 \leq \underline{z}_{t,ij} \leq z^e_{t,ij} \leq \bar{z}_{t,ij}, \tag{15.8}$$

which reflects physical constraints on production. On the other hand, trading of energy is not restricted in the basic superhedging setup.

For certain energy sources (e.g., stored water in a certain reservoir) one may also consider random inflows, denoted by $d_{t,j} \geq 0$. Further inflows can result from intermittent electricity production by renewable sources like wind or solar power.

Note that in this framework it is also possible to model energy-related contracts with physical delivery, if the lower and upper bounds for conversion and the related conversion costs are modeled time dependent or depending on other variables of the system. As an example consider an electricity future j: conversion between electricity deliverable by the future contract and actual electricity happens during the delivery period. The related conversion factors are described by $\eta_{j0} = 1$. For the seller of the delivery contract we have during delivery $\bar{z}_t = \underline{z}_t = -x_{t,j}/n_t$, where n_t is the number of remaining exercise dates. Outside the delivery period we have $\bar{z}_t = \underline{z}_t = 0$. Finally, the delivery price is modeled by using the operating cost factors $\gamma_{t,j0}$.

In addition to forms of energy and physically settled contracts we use a cash position $x_{t,0}$ with an interest yield of $r_f > 0$ and include financial assets and contracts $\mathscr{I} = \{1,\dots,I\}$ with prices $S^f_{t,i}$, paying cash flows $C^f_{t,i}$ at time t. While typical financial assets are not in the focus for pure energy-related valuation problems, energy derivatives with financial settlement can be modeled in this way. As an example,

an electricity future contract i for fuel j with strike price K_i pays $S^e_{t,j} - K_i$ during delivery. Holdings of financial contracts are denoted by $x^f_{t,i}$ and are not restricted; in particular there are no shortselling constraints.

In the extended model we use physical and financial contracts to hedge an OTC contract that is defined by cash flows C_t and physical flows of energy B_t, with $B_{t,j}$ denoting the flow of energy j at time t in MWh. Inflows are reflected by positive and outflows by negative values. Both C_t and D_t can be random and can depend on other variables of the system.

For a simple delivery contract, D_t (delivery of one commodity) and C_t are constant during the whole time of delivery. A swap contract for different forms of energy i, j can be modeled by setting $D_{t,i} \geq 0$ and $D_{t,j} \leq 0$.

Table 15.1 gives an overview of the possible conversions within the proposed framework.

Table 15.1 Possible conversions, including cash

	1.	2.	3.	4.	5.	6.	7.
1. Cash		x	x		x	x	x
2. Energy commodities	x	x			x		
3. Phys. en. contracts	x	x					
4. (Stored water)					x		
5. Electricity	x			x			
6. Phys. el. contracts	x				x		
7. Financial contracts	x						

15.3.2 Formulation of the Extended Valuation Model

Using the outlined notation we can now formally describe the financial sellers' problem to energy markets. We do not consider transaction costs on energy related and financial markets. Hence it is possible to give an LP formulation.

Stored energy starts with an initial storage x^0. In the subsequent periods storages are changed by buying and selling energy, by conversion between energy forms and contractual deliveries of the physical contracts in the portfolio and of the OTC contract under consideration. For all energy contracts except electricity ($j \neq 0$) we formulate this as

$$x^e_{0,j} \leq x^0 + y^e_{0,j} + d_{0,j} \tag{15.9}$$

and

$$x^e_{t,j} \leq x^e_{t-1,j} + y^e_{t,j} + \sum_{i=0}^{J} \eta_{ij} z^e_{t-1,ij} - \sum_{i=1}^{J} z_{t-1,ji} + d_{t,j} + D_{t,j}. \tag{15.10}$$

Note that the index in the first sum begins at $i = 0$. This allows for pumping, if energy j refers to water, stored in a reservoir.

Optimization will lead to boundary solutions in (15.9) and (15.10). For electricity (which is not storable) we require

$$0 = y^e_{0,0} + d_{0,0} \tag{15.11}$$

$$0 = y^e_{t,0} + \sum_{i=1}^{J} \eta_{i0} z^e_{t-1,i0} + d_{t,0} + D_{t,0}. \tag{15.12}$$

Note that (15.9)–(15.12) refer to the point in time immediately before the next production period $[t, t+1)$ begins and recall that storage is also constrained by (15.8), which models physical restrictions as well as contractual limits for physical contracts.

In a model with discrete time, only energy stored at the beginning of a period can be used for conversion during the period. Therefore we introduce the following constraints:

$$\sum_{j=1}^{J} z^e_{t,ij} \le x^e_{t,i}. \tag{15.13}$$

The cash account is $x^f_{t,0}$. Cash is considered after buying and selling energy and settling all types of contract but before actually converting between energies. Hence with an initial amount w of cash just before the transactions at time zero to be effectuated, $x^f_{0,0}$ must fulfill

$$x^f_{0,0} \le w - \sum_{j=0}^{J} S^e_{0,j} y^e_{0,j} - \sum_{i=0}^{I} S^f_{0,i} x^f_{0,i}. \tag{15.14}$$

In the subsequent periods $t > 0$ the cash position accumulates gains and subtracts costs from buying and selling energy and financial contracts and the cash flows from the OTC contract under consideration. Furthermore, it has to account for interest on cash, costs for energy conversion, storage costs, and all cash flows from financial contracts. This results in

$$\begin{aligned} x^f_{t,0} \le\ & (1+r_f) x^f_{t-1,0} \\ & - \sum_{j=0}^{J} S^e_{t,j} y^e_{t,j} - \sum_{i=1}^{I} S^f_{t,i} (x^f_{t,i} - x^f_{t-1,i}) + \sum_{i=1}^{I} C^f_{t,i} + C_t \\ & - \sum_{i=0}^{J} \sum_{j=0}^{J} \gamma_{t,ij} z_{t,ij} \\ & - \sum_{j=1}^{J} \zeta_j \frac{(x^e_{t,j} + x^e_{t-1,j})}{2}. \end{aligned} \tag{15.15}$$

Note that we do not use a nonnegativity constraint on the cash position, that is, borrowing money is allowed.

Finally the inequality

$$x^f_{T,0} + \sum_{j=1}^{J} S^e_{T,j} x^e_{t,j} + \sum_{i=1}^{I} S^f_{T,i} x^f_{T,i} \geq 0 \tag{15.16}$$

ensures that we search for the smallest initial payment w and a related hedging strategy, such that the asset value—consisting of the final cash position $x^f_{T,0}$ and all physical and financial contracts—is nonnegative after handling the OTC contract under consideration.

Based on the previous considerations and using the hedging approach, the valuation problem can be formulated as the following optimization problem:

$$\left\| \begin{array}{l} \text{Minimize (in } x^e, x^f, y, z \text{ and } w\text{): } w \\ \text{subject to} \\ \text{constraints } (15.7)\text{–}(15.16), \\ x^e_t, x^f_t, y_t, z_t \text{ are nonanticipative.} \end{array} \right. \tag{15.17}$$

In principle, both the seller's and the buyer's hedging problem have the same form (15.17). However, the typical situation in energy markets, the buyer's set of possibilities in producing, trading, or hedging is usually restricted. Thus a natural asymmetry between the contracting partners occurs.

Another key difference lies in the fact that the streams C_t and D_t have different signs for the two participants, e.g., for a simple delivery contract for one commodity, the physical flow D_t is negative for the seller and positive for the buyer, while the opposite holds for the cash flow C_t. Hence even if the buyer would have access to all types of financial and commodity contracts, there is still another source of asymmetry: it is impossible to just change the signs of the flows of commodities to get the picture of the other contractor: usually efficiencies are not symmetric, i.e., $\eta_{ij} \neq \eta_{ji}$. Consider e.g., the production of electricity: for pumped turbines it is possible to use electricity for storing it in higher reservoirs, but with low efficiency, compared to the efficiency of producing electricity from stored water. As an extreme example of asymmetry it is not possible to produce fuel from electricity delivered by a contract, while fuel clearly can be used to produce electricity with some positive efficiency. Moreover, by the existence of bounds in production the problem does not scalarize.

15.4 Acceptability Replaces Non-replicability

Clearly, the principle of (super)replication is one of the cornerstones of modern finance. On the other hand it might be too strong under some circumstances. As was already said, it may lead to very large (even infinite) upper prices in incomplete markets.

This is especially true in the case of electricity markets: generating companies have the equipment to buy, store, and use fuel to generate electricity in order to satisfy even very complicated contractual terms. Physical constraints are present but only mildly affect the ability to hedge the flows related to the contract. On the other hand other market participants do not own the same equipment. Pure traders have access to electricity exchanges or pools, and hence to the full spectrum of financially settled contracts, but are not able to produce electricity. End consumers do not even have access to exchange markets. While contractual energy flows can be very complicated for OTC contracts, electricity is not storable and there are only very few instruments (i.e., contracts such as base and peak futures) available to partially hedge a specific contract. So even for a trader with access to an electricity exchange, hedging is difficult and will in general work only approximately—there will always remain residual electricity flows that have to be settled by buying spot electricity without protection from future contracts.

For these reasons we analyze the pricing problem by the notion of acceptability: it is wanted that the difference between the optimal hedge and the cash-flow process is acceptable for the seller, which—in the most basic formulation—means that inequality (15.16) is replaced by

$$\mathscr{A}(C_T + \sum_{j=1}^{J} S^e_{T,j} x^e_{t,j} + \sum_{i=1}^{I} S^f_{T,i} x^f_{T,i}) \geq 0, \tag{16'}$$

where $\mathscr{A}$ is an acceptability functional (see below). In this way, unfavorable scenarios are not avoided completely at the end. Instead, the loss distribution is restricted by the acceptability functional, such that only unfavorable outcomes with small probability are acceptable.

The resulting optimization problem for acceptability pricing is a modification of (15.17) and can be written as

$$\left\| \begin{array}{l} \text{Minimize (in } x^e, x^f, y, z \text{ and } w\text{): } w \\ \text{subject to} \\ \text{constraints (15.7)–(15.15)}, \\ \mathscr{A}(C_T + \sum_{j=1}^{J} S^e_{T,j} x^e_{t,j} + \sum_{i=1}^{I} S^f_{T,i} x^f_{T,i}) \geq 0, \\ x^e_t, x^f_t, y_t, z_t \text{ are nonanticipative.} \end{array} \right. \tag{18}$$

15.4.1 Acceptability Functionals

A *probability functional* is an extended real-valued function defined on some random space or on a suitable subset of a random space. Examples are well-known functionals like the expectation, the median, value-at-risk, average (or "conditional") value-at-risk, and variance. If the value of a probability functional depends

only on the distribution of the random variable under consideration, it is called *version independent*. If a functional is interpreted in the sense that higher values are preferable to lower values, we call it an *acceptability-type functional*.

Acceptability functionals are probability functionals $\mathscr{A}$, defined on a linear $\mathscr{Y}$ space of random variables on $(\Omega, \mathscr{F}, \mathbb{P})$, such that the following properties are true for all $X, Y \in \mathscr{Y}$:

(A1) Concavity. $\mathscr{A}(\lambda \cdot X + (1-\lambda)\cdot Y) \geq \lambda \cdot \mathscr{A}(X) + (1-\lambda)\cdot \mathscr{A}(Y)$ holds for $\lambda \in [0,1]$.
(A2) Monotonicity. $X \leq Y\ a.s. \Rightarrow \mathscr{A}(X) \leq \mathscr{A}(Y)$.

Often (see, e.g., [17]), acceptability functionals are defined by including the translation equivariance property:

(A3) Translation equivariance. $\mathscr{A}(Y+c) = \mathscr{A}(Y) + c$ holds for all constants c.

An acceptability functional is called *positively homogeneous*, if it satisfies the condition $\mathscr{A}(\lambda Y) = \lambda \cdot \mathscr{A}(Y)$ for all $\lambda \geq 0$. It is called strict, if $\mathscr{A}(Y) \leq \mathbb{E}(Y)$ holds. Recall that for an acceptability functional $\mathscr{A}$ and a random loss Y the valuation $-\mathscr{A}(-Y)$ is a coherent risk measure in the sense of [1].

Throughout this paper we will consider only acceptability functionals with $\mathscr{A}(0) = 0$. If necessary, this can be achieved easily by relocating the functional. Under rather mild conditions (upper semicontinuity) an acceptability functional $\mathscr{A}_t$ has a dual representation

$$\mathscr{A}(Y) = \inf\{\mathbb{E}(YZ) - \mathscr{A}^+(Z) : Z \in \mathscr{Y}^*\},$$

where Z is an element of the dual space $\mathscr{Y}^*$ and $\mathscr{A}^+$ is the concave conjugate (in the sense of Fenchel–Moreau–Rockafellar, see [20]) of the functional $\mathscr{A}$. If the functional is positively homogeneous, then the dual representation simplifies to

$$\mathscr{A}(Y) = \inf\{\mathbb{E}(YZ) : Z \in \mathscr{Z}\}, \tag{19}$$

where $\mathscr{Z}$ is a convex subset of $\mathscr{Y}^*$.

An important, but simple, example for a positively homogeneous acceptability functional is the *average value-at-risk*. For a random variable Y with distribution function G_Y it is defined by $\mathbb{AV@R}_\alpha(Y) = \frac{1}{\alpha}\int_0^\alpha G_Y^{-1}(u)du$ and is also known as conditional value-at-risk or tail value-at-risk. Its conjugate representation is given as follows [see [17], Theorem 2.34 (ii)]:

$$\mathbb{AV@R}_\alpha(Y) = \inf\left\{\mathbb{E}(Y\cdot Z) : \mathbb{E}(Z) = 1,\ 0 \leq Z \leq \frac{1}{\alpha}\right\}. \tag{20}$$

In our examples we will use the average *value-at-risk*, because it is closely related to the notion of value-at-risk, the most important risk measure in practice. The value-at-risk of a random variable X at confidence level $0 \leq \alpha \leq 1$ is basically defined as the value-at-risk of the related distribution (see, e.g., [14]), which is given by $\mathbb{V@R}_\alpha(Y) = \inf\{v : P\{Y \leq v\} \geq \alpha\}$.

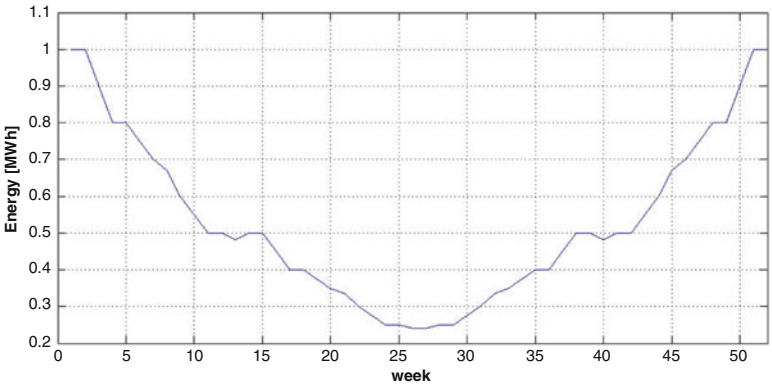

Fig. 15.2 Acceptability pricing: delivery pattern D_t over the 52 weeks of a year

Using the value-at-risk in (16′) is called quantile hedging/pricing; see, e.g., [8]. In this case $\mathbb{V}@\mathrm{R}_\alpha(Y) \geq q$ is equivalent to $P\{Y < q\} \leq \alpha$. $\mathbb{V}@\mathrm{R}$ is monotone, but unfortunately the mapping $Y \mapsto \mathbb{V}@\mathrm{R}_\alpha(Y)$ is not convex and nonsmooth. This makes the usage of constraint (16′) in an optimization problem very difficult.

On the other hand $\mathbb{A}\mathrm{V}@\mathrm{R}$ is an acceptability functional and is a concave minorant of the $\mathbb{V}@\mathrm{R}$. As an alternative to using quantiles or acceptability functionals, one may also consider utility functions U and accept a contract, if $\mathbb{E}[U(Y)] \geq q$. However, the price will then depend on the choice of the entire utility function while in quantile pricing only two parameters, the threshold q and the confidence level $1 - \alpha$, have to be set by the management.

The following example illustrates acceptability pricing, using the average value-at-risk.

Example 1. We consider a planning horizon of 1 year (52 weeks). Electricity spot prices are modeled by geometric Brownian motion with jumps (GBMJ), estimated from EEX Phelix hourly electricity prices (hourly, 09/2008–12/2011, Bloomberg). The pricing model was reformulated and solved on a stochastic tree, generated from the GBMJ model.

The hedging opportunities are represented by four future contracts, related to the quarters of the year, i.e., each of the futures delivers a constant amount of electric energy during one of the quarters. The delivery pattern of the contract to be valued is shown in Fig. 15.2.

Using problem (18) the acceptability price is calculated for a pure trader meaning that only wholesale base quarter future contracts can be used for hedging for different values of the $\mathbb{A}\mathrm{V}@\mathrm{R}$-parameter α, shown in Fig. 15.3. The related optimal hedging strategies can be seen in Fig. 15.4. Finally, Fig. 15.5 shows the density of the optimized distribution of profits and the value-at-risk at level $\alpha = 0.1$.

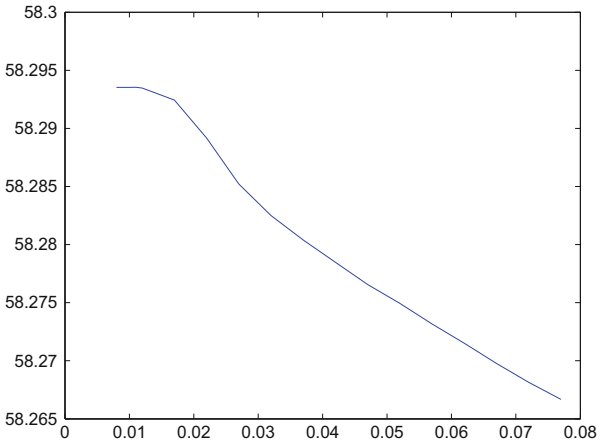

Fig. 15.3 Acceptability pricing: the price of 1 MWh as a function of the acceptance level α

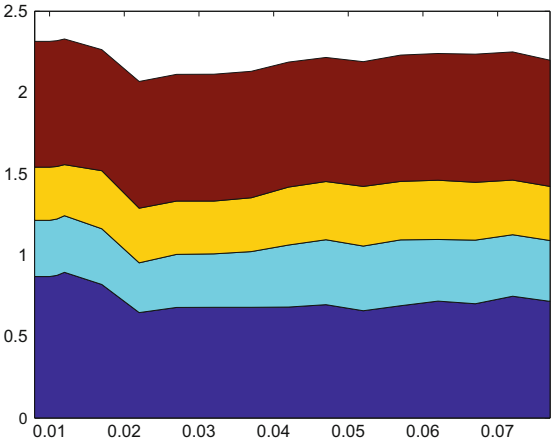

Fig. 15.4 Acceptability pricing: optimal hedges as a function of the acceptance level α

15.4.2 Acceptability Pricing for Financial Contracts

For purely financial contracts, the acceptability upper pricing problem can be investigated in more detail: it is the following variant of the replication problem (15.1). The upper price π_u is the minimal value of

$$\left\| \begin{array}{l} \text{Minimize (in } x \text{ and } w\text{): } w \\ \text{subject to} \\ Y_0^{w,x} = w, \\ Y_{t-}^{w,x} - Y_t^{w,x} \geq c_t \qquad t = 1, \ldots, T, \\ \mathscr{A}(Y_T^{w,x}) \geq 0, \\ x_t \text{ is nonanticipative,} \end{array} \right. \tag{21}$$

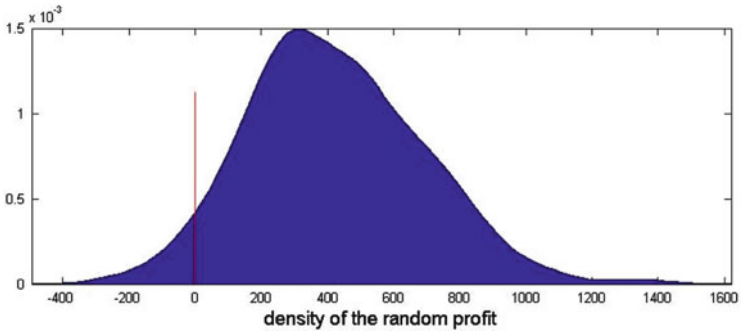

Fig. 15.5 Acceptability pricing: density of the profit variable

which takes the following form in the linear setup:

$$\left\| \begin{array}{l} \text{Minimize (in } x \text{ and } w\text{): } w \\ \text{subject to} \\ x_0 S_0 \le 0, \\ x_{t-1} S_t - x_t S_t - c_t \ge 0; t = 1, \ldots, T, \\ \mathscr{A}_T(x_T S_T) \ge 0. \end{array} \right. \tag{22}$$

If the functional $\mathscr{A}$ is given by representation (19), then problem (22) has a dual given by

$$\left\| \begin{array}{l} \text{Maximize (in } Z_t\text{) } \sum_{t=1}^{T} \mathbb{E}(c_t Z_t) \\ \text{subject to} \\ \mathbb{E}(S_{t+1} Z_{t+1} | \mathscr{F}_t) = Z_t S_t, \\ Z_t \ge 0; \quad t = 1, \ldots, T-1, \\ Z_T \in \mathscr{Z}. \end{array} \right. \tag{23}$$

The latter problem can be reformulated in an analogous way as in (15.5): Let $\tilde{c}_t = c_t / S_{t,0}$ and $\tilde{S}_t = S_t / S_{t,0}$, where $S_{t,0}$ is the riskless investment. Then the acceptability upper price π_u is given by

$$\max\{\sum_{t=1}^{T} \mathbb{E}_Q(\tilde{c}_t) : \ (\tilde{S}_t) \text{ is an (equivalent) martingale under } Q \text{ s. t. } \frac{dQ}{dP} \in \mathscr{Z}\}. \tag{24}$$

Similarly, the lower price π_ℓ is

$$\min\{\sum_{t=1}^{T} \mathbb{E}_Q(\tilde{c}_t) : \ (\tilde{S}_t) \text{ is an (equivalent) martingale under } Q \text{ s. t. } \frac{dQ}{dP} \in \mathscr{Z}\}. \tag{25}$$

Denote by $\pi_u(\mathscr{Z})$ the upper price in dependency of the considered acceptability functional $\mathscr{A}$ with dual set $\mathscr{Z}$. Notice that

$$\mathscr{Z}_1 \subseteq \mathscr{Z}_2 \quad \text{implies that} \quad \pi_u(\mathscr{Z}_1) \le \pi_u(\mathscr{Z}_2).$$

The largest price is gotten when full (super)replication is required, meaning that $\mathscr{Z}$ must be equal to all nonnegative random variables, compare (15.5). A smaller and more realistic price is obtained, if the acceptability functional is, e.g., the average value-at-risk $\mathbb{AV@R}_\alpha$, with $\mathscr{Z} = \{Z : 0 \le Z \le \frac{1}{\alpha}\}$. The smallest price is given by the choice $\mathscr{Z} = \{1\}$, which corresponds to the acceptability requirement $\mathbb{E}(Y_T^{w,x}) \ge 0$. This simple pricing rule is related to the concept of expected net present value (ENPV) of a contract and can be seen as the absolute minimum price for avoiding bankruptcy. However no seller will be willing to contract on this basis.

15.4.3 From Acceptability Pricing to Indifference Pricing

Acceptability pricing allows a meaningful valuation of contracts, even if full replication of the related flows is not possible or too expensive. However there is another difficulty remaining: the equipment of, e.g., a producer of electricity is never dedicated just to the production of the OTC contract under consideration. From the standpoint of production, delivering new contractual cash and energy flows is always an addendum to previously planned decisions. This means that the value of a contract should be valuated relative to the optimal management of all the other contracts which are already in the portfolio of the seller.

This idea leads to the notion of indifference pricing: the indifference principle states that the seller of a product compares his optimal decisions with and without the contract and then requests a price such that he is at least not worse off when closing the contract. This idea goes back to insurance mathematics (see [4]) but has been used for pricing a wide diversity of financial contracts in recent years, e.g., [5] for an overview.

In order to model the indifference price approach, assume that the total energy deliveries of the actual portfolio are D_t^{old} and the total cash flows out of this portfolio are C_t^{old}. These cash flows must include also the upfront payments at time 0. The additional contract, for which a price is not yet determined, is given by D_t respectively C_t. Indifference pricing happens in two steps:

- Determination of the acceptability of the actual portfolio. To this end, the following problem is solved:

$$\left\| \begin{array}{l} \text{Maximize (in } x,y,z \text{ and } w\text{): } \mathscr{A}(C_T^{old} + \sum_{j=1}^{J} S_{T,j}^e x_{t,j}^e + \sum_{i=1}^{I} S_{T,i}^f x_{T,i}^f) \\ \text{subject to} \\ \text{constraints } (15.7)\text{–}(15.16), \\ x_t, y_t, z_t \text{ are nonanticipative.} \end{array} \right. \tag{26}$$

  Here the equations are based on D_t^{old} respectively C_t^{old}. The optimal value of this optimization problem, that is, the acceptability level of the actual (old) portfolio, is denoted by a_0.

- Determination of the indifference price of the additional contract. Let the new total deliveries be $D_t^{new} = D_t^{old} + D_t$ and the new cash flows (without the upfront payment for the additional contract) be $C_t^{new} = C_t^{old} + C_t$. The price of the additional contract is denoted by $x_{0,0}$. It is determined by the following problem:

$$\left\| \begin{array}{l} \text{Minimize (in } x,y,z \text{ and } w\text{): } w \\ \text{subject to} \\ \mathscr{A}(C_T^{new} + \sum_{j=1}^{J} S_{T,j}^{e} x_{t,j}^{e} + \sum_{i=1}^{I} S_{T,i}^{f} x_{T,i}^{f}) \geq a_0 \\ \text{and constraints } (15.7)\text{–}(15.15), \\ x_t, y_t, z_t \text{ are nonanticipative.} \end{array} \right. \tag{27}$$

Of course, here the equations are based on D_t^{new} respectively C_t^{new}.

The following example compares indifference pricing with acceptability pricing for a simple setup with one thermal generation unit and a fixed delivery contract to be valuated.

Example 2. We consider an electricity producer, who has available a single combined cycle plant that is able to use both oil and gas. The machine has maximum power production of 410 MW and efficiencies of 0.575 (gas) and 0.57 (oil). Both fuels can be stored up to some amount ($1.5 \cdot 10^6$ MWh) at storage costs 0.2 Euro/MWh/h. We do not consider future contracts in this setup; hence hedging is possible only by buying fuel at appropriate points in time. Again we use electricity prices and weekly decision periods as described in Example 1. Gas prices are estimated (following [12]) by GBMJ from GPL spot prices (hourly, 04/2007–12/2011, Bloomberg) and oil spot prices for Brent Crude prices (daily, 05/2003–12/2011, Bloomberg).

We valuate a simple delivery contract, which binds the producer to supply a fixed amount of energy, the contract size in MWh, during each stage of the planning problem. The producer is free to buy and store fuel and to produce electric energy for the contract and also for selling it at the spot market. The value of the contract per MWh contains variable operating costs. From this we calculate a contract value per MWh that also includes an amount of coverage for fixed cost, which is proportional to the mean workload of the production unit during the planning horizon.

Within this setup we compare the superhedging approach to indifference pricing. Superhedging is possible for a producer, if the contract size does not exceed the capacity of the combined cycle turbine. For indifference pricing, we use the average value-at-risk at level $\alpha = 0.05$ as the related acceptability functional.

See Fig. 15.6 for the main results. As described above, superhedging leads to contractual deliveries, but does not account for alternative usages of the machine, whereas indifference pricing does. This is the reason why the superhedging price might be considered as too low in this case. Superhedging and indifference pricing also show different amounts of fixed cost, because the related strategies use non-contractual electricity production at different levels.

15.5 Flexible Contracts: Swing Option Pricing

Electricity *swing options* give their buyers the right to obtain electricity at a fixed price K per MWh during some delivery period. The price is set by the seller at contract formation, while the actual purchase quantities can be chosen (within some contractual range) by the buyer during the whole delivery period. Swing options are also known as *flexible nomination contracts*, *take-or-pay contracts*, or *virtual power plants*. See, e.g., [2, 11, 18, 19].

So far, we considered delivery contracts with delivered quantities that were either fixed in advance or depending on some observable (possibly stochastic) variables. Swing options are different, because the delivered quantities are decisions of the option buyer, and the seller has to account for this fact, when making the pricing decision. Hence, two questions are important, when considering swing options:

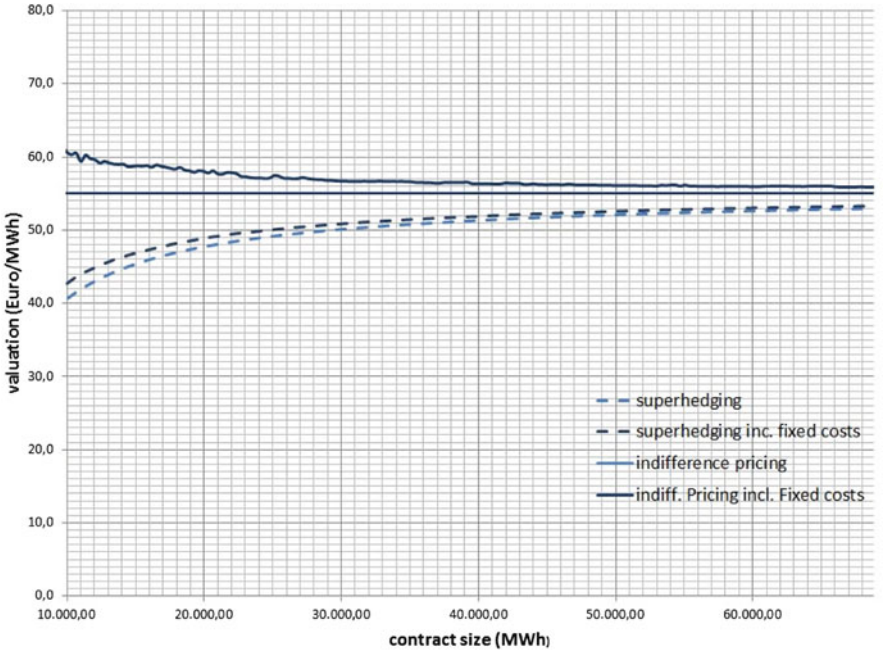

Fig. 15.6 Superhedging and indifference pricing

- The buyer's view: given the exercise price K, what is the optimal demand strategy of the buyer, and what is the resulting value of a swing option?
- The seller's view: what exercise price K should be offered by the seller, in view of the related optimal decision of the buyer?

Clearly the second question goes beyond the pure valuation issues raised and analyzed in the previous sections. While different approaches have been used to answer the first question, we remain in the framework of stochastic optimization

and base our elaboration of the buyer's problem on the method proposed first in [9], which was developed further in [10]. For the second question we build on these results and follow the approach in [3], which was extended in several directions by [13].

15.5.1 The Buyer's View

Again, we consider points in time $t \in \{0,1,2,\dots,T\}$. The delivery price K is fixed at $t=0$ and delivery (for a single commodity) is possible during some periods $t \in \{t_D,\dots,T\}$. The buyer specifies the actual consumption from the contract, D_t, one period before delivery. The quantities bought for the tth period are denoted by y_t, where $t \in \mathscr{T} = \{0,1,\dots,T-1\}$.

The exact strategy of the buyer clearly depends on his own liabilities (e.g., a producer of aluminum will behave differently from a pure trader) and his access to electricity spot markets and other parts of energy markets. We follow [3, 10] and [13], and model a trader with access to the electricity spot market, which is in some sense the worst case from the standpoint of a swing option seller, because the trader is not restricted by own liabilities: the basic problem lies in the fact that the buyer buys swing electricity at the delivery price if he thinks that spot prices will be high and does not buy if he thinks that spot prices will be low. If the buyer is right, this means for the seller that he will have to deliver when prices are high, which clearly is inconvenient.

In this framework, the trader solves the following optimization problem to find an optimal strategy—a consumption pattern $D = (D_0,\cdots,D_{T-1})$ such that D_t is deliverable during period $(t,t+1]$—for a swing option contract with given exercise price K. It is assumed that both the consumption in each period (base line schedule) and the accumulated consumption over the whole delivery period are restricted by lower and upper bounds and that the trader sells any consumption from the contract at the electricity spot market. By $S^e_{t,0}$ we denote the electricity spot prices; $\underline{e}_t$ and $\overline{e}_t$ represent lower and upper bounds for consumption in each period and $\underline{E}$, $\overline{E}$ refer to lower and upper bounds for the cumulated consumption.

$$\left\|\begin{array}{l} \max_D \sum_{t=0}^{T-1} \mathbb{E}\left[D_t\left(S^e_{t+1,0}-K\right)\right] \\ \text{subject to } \underline{e}_t \le D_t \le \overline{e}_t,\ \forall t \in \{0,\dots,T-1\}, \\ \underline{E} \le \sum_{t=0}^{T-1} D_t \le \overline{E}, \\ D_t \ge 0,\ \ \forall t \in \mathscr{T}, \\ \sum_{t=0}^{T-1} \mathbb{E}\left[D_t\left(S^e_{t+1,0}-K\right)\right] \ge 0, \\ D_t \text{ is nonanticipative.} \end{array}\right. \tag{28}$$

As pointed out in [10] the optimal value of this problem can be seen as the value of the swing option from the buyer's perspective, as long as it is not negative. If the optimal value is negative, the contract will not be concluded, which justifies the second to last constraint.

The formulation (28) is used in [3, 13] and can be extended in various directions. In particular, ramping constraints with *ratchets* ρ_t can be modeled by

$$-\rho_t \cdot \Delta \le y_t - y_{t-1} \le \rho_t \cdot \Delta, \tag{29}$$

where Δ is the length of the time periods.

15.5.2 The Seller's View

While the buyer's decision problem (28) does not make use of the hedging or acceptability concepts discussed before, the sellers decision is again modeled by acceptability pricing: the seller searches for the minimal delivery price and related hedging and production decisions such that the resulting profit and loss distribution remains acceptable.

The seller's decision problem is similar to (18): the decision variables are augmented by the strike price K and the contractual cashflows C_t are redefined by $C_t = D_t \cdot K$. In particular this means that (15.15) is replaced by

$$\begin{aligned} x^f_{t,0} \le\ & (1+r_f)x^f_{t-1,0} \\ & - \sum_{j=0}^{J} S^e_{t,j} y^e_{t,j} - \sum_{i=1}^{I} S^f_{t,i}(x^f_{t,i} - x^f_{t-1,i}) + \sum_{i=1}^{I} C^f_{t,i} + D_t \cdot K \\ & - \sum_{i=0}^{J}\sum_{j=0}^{J} \gamma_{t,ij} z_{t,ij} \\ & - \sum_{j=1}^{J} \zeta_j \frac{(x^e_{t,j} + x^e_{t-1,j})}{2}. \end{aligned} \tag{30}$$

Unfortunately, acceptability pricing by an extended version of (15.15) cannot be used directly, because the formulation would include the buyer's decisions D_t which are not decision variables of the seller. Instead, the problem has to be reformulated in the framework of bi-level optimization. The papers [3, 13] propose this approach but use simplified versions of this problem, both simplifying production decisions: the first does not model production, while the second uses an internal price for accounting between a production and a trading department as a proxy. The latter paper also gives a broad survey of related models and methods.

The basic problem in bi-level optimization lies in the fact that given a strike price K the buyer's problem (28) may have non-unique optimal decisions. Using the optimistic approach of bi-level optimization and assuming that the lower level chooses among its optimal decisions the best one from the seller's point of view (see, e.g., [7], also for the alternative—the pessimistic approach) the decision problem can be formulated as

$$\left\| \begin{array}{l} \text{Minimize (in } K, x^e, x^f, y, z, w \text{ and D): } w \\ \text{subject to} \\ \text{constraints (15.7)–(15.14) and (30)}, \\ \mathscr{A}(C_T + \sum_{j=1}^{J} S^e_{T,j} x^e_{t,j} + \sum_{i=1}^{I} S^f_{T,i} x^f_{T,i}) \geq 0, \\ D \in D^*_K, \\ x^e_t, x^f_t, y_t, z_t \text{ are nonanticipative,} \end{array} \right. \tag{31}$$

where D^*_K denotes the argmin-set (i.e., the set of optimal solutions), given the strike price K, of the buyer's problem (28).

Even in the simplest case bi-level problems like (31) are nonconvex, and strongly NP-hard. See, e.g., [6, 15] for necessary optimality conditions. A specific intricacy of bi-level problems lies in the fact that their feasible set can be disconnected.

Typical standard approaches for stochastic bi-level optimization are stochastic quasi-(sub)gradient methods and the MPEC approach. The first one is applicable only if the argmin-set of the buyer's problem is guaranteed to be a singleton for all relevant prices K. The latter approach consists in formulating the KKT conditions for the buyer's problem and to include them into the upper-level problem in order to code the argmin-set of the buyer's problem. This is very common but hard to use for multistage problems like (28), because complementarity conditions for each node have to be included, which makes the resulting formulation hard to solve. Kovacevic and Pflug [13] give an overview and propose some new algorithms, building on the fact that the optimistic bi-level problem can be approximated by an LP if polyhedral acceptability functionals like the average value-at-risk are used for $\mathscr{A}$ and that all decisions of both the buyer and the seller are dominated by the seller's decision regarding the strike price K.

The following example illustrates the bargaining situations between buyer and seller.

Example 3. We set $\overline{e}_t \equiv 5$, $\overline{E} \equiv 50$. The $\mathbb{A}$V@R-parameter is $\alpha = 0.15$ and the minimum $\mathbb{A}$V@R$_\alpha$-requirement $q = -20$. The spot price process S_t is modeled by a stochastic tree with 6 (monthly) stages. The generating unit and the fuels are as in the previous examples. In addition, we use three futures products (with exercise periods of length 2) for hedging. Three scenarios for lower bounds are considered: $\underline{e} = 0$, $\underline{e} = 0.4\overline{e}$, and $\underline{e} = 0.8\overline{e}$. Furthermore we use $\underline{E} = 0$ and $\overline{E} = 5\overline{e}$.

Figure 15.7 shows effects of these scenarios. In particular increasing lower bounds reduces the value of the contract for the buyer, while it increases the value for the seller. However, the drawback of increasing lower bounds is that also the largest feasible strike price decreases. Note that the feasible prices in this example are comparably small, because they refer to summer months, and the $\mathbb{A}$V@R-requirements are relatively mild.

15.6 Conclusions

We have shown that fair pricing of energy contracts is a difficult task, which can be accomplished by solving various multistage optimization problems. While (super) replication requires the solution of a deterministic program, acceptance and indifference pricing are based on stochastic programs. These programs can be quite complex, especially if the full available portfolio of the contract seller is modeled. For flexible contracts, it is even necessary to solve a stochastic bi-level program, since in this case the optimal pricing must be embedded into a game-theoretic model of the leader-follower type.

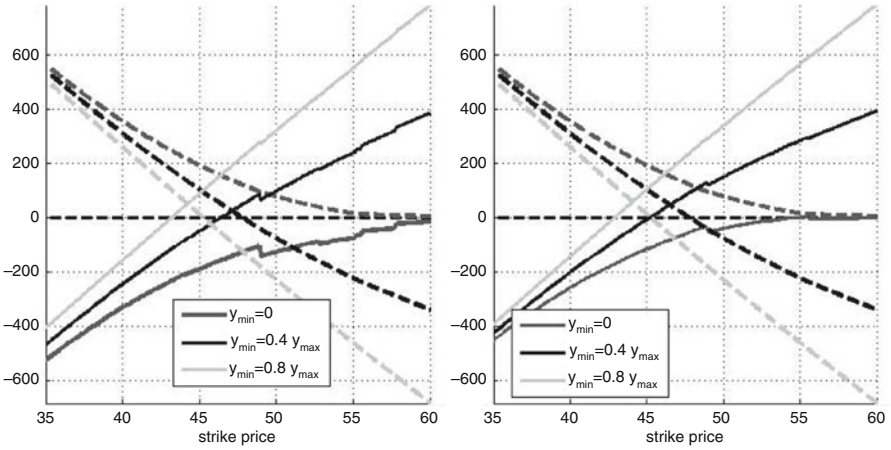

Fig. 15.7 Swing option pricing—the bargaining situation: the *left part* shows the buyer's expectation (*dashed*) versus the seller's average value-at-risk (*solid*). The *right part* shows the buyer's expectation (*dashed*) versus the seller's expectation

References

1. Artzner P, Delbaen F, Eber JM, Heath D (1999) Coherent measures of risk. Math Financ 9:203–228
2. Barbieri A, Garman M (2002) Understanding the valuation of swing contracts. Energy Power Risk Manag, FEA, Tech. Rep.
3. Broussev N, Pflug G (2009) Electricity swing options: behavioral models and pricing. Eur J Oper Res 197(39):1041–1050
4. Bühlmann H (1972) Mathematical risk theory. Die Grundlehren der mathematischen Wissenschaften, Band 172. Springer, New York
5. Carmona R (2009) Indifference pricing: theory and applications. Princeton series in financial engineering. Princeton University Press, Princeton

6. Dempe S (1992) A necessary and a sufficient optimality condition for bilevel programming problems. Optimization 25:341–354
7. Dempe S (2002) Foundations of bilevel programming. Kluwer Academic Publishers, Dordrecht
8. Föllmer H, Leukert P (1999) Quantile hedging. Financ Stoch 3:251–273
9. Frauendorfer K, Güssow J, Haarbrücker G, Kuhn D (2005) Stochastische optimierung im energiehandel: entscheidungsunterstützung und bewertung für das portfoliomanagement. Zeitschrift für Energie, Markt, Wettbewerb 1:59–66
10. Haarbrücker G, Kuhn D (2009) Valuation of electricity swing options by multistage stochastic programming. Automatica 45:889–899
11. Kaminski V, Gibner S (1995) Exotic options. In: Kaminski V (ed) Managing energy price risk. Risk Publications, London, pp 117–148
12. Kovacevic R, Paraschiv F (2012) Medium-term planning for thermal electricity production, OR Spectrum, Doi:10.1007/500291-013-0340-9
13. Kovacevic RM, Pflug GC (2013) Electricity swing option pricing by stochastic bilevel optimization: a survey and new approaches, available ref www.speps.org
14. McNeil AJ, Frey R, Embrechts P (2005) Quantitative risk management - concepts, techniques and tools. Princeton series in finance. Princeton University Press, Princeton
15. Outrata J (1993) Necessary optimality conditions for Stackelberg problems. J Optim 76:305–320
16. Pennanen T (2012) Introduction to convex optimization in financial markets. Math Program 134:157–186
17. Pflug G, Römisch W (2007) Modeling, measuring and managing risk. World Scientific, Singapore
18. Pilipović D (2007) Energy risk: valuing and managing energy derivatives, 2nd edn. McGraw-Hill Professionalition, New York
19. Pilipović D, Wengler J (1998) Getting into the swing. Energy Power Risk Manag 2:22–24
20. Rockafellar R (1974) Conjugate duality and optimization. In: CBMS-NSF regional conference series in applied mathematics, vol 16. SIAM, Philadelphia

Chapter 16
Energy Derivatives with Volume Controls

Fred Espen Benth and Marcus Eriksson

Abstract We analyse two classes of power derivatives with volume control, tolling agreements and flexible load contracts. Under certain assumptions, we can price a tolling agreement by resorting to theory of flexible load contracts, when using the fuel cost as numeraire in the power price. Tolling agreements can be priced as a strip of spread options under simple set of controls. Finally, we prove a general theory based on dynamic programming for these two classes of derivatives. We base our theory on price dynamics driven by Brownian motion.

16.1 Introduction

Electricity can be generated by a fuel like gas or coal or nuclear or from renewable sources like wind, sun or water. In most developed economies, the electricity producers sell their production in open power markets where they face a considerable price risk. But, equally important is the volume risk.

As low demand for power goes along with low prices, the producer will have a double effect on her revenues. On the other side, retailers must pay high prices when the demand is high, typically. Hedging instruments that cover both price and volume risk will provide an insurance against undesirable market situations like these. Such derivatives are traded in the power markets and include protection against low electricity prices, high fuel costs and/or demand uncertainty. In this chapter we will focus on tolling agreements and flexible load contracts.

With the development of renewable energy production, like wind power and solar, the volume risk also comes on the supply side. A wind mill generates power only when there is wind, with an effect being a function of the wind speed.

F.E. Benth (✉) • M. Eriksson
Centre of Mathematics for Applications (CMA) and Department of Mathematics,
University of Oslo, P.O. Box 1053 Blindern, N-0316 Oslo, Norway
e-mail: fredb@math.uio.no; mkerikss@math.uio.no

R.M. Kovacevic et al. (eds.), *Handbook of Risk Management in Energy Production and Trading*, International Series in Operations Research & Management Science 199,
DOI 10.1007/978-1-4614-9035-7_16,

Hence, the wind speed determines the volume power that can be produced, and one cannot necessarily generate more when prices are favourable. In this way, we might say that there is an opportunity loss for a wind mill producer if there are high prices in the power market, but no wind.

A tolling agreement is a financial contract that mimics the operation of a gas or coal-fired power plant. We analyse this contract in the case of power and fuel prices following a co-integrated Brownian motion-driven dynamics. We relate the contract to spread options, for which we provide an analytical pricing formula for our specific set-up. This will be an extension of the well-known Margrabe formula (see [6]).

Next, we show that the tolling agreement can be viewed as a flexible load contract. This class of derivatives gives the holder the right to buy power at a fixed price within given volume constraints. For example, the holder can buy power up to a certain total volume over a year. The tolling agreement, with rather general volume constraints, can be mapped into such a derivative, where the power price is using the fuel as a numeraire.

Finally, we study the two classes of derivatives for more general price processes applying dynamic programming. We associate the Hamilton–Jacobi–Bellman equations and provide verification theorems for the value function (or the price of the contract) and its optimal execution. These results rely heavily on recent developments presented in the papers [1, 2].

We present our analysis as follows. In the next section we analyse tolling agreements and derive an extension of Margrabe's formula . Then, in Sect. 16.3 we show that a tolling agreement is a flexible load contract. Section 16.4 deals with these contracts using dynamic programming.

16.2 Tolling Agreements

Let $P(s)$ be the price of power at time s and $C(s)$ be the price of a fuel (like coal or gas, say). If $u(s)$ is the production rate at time s, then the t-value of a power plant which is assumed to operate until time T will be

$$\int_t^T \mathrm{e}^{-r(s-t)}(P(s) - hC(s))u(s)\,ds. \tag{16.1}$$

Here, $t \leq T$ and the production rate u is an adapted control which is supposed to satisfy $u(s) \in [0, \bar{u}]$, where $\bar{u} < \infty$ is the maximal production rate from the power plant. Furthermore, r is a nonrandom (fixed) interest rate and $h > 0$ is the *heat rate*, e.g. the factor converting the units of the fuel into units of power.

A tolling agreement is now a financial option contract mimicking the ownership of a power plant fuelled by coal or gas. The holder of the tolling agreement will be paid at time T the amount given in (16.1), where u is the production rate decided by the holder. Obviously, the holder tries to maximize the payoff, which means that she

is maximizing over all *admissible* controls u. We say that a control u is admissible if it is adapted and bounded as above and denote the set of such controls by $\mathscr{A}(t,T)$.

The tolling agreement is paying the same as the profit/loss from operating a power plant. In this sense, the option is giving the holder access to the power plant in a virtual sense, without actually having to operate one. A tolling agreement could also be viewed as a real option valuation of a power plant project.

The price of a tolling agreement is

$$V(t) = \sup_{u\in\mathscr{A}(t,T)} \mathbb{E}_Q\left[\int_t^T \mathrm{e}^{-r(s-t)}(P(s)-hC(s))u(s)\,ds\,|\,\mathscr{F}_t\right]. \tag{16.2}$$

Here, Q is some pricing measure, incorporating the risk premium in the market. We are interested in both the optimal strategy u^* such that

$$V(t) = \mathbb{E}_Q\left[\int_t^T \mathrm{e}^{-r(s-t)}(P(s)-hC(s))u^*(s)\,ds\,|\,\mathscr{F}_t\right] \tag{16.3}$$

and the price $V(t)$ of the tolling agreement.

Note that the tolling agreement contract is an option paying out the accumulated present value of the difference between the power and fuel price. The payment stream is scaled by the production rate and thus is an option paying money according to price levels and volume decisions.

As long as there are no other constraints on the rate u than it should be adapted and bounded in the interval $[0,\bar{u}]$, it is obvious that the holder will produce at a maximal rate whenever this is advantageous, while turning off the production completely if it produces a negative payment. Thus, the optimal control is given by

$$u^*(s) = \bar{u}\mathbf{1}(P(s) > hC(s)), \tag{16.4}$$

where $\mathbf{1}(\cdot)$ is the indicator function. We show that this is indeed the optimal control in the following lemma:

Lemma 1. *The control $u^*(s)$ in (16.4) is admissible and is such that (16.3) holds. Hence, u^* is the optimal control.*

Proof. As the processes $P(s)$ and $C(s)$ are $\mathscr{F}_s$-measurable, so is $u^*(s)$. Moreover, we trivially have that $u^*(s)\in[0,\bar{u}]$. Thus, $u^*(s)\in\mathscr{A}(t,T)$.

Let u be a constant such that $u\in[0,\bar{u}]$. If $P-hC>0$, then $(P-hC)u\le(P-hC)\bar{u}$. If $P-hC<0$, then trivially $(P-hC)u\le 0$. Here, P and C are simply two constants. Thus, we find that for any admissible control $u\in\mathscr{A}(t,T)$

$$\int_t^T \mathrm{e}^{-r(s-t)}(P(s)-hC(s))u(s)\,ds \le \int_t^T \mathrm{e}^{-r(s-t)}(P(s)-hC(s))\bar{u}\mathbf{1}(P(s)>hC(s))\,ds,$$

and the lemma follows. □

From this result, we can express the value function of a tolling agreement as

$$V(t) = \bar{u}\mathbb{E}_Q\left[\int_t^T \mathrm{e}^{-r(s-t)}(P(s) - hC(s))^+ \, ds \,|\, \mathscr{F}_t\right]$$
$$= \bar{u}\int_t^T \mathrm{e}^{-r(s-t)}\mathbb{E}_Q\left[(P(s) - hC(s))^+ \,|\, \mathscr{F}_t\right] ds\,.$$

Hence, a tolling agreement is nothing but a strip of European call options with zero strike on the spread between power and fuel prices.

16.2.1 Pricing of Spread Options in the Energy Market

Let us study the pricing of spread options in the energy market for a particular bivariate exponential Gaussian process. We specify the dynamics as

$$P(t) = \Lambda_P(t)\exp\left(X(t) + Y_P(t)\right), \tag{16.5}$$
$$C(t) = \Lambda_C(t)\exp\left(kX(t) + Y_C(t)\right), \tag{16.6}$$

where $\Lambda_i(t)$ for $i = P,C$ are deterministic seasonality functions, assumed to be measurable and of at least polynomial growth, and k is a constant. Furthermore, $X(t)$ is a drifted Brownian motion

$$X(t) = \mu t + \sigma B(t)\,,$$

while $Y_i(t)$, $i = P,C$ are two correlated Ornstein–Uhlenbeck processes defined as

$$dY_i(t) = (\theta_i - \alpha_i Y_i(t))\, dt + \sigma_i\, dW_i(t) \tag{16.7}$$

for $i = P,C$. Here, B, W_P and W_C are three correlated Brownian motions such that $\mathbb{E}[dW_i(t)dB(t)] = \rho_i\, dt$ and $\mathbb{E}[dW_P(t)dW_C(t)] = \rho\, dt$, $\rho, \rho_i \in [-1,1]$. To make matters technically simpler, we suppose the model is already stated under the pricing measure Q.

Such two-factor models have been extensively used in the literature for commodity pricing; see for example Lucia and Schwartz [5] for modelling of power spot prices or Schwartz and Smith [8] for modelling of oil and gas. The non-stationary factor X models the long-term mean, towards which the short-term factor Y is mean reverting. The short-term fluctuations of prices are due to supply and demand imbalances, while the long-term movements are coming from macro-economic variables like inflation, limitations in reserves and technological innovations.

The solution of $Y_i, i = P,C$ is given by

$$Y_i(t) = Y_i(0)\mathrm{e}^{-\alpha_i t} + \frac{\theta_i}{\alpha_i}(1 - \mathrm{e}^{-\alpha_i t}) + \int_0^t \sigma_i \mathrm{e}^{-\alpha_i(t-s)}\, dW_i(s)\,. \tag{16.8}$$

We see from the model that the deseasonalized spot prices are co-integrated in their difference, that is,

$$\ln(P(t)/\Lambda_P(t)) - \frac{1}{k}\ln(C(t)/\Lambda_C(t)) = Y_P(t) - \frac{1}{k}Y_C(t),$$

where the right-hand side is a difference of two OU processes, which are stationary in the sense that the difference converges to a normal distribution with time going to infinity. From this relationship we see that the power prices P can be expressed in terms of the fuel prices C.

The variance-covariance matrix of the Brownian motions B, W_P and W_C is

$$\begin{bmatrix} 1 & \rho_P & \rho_C \\ \rho_P & 1 & \rho \\ \rho_C & \rho & 1 \end{bmatrix},$$

which is positive definite if and only if

$$\rho_P^2 + \rho_C^2 + \rho^2 < 1 + 2\rho\rho_P\rho_C\,.$$

We assume this to hold for the correlation coefficients, and note that in this case we can express the Brownian motions B, W_P and W_C in terms of a three-dimensional Brownian motion $\mathbf{B}(t)$ with independent coordinates. That is, it holds that

$$\begin{bmatrix} dB \\ dW_P \\ dW_C \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ \rho_P & \sqrt{1-\rho_P^2} & 0 \\ \rho_C & \frac{\rho-\rho_P\rho_C}{\sqrt{1-\rho_P^2}} & \frac{\sqrt{1-\rho_P^2-\rho_C^2-\rho^2+2\rho\rho_P\rho_C}}{\sqrt{1-\rho_P^2}} \end{bmatrix} d\mathbf{B}(t).$$

We denote the above matrix C. But then we can represent the two energy price dynamics as

$$P(t) = \exp\left(\mu_P(t) + \int_0^t \mathbf{a}_P(t-s)'\,d\mathbf{B}(s)\right) \tag{16.9}$$

$$C(t) = \exp\left(\mu_C(t) + \int_0^t \mathbf{a}_C(t-s)'\,d\mathbf{B}(s)\right) \tag{16.10}$$

for

$$\mu_P(t) = \ln\Lambda_P(t) + X(0) + \mu t + Y_P(0)\mathrm{e}^{-\alpha_P t} + \frac{\theta_P}{\alpha_P}(1-\mathrm{e}^{-\alpha_P t}) \tag{16.11}$$

$$\mu_C(t) = \ln\Lambda_C(t) + kX(0) + k\mu t + Y_C(0)\mathrm{e}^{-\alpha_C t} + \frac{\theta_C}{\alpha_C}(1-\mathrm{e}^{-\alpha_C t}) \tag{16.12}$$

and

$$\mathbf{a}_P(x)' = [\sigma, \sigma_P\mathrm{e}^{-\alpha_P x}, 0]C \tag{16.13}$$

$$\mathbf{a}_C(x)' = [k\sigma, 0, \sigma_C\mathrm{e}^{-\alpha_C x}]C. \tag{16.14}$$

Here, $\mathbf{x}'$ denotes the transpose of a vector. In the next proposition we state a generalization of the Margrabe formula for the model in (16.9) and (16.10) with general assumptions on the parameters μ_i and $\mathbf{a}_i$. Indeed, we suppose that $\mathbf{B}$ is a d-dimensional Brownian motion and that the parameters are naturally extended to this situation.

Proposition 1. *Suppose that P and C have dynamics given by (16.9) and (16.10) where $\mu_i(t)$ is a bounded function on each finite time interval and $\mathbf{a}_i(x)$ is an $\mathbb{R}^d$-valued measurable function which is square integrable on each finite time interval, $i = P,C$. Then the spread option price at time t on a contract that pays $(P(s) - hC(s))^+$ at time $s \geq t$ is given by*

$$\begin{aligned}\tilde{V}(t,s) = \exp&\left(\mu_P(s) - r(s-t) + \int_0^t \mathbf{a}_P(s-u)'\, d\mathbf{B}(u)\right.\\ &\left.+\frac{1}{2}\int_0^{s-t} \mathbf{a}_P(u)'\mathbf{a}_P(u)\, du\right)\Phi(d_1)\\ &- h\exp\left(\mu_C(s) - r(s-t) + \int_0^t \mathbf{a}_C(s-u)'\, d\mathbf{B}(u)\right.\\ &\left.+\frac{1}{2}\int_0^{s-t} \mathbf{a}_C(u)'\mathbf{a}_C(u)\, du\right)\Phi(d_2),\end{aligned}$$

where $d_1 = d_2 + \sigma_a(s-t)$,

$$\begin{aligned}d_2 = \frac{1}{\sigma_a(s-t)}&\left[\mu_P(s) - \mu_C(s) + \int_0^t \mathbf{a}_P(s-u)' - \mathbf{a}_C(s-u)'\, d\mathbf{B}(u)\right.\\ &\left.+\int_0^{s-t} (\mathbf{a}_P(u)' - \mathbf{a}_C(u)')\mathbf{a}_C(u)\, du - \ln h\right],\end{aligned}$$

$$\sigma_a^2(x) = \int_0^x (\mathbf{a}_P(u)' - \mathbf{a}_C(u)')(\mathbf{a}_P(u) - \mathbf{a}_C(u))\, du,$$

and Φ being the cumulative standard normal distribution function.

Proof. Since $(P(s) - hC(s))^+ = C(s)(P(s)/C(s) - h)^+$, we find from the dynamics of $C(s)$ that

$$\begin{aligned}\mathbb{E}_Q\left[(P(s) - hC(s))^+ \mid \mathscr{F}_t\right] = &\exp\left(\mu_C(s) + \frac{1}{2}\int_0^s \mathbf{a}_C(s-u)'\mathbf{a}_C(s-u)\, du\right)\\ &\times \mathbb{E}_Q\left[Z(s)\left(\frac{P(s)}{C(s)} - h\right)^+ \mid \mathscr{F}_t\right].\end{aligned}$$

Here, the martingale process $Z(v)$ for $v \leq s$ is defined as

$$Z(v) = \exp\left(\int_0^v \mathbf{a}_C(s-u)'\, d\mathbf{B}(u) - \frac{1}{2}\int_0^v \mathbf{a}_C(s-u)'\mathbf{a}_C(s-u)\, du\right).$$

But by Girsanov's theorem (see Øksendal [7], p. 162) there exists a probability measure $Q_C \sim Q$ with density process $Z(v)$ and where

$$d\mathbf{W}(u) = -\mathbf{a}_C(s-u)\, du + d\mathbf{B}(u)$$

is a d-dimensional Q_C-Brownian motion $u \in [0,s]$. Applying Bayes' theorem (see [4]) we obtain

$$\begin{aligned}
&\mathbb{E}_Q\left[Z(s)\left(\frac{P(s)}{C(s)}-h\right)^+ \mid \mathscr{F}_t\right] \\
&\quad = \mathbb{E}_Q[Z(s)\mid\mathscr{F}_t]\mathbb{E}_Q\left[\frac{Z(s)}{\mathbb{E}_Q[Z(s)\mid\mathscr{F}_t]}\left(\frac{P(s)}{C(s)}-h\right)^+ \mid \mathscr{F}_t\right] \\
&\quad = Z(t)\mathbb{E}_{Q_C}\left[\left(\frac{P(s)}{C(s)}-h\right)^+ \mid \mathscr{F}_t\right].
\end{aligned}$$

But from the dynamics of P and C in (16.9) and (16.10), we find the Q_C-dynamics to be

$$\begin{aligned}
\frac{P(s)}{C(s)} &= \exp\left(\mu_P(s)-\mu_C(s)+\int_t^s(\mathbf{a}_P(s-u)'-\mathbf{a}_C(s-u)')\mathbf{a}_C(s-u)\,du\right) \\
&\quad\times\exp\left(\int_t^s \mathbf{a}_P(s-u)'-\mathbf{a}_C(s-u)'\,d\mathbf{W}(u)\right).
\end{aligned}$$

The first exponential is $\mathscr{F}_t$-measurable, while the Itô integral in the second exponential is normally distributed with mean zero and variance equal to $\sigma_a^2(s-t)$. The proposition follows after computing the expectation using standard properties of the normal density function and rearranging terms. □

The extended Margrabe formula above contains the terms $\int_0^t \mathbf{a}_i(s-u)'\,d\mathbf{B}(u)$ for $i=P,C$, which for $s>t$ is not equal to the analogous terms $\int_0^t \mathbf{a}_i(t-u)'\,d\mathbf{B}(u)$ in the spot prices $P(t)$ and $C(t)$. Hence, in this general form we cannot express the option price in terms of the underlying spot. However, if we go back to the special case of a co-integrated spot price dynamics in (16.5) and (16.6) with X being a drifted Brownian motion and Y_i, $i=P,C$ correlated Ornstein–Uhlenbeck processes, we may simplify to at least obtain an explicit dependency on the factors $X(t)$ and $Y_i(t), i=P,C$. Using the above link to the general model, we get the spread option price in the specific case of the dynamics in (16.5) and (16.6) as

$$\begin{aligned}
\tilde{V}(t,s) &= \Lambda_P(s)\exp\left(X(t)+\mathrm{e}^{-\alpha_P(s-t)}Y_P(t)+\tilde{\theta}_P(s-t)\right)\Phi(d_1) \\
&\quad -h\Lambda_C(s)\exp\left(kX(t)+\mathrm{e}^{-\alpha_C(s-t)}Y_C(t)+\tilde{\theta}_C(s-t)\right)\Phi(d_2),
\end{aligned}$$

where $d_1=d_2+\sigma_a(s-t)$,

$$\begin{aligned}
\sigma_a^2(x) &= (1-k)^2\sigma^2 x+2(1-k)\rho_P\sigma\frac{\sigma_P}{\alpha_P}(1-\mathrm{e}^{-\alpha_P x})-2(1-k)\rho_C\sigma\frac{\sigma_C}{\alpha_C}(1-\mathrm{e}^{-\alpha_C x}) \\
&\quad+\frac{\sigma_P^2}{2\alpha_P}(1-\mathrm{e}^{-\alpha_P x})-2\rho\frac{\sigma_P\sigma_C}{\alpha_P+\alpha_C}(1-\mathrm{e}^{-(\alpha_P+\alpha_C)x})+\frac{\sigma_C^2}{2\alpha_C}(1-\mathrm{e}^{-2\alpha_C x}),
\end{aligned}$$

$$\tilde{\theta}_P(x) = (\mu-r+\frac{1}{2}\sigma^2)x+\frac{1}{\alpha_P}(\theta_P+\rho_P\sigma\sigma_P)(1-\mathrm{e}^{-\alpha_P x})+\frac{\sigma_P^2}{4\alpha_P}(1-\mathrm{e}^{-2\alpha_P x})$$

$$\tilde{\theta}_C(x) = (k\mu-r+\frac{1}{2}k^2\sigma^2)x+\frac{1}{\alpha_C}(\theta_C+\rho_C k\sigma\sigma_C)(1-\mathrm{e}^{-\alpha_C x})+\frac{\sigma_C^2}{4\alpha_C}(1-\mathrm{e}^{-2\alpha_C x}),$$

and

$$d_2 = \frac{1}{\sigma_a(s-t)}\Big[\ln(\Lambda_P(s)/\Lambda_C(s)) + (1-k)X(t) \\ + \mathrm{e}^{-\alpha_P(s-t)}Y_P(t) - \mathrm{e}^{-\alpha_C(s-t)}Y_C(t) - \ln h + \tilde{\theta}(s-t)\Big].$$

Here,

$$\tilde{\theta}(x) = (1-k)(\mu + k\sigma^2)x + \frac{\theta_P}{\alpha_P}(1-\mathrm{e}^{-\alpha_P x}) - \frac{\theta_C}{\alpha_C}(1-\mathrm{e}^{-\alpha_C x}) \\ + k\sigma\rho_P\frac{\sigma_P}{\alpha_P}(1-\mathrm{e}^{-\alpha_P x}) + (1-2k)\sigma\rho_C\frac{\sigma_C}{\alpha_C}(1-\mathrm{e}^{-\alpha_C x}) \\ + \rho\frac{\sigma_P\sigma_C}{\alpha_P+\alpha_C}(1-\mathrm{e}^{-(\alpha_P+\alpha_C)x}) - \frac{\sigma_C^2}{2\alpha_C}(1-\mathrm{e}^{-2\alpha_C x}).$$

Notice that the total spread volatility $\sigma_a(x)$ is not depending on ρ and σ when $k=1$.

We can relate the factors in the spread option price to the forward price associated to the two energies P and C. From the general price dynamics, we find the forward price for a contract delivering P at time $s \geq t$ as

$$\begin{aligned} f_P(t,s) &= \mathbb{E}_Q[P(s)\,|\,\mathscr{F}_t] \\ &= \mathbb{E}_Q\left[\exp\left(\mu_P(s) + \int_0^s \mathbf{a}_P(s-u)'\,d\mathbf{B}(u)\right)|\,\mathscr{F}_t\right] \\ &= \exp\left(\mu_P(s) + \int_0^t \mathbf{a}_P(s-u)'\,d\mathbf{B}(u)\right)\mathbb{E}_Q\left[\exp\left(\int_t^s \mathbf{a}_P(s-u)'\,d\mathbf{B}(u)\right)\right] \\ &= \exp\left(\mu_P(s) + \int_0^t \mathbf{a}_P(s-u)'\,d\mathbf{B}(u) + \frac{1}{2}\int_0^{s-t}\mathbf{a}_P(u)'\mathbf{a}_P(u)\,du\right), \end{aligned}$$

where we have applied the $\mathscr{F}_t$-measurability of the Itô integral and the independent increment property of Brownian motion. Analogously, it follows that

$$f_C(t,s) = \exp\left(\mu_C(s) + \int_0^t \mathbf{a}_C(s-u)'\,d\mathbf{B}(u) + \frac{1}{2}\int_0^{s-t}\mathbf{a}_C(u)'\mathbf{a}_C(u)\,du\right). \quad (16.15)$$

Inspecting Proposition 1, we obtain the spread price expressed in terms of the forward prices as

$$\tilde{V}(s,t) = \mathrm{e}^{-r(s-t)}\left\{f_P(t,s)\Phi(d_1) - hf_C(t,s)\Phi(d_2)\right\}$$

with $d_1 = d_2 + \sigma_a(s-t)$, σ_a defined in Proposition 1 and

$$d_2 = \frac{\ln f_P(t,s) - \ln f_C(t,s) - \ln h - \frac{1}{2}\Sigma^2(t-s)}{\sigma_a(s-t)}.$$

That the price of an option on the spread of the spots can be expressed in terms of the respective forward prices is not a surprise since indeed the contract can be viewed as an option on the spread of the forwards from the identities $f_P(s,s) = P(s)$ and $f_C(s,s) = C(s)$. The advantage is that we can mark the option price to observable market prices being the forwards. Furthermore, this is also the starting point for developing a hedge of the contract. For example, the deltas are easily computed to be

$$\frac{\partial \tilde{V}(t,s)}{\partial f_P(t,s)} = \mathrm{e}^{-r(s-t)}\Phi(d_1),$$
$$\frac{\partial \tilde{V}(t,s)}{\partial f_C(t,s)} = -h\mathrm{e}^{-r(s-t)}\Phi(d_2),$$

via direct differentiation and usage of the expressions for d_1 and d_2.

16.3 Flexible Load Contracts

One may easily imagine situations where there are additional constraints on the set of possible controls u in $\mathscr{A}(t,T)$ defining the problem (16.2). For example, a gas-fired power plant cannot simply switch on or off the production, as it takes time to close the plant and time to restart it (the so-called ramping times). Also, it might be situations where one has to make the plant run for regulatory reasons. Thus, we may be interested in considering a strictly smaller set of admissible controls $\tilde{\mathscr{A}}(t,T) \subset \mathscr{A}(t,T)$ defined in connection with tolling agreements. Thus, we consider the stochastic control problem

$$V(t) = \sup_{u \in \tilde{\mathscr{A}}(t,T)} \mathbb{E}_Q\left[\int_t^T \mathrm{e}^{-r(s-t)}(P(s) - hC(s))u(s)\,ds \,|\, \mathscr{F}_t\right]. \tag{16.16}$$

Below, we translate the stochastic control problem (16.16) into a structure which resembles a call option payoff. In fact, as it turns out, one may reduce the problem into controlling only the power price, using the fuel price C as a numeraire. Hence, instead of facing a bivariate stochastic control problem, we can recast it as a univariate one, which is considerably simpler to analyse. In Sect. 16.4 we give a rigorous treatment of the univariate and bivariate stochastic control problem in a general setting. Furthermore, we analyse different choices of the restricted control set $\tilde{\mathscr{A}}(t,T)$.

The reduction from the bivariate to the univariate case leads us to another set of swing options called *flexible load contracts*. Flexible load contracts give the holder the right to buy power at a given price with given volume constraints over a set of pre-defined hours.

We state and prove the following proposition on the stochastic control problem (16.16).

Proposition 2. *Let P and C have dynamics as in (16.9) and (16.10), respectively. It holds for the stochastic control problem (16.16)*

$$V(t) = f_C(t,T)g^{-1}(T) \sup_{u\in\mathscr{A}(t,T)} \mathbb{E}_{Q_C}\left[\int_t^T \mathrm{e}^{-r(T-s)} g(s)\left(\tilde{P}(s)-h\right)u(s)\,ds\,|\,\mathscr{F}_t\right],$$

where

$$g(t) = \exp\left(\mu_C(t) - \frac{1}{2}\int_0^t \mathbf{a}_C(t-u)'\mathbf{a}_C(t-u)\,du\right),$$

the probability Q_C is equivalent to Q with density process

$$Z(t) = \exp\left(\int_0^t \mathbf{a}_C(T-u)'\,d\mathbf{B}(u) - \frac{1}{2}\int_0^t \mathbf{a}_C(T-u)'\mathbf{a}_C(T-u)\,du\right),$$

and $d\mathbf{W}(u) = -\mathbf{a}_C(T-u)\,du + d\mathbf{B}(u)$ is a Q_C-Brownian motion for $u \le T$. Moreover, the price dynamics of $\tilde{P}(s) = P(s)/C(s)$ is given by

$$\tilde{P}(s) = \exp\left(\mu_P(s) - \mu_C(s) + \int_0^s (\mathbf{a}_P(u)' - \mathbf{a}_C(u)')\mathbf{a}_C(u)\,du \right.$$
$$\left. + \int_0^s \mathbf{a}_P(s-u)' - \mathbf{a}_C(s-u)'\,d\mathbf{W}(u)\right).$$

Proof. Since $C(s) = C(T)C(s)/C(T)$, we find

$$\begin{aligned} J(t) &:= \mathbb{E}_Q\left[\int_t^T \mathrm{e}^{-r(T-s)}(P(s)-hC(s))u(s)\,ds\,|\,\mathscr{F}_t\right] \\ &= \mathbb{E}_Q\left[C(T)\int_t^T \mathrm{e}^{-r(t-s)}\frac{C(s)}{C(T)}\left(\tilde{P}(s)-h\right)u(s)\,ds\,|\,\mathscr{F}_t\right]. \end{aligned}$$

But, from the dynamics of C in (16.10), $C(T) = g(T)Z(T)$, with g and Z defined in the proposition. The process $Z(t)$ is a martingale and by Girsanov's theorem (see [7]) is the density process of an equivalent probability Q_C for which $\mathbf{W}(u)$ defined in the proposition is a Brownian motion, $u \le T$. Hence, it follows from Bayes' formula (see [4]) followed by Fubini–Tonelli's theorem that

$$\begin{aligned} J(t) &= Z(t)g(T)\mathbb{E}_{Q_C}\left[\int_t^T \mathrm{e}^{-r(T-s)}\frac{C(s)}{C(T)}\left(\tilde{P}(s)-h\right)u(s)\,ds\,|\,\mathscr{F}_t\right] \\ &= Z(t)g(T)\int_t^T \mathrm{e}^{r(T-s)}\frac{g(s)}{g(T)}\mathbb{E}_{Q_C}\left[\frac{Z(s)}{Z(T)}\left(\tilde{P}(s)-h\right)u(s)\,|\,\mathscr{F}_t\right]ds \\ &= Z(t)\int_t^T \mathrm{e}^{-r(T-s)}g(s)\mathbb{E}_{Q_C}\left[\mathbb{E}_{Q_C}\left[\frac{Z(s)}{Z(T)}\,|\,\mathscr{F}_s\right]\left(\tilde{P}(s)-h\right)u(s)\,|\,\mathscr{F}_t\right]ds. \end{aligned}$$

In the last equality we applied the tower property of conditional expectation together with the $\mathscr{F}_s$-measurability of $(\tilde{P}(s)-h)u(s)$. Invoking Bayes' formula once again, we find

$$\mathbb{E}_{Q_C}\left[\frac{Z(s)}{Z(T)}\right] = \frac{\mathbb{E}_Q\left[\frac{Z(s)}{Z(T)}Z(T)\,|\,\mathscr{F}_s\right]}{\mathbb{E}_Q[Z(T)\,|\,\mathscr{F}_s]} = 1$$

by the martingale property of Z. We conclude that

$$J(t) = Z(t)\mathbb{E}_{Q_C}\left[\int_t^T \mathrm{e}^{-r(T-s)}g(s)\left(\tilde{P}(s)-h\right)u(s)\,ds\,|\,\mathscr{F}_t\right].$$

By the definition of the dynamics of P and C, and the Girsanov change of probability, we find the dynamics of $\tilde{P}$ under Q_C easily. Using the definition of the forward price $f_C(t,T)$ in (16.15) completes the proof. □

From this proposition we may view a tolling agreement on P and C with dynamics (16.9) and (16.10), respectively, and controls u in $\tilde{\mathscr{A}}(t,T)$ as a flexible load contract on P/C. Flexible load contracts are interesting in their own right, being popular derivatives in the power market.

Note that if $\tilde{\mathscr{A}}(t,T) = \mathscr{A}(t,T)$, it is simple to see that the optimal control for the problem

$$\sup_{u\in\tilde{\mathscr{A}}(t,T)} \mathbb{E}_{Q_C}\left[\int_t^T \mathrm{e}^{-r(T-s)}g(s)\left(\tilde{P}(s)-h\right)u(s)\,ds\,|\,\mathscr{F}_t\right] \tag{16.17}$$

is given by a bang-bang type, namely

$$u^*(s) = \bar{u}\mathbf{1}(\tilde{P}(s) > h). \tag{16.18}$$

However, since $\tilde{P}(s) = P(s)/C(s)$, this coincides, not surprisingly, with the control in (16.4). Moreover, we can move the expectation operator inside and relate the value function V to the price of a strip of call options on $\tilde{P}$ with exercise time $s \in [t,T]$ and strike h. However, again using that $\tilde{P}$ is the ratio between the power and fuel price, we can compute an analytic price that will be equal to the one in Proposition 1.

16.4 The Dynamic Programming Approach

In this section we use the dynamic programming approach to consider the univariate and bivariate stochastic control problems of the previous sections in a more general context. We give a rigorous statement of the univariate case, prove some general properties and derive the associated Hamilton–Jacobi–Bellman (HJB) equation as well as a verification theorem. Then we extend those results to a multidimensional model (which, in particular, covers the case of tolling agreements), which we relate to the bivariate case in Sect. 16.2. Finally, we discuss the delicate issue of boundary conditions which are important in numerical solutions of the control problem.

16.4.1 Univariate Case

Let $P(t)$ be the unique strong solution to the stochastic differential equation

$$dP(t) = \mu(t,P(t))dt + \sigma(t,P(t))dW(t), \tag{16.19}$$

where W is a one-dimensional Brownian motion, $\mu, \sigma : [0,T] \times \mathbb{R} \mapsto \mathbb{R}$ being continuous functions satisfying a uniform Lipschitz condition in the second argument (see [7]) and σ taking positive values. Referring to the discussions in the sections above, we may view P as the power price or the power price with the fuel as numeraire. The exposition that follows is closely based on the analysis in [1].

We focus on a flexible load contract with maximal volume constraint on the controls $u \in \mathscr{A}(t,T)$. To this end, define for $u \in \mathscr{A}(t,T)$

$$dZ(t) := u(t)dt. \tag{16.20}$$

The process Z measures the total volume produced until time t using the production rate u. We define the restricted control set $\tilde{\mathscr{A}}(t,T)$ as the set of controls $u \in \mathscr{A}(t,T)$ with the additional constraint that $Z(T) \in [0,M]$, where $M > 0, \bar{u} > 0$ are constants. Hence, we have bounded production rates as well as a total volume constraint M that must be satisfied by the production rate.

It is convenient to extend the state space of the control problem (16.16) to include the total volume process Z. Thus, we define the value function $V(t,z,p)$, defined on $\mathscr{S} := [0,T] \times [0,M] \times \mathbb{R}$, as

$$V(t,z,p) = \sup_{u \in \tilde{\mathscr{A}}(t,T)} \mathbb{E}\left[\int_t^T \mathrm{e}^{-r(s-t)} (P(s) - h) u(s) ds \,|\, P(t) = p, Z(t) = z \right]. \tag{16.21}$$

Note that the process Z appears in the control set. However, we may also substitute $u(s)\,ds$ by $dZ(s)$ in the formulation above to get it explicitly appearing in the value function. Note also that $h > 0$ is the strike price in the flexible load contract. We also observe that the state processes P and Z are Markovian, which explains why we condition on $P(t) = p, Z(t) = z$ rather than the filtration $\mathscr{F}_t$ in the definition of the value function in (16.21).

16.4.1.1 Some General Results

We now state some general results about the value function. First, observe that if $z = M$, $\tilde{\mathscr{A}}(t,T)$ consists of only $u = 0$. Hence,

$$V(t,M,p) = 0, t \leq T, p \geq 0. \tag{16.22}$$

Therefore, the maximal constraint comes in as a boundary condition on the value function. Also, we have $V(T,z,p) = 0$ for $(z,p) \in [0,T] \times \mathbb{R}$.

Observe that the maximum volume constraint is absent if $M - z \geq \bar{u}(T-t)$ as it will be reached trivially. In this case we have the following result.

Proposition 3. *If $M - z \geq \bar{u}(T-t)$, an optimal control is given by*

$$u^*(s) = \bar{u}\mathbf{1}(P(s) - h > 0)\,. \tag{16.23}$$

Proof. Clearly $u^*(s)$ is admissible. For any $u \in \mathscr{A}(t,T)$ we have

$$\begin{aligned} J(t,z,x:u) :&= \mathbb{E}\left[\int_t^T \mathrm{e}^{-r(s-t)}(P(s)-h)u(s)\,ds\,|\,P(t)=p, Z(t)=z\right] \\ &= \mathbb{E}\left[\int_t^T \mathrm{e}^{-r(s-t)}(P(s)-h)u(s)\mathbf{1}(P(s)-h>0)\,ds\,|\,P(t)=p, Z(t)=z\right] \\ &+ \mathbb{E}\left[\int_t^T \mathrm{e}^{-r(s-t)}(P(s)-h)u(s)\mathbf{1}(P(s)-h\leq 0)ds\,|\,P(t)=p, Z(t)=z\right] \\ &\leq \mathbb{E}\left[\int_t^T \mathrm{e}^{-r(s-t)}(P(s)-h)u^*(s)\,ds\,|\,P(t)=p, Z(t)=z\right], \end{aligned} \tag{16.24}$$

and the result follows. □

With the optimal control (16.23) we see that the value of the contract is independent of the amount of power produced. As a consequence we get

Lemma 2. *If $M - z \geq \bar{u}(T-t)$, then the marginal value of the value function with respect to the production volume becomes*

$$\frac{\partial V(t,z,p)}{\partial z} = 0\,, z < M - \bar{u}(T-t)\,. \tag{16.25}$$

The immediate interpretation of this result is that the holder of the flexible load contract cannot lose value by producing more power. This is intuitively clear, since he is allowed to produce as much (or little) as he wants when the maximal volume constraint is not effective. However, in the presence of an effective volume constraint, we find a different result on the marginal value.

Proposition 4. *If $M - z < \bar{u}(T-t)$, then the marginal value of the value function with respect to the production volume satisfies*

$$\frac{\partial V(t,z,p)}{\partial z} \leq 0\,, z > M - \bar{u}(T-t)\,. \tag{16.26}$$

Since the holder has a maximal amount of power that can be produced until maturity the set of admissible control rates decreases as more power is produced. Hence, the value function decreases with respect to an increasing z, meaning more production has taken place. For a proof we refer to [1], Proposition 3.3.

When the maximal constraint was absent we could find an optimal production rate explicitly according to Proposition 3. In the presence of an effective maximal

constraint, we use Bellman's principle of optimality to derive an HJB equation and via a verification theorem we are able to show that its solution (under appropriate assumptions) coincides with the optimal value of the contract. This is our next task.

16.4.1.2 Necessary Conditions

To study Problem (16.21) under an effective total volume constraint, we first derive the associated HJB equation. This is done by appealing to the *dynamic programming principle* (see, e.g., [3]), which is expressed in the present situation as: for all times $0 \leq t < w \leq T$, the condition

$$V(t,Z(t),P(t)) = \sup_{u\in\mathscr{A}(t,T)} \mathbb{E}\left[\int_t^w \mathrm{e}^{-r(s-t)}(P(s)-h)u(s)ds \right. \\ \left. +\mathrm{e}^{-r(w-t)}V(w,Z(w),P(w))\,|\,P(t),Z(t)\right] \quad (16.27)$$

holds. Rewrite (16.27) as

$$\sup_{u\in\mathscr{A}(t,T)} \mathbb{E}\left[\int_t^w \mathrm{e}^{-rs}(P(s)-h)u(s)\,ds \right. \\ \left. +\left(\mathrm{e}^{-rw}V(w,Z(w),P(w)) - \mathrm{e}^{-rt}V(t,Z(t),P(t))\right)\,|\,P(t),Z(t)\right] = 0. \quad (16.28)$$

To proceed, we assume that V is a smooth function, that is, $V \in C^{1,1,2}(\mathscr{S})$. Applying the Itô formula to the process $t \mapsto \mathrm{e}^{-rt}V(t,Z(t),P(t))$ yields

$$\begin{aligned}
&\mathrm{e}^{rt}d(\mathrm{e}^{-rt}V(t,Z(t),P(t))) \\
&\quad = (V_t(t,Z(t),P_t) - rV(t,Z(t),P(t)) + u(t)V_z(t,Z(t),P(t)))\,dt \\
&\quad + V_p(t,Z(t),P(t))\,(\mu(t,P(t))\,dt + \sigma(t,P(t))\,dW(t)) \\
&\quad + \frac{1}{2}\sigma^2(t,P(t))V_{pp}(t,Z(t),P(t))\,dt.
\end{aligned}$$

Here, we have used the notation $V_t = \partial V/\partial t$, $V_k = \partial V/\partial k$ for $k = p,z$ and $V_{pp} = \partial^2 V/\partial p^2$. Using this, we can rewrite (16.28) as

$$\sup_{u\in\mathscr{A}(t,T)} \mathbb{E}\left[\frac{1}{w-t}\int_t^w \mathrm{e}^{-rs}((P(s)-h)u(s) + (\mathscr{L}-r)V(s,Z(s),P(s)) \right. \\ \left. +u(s)V_z(s,Z(s),P(s)))ds + (Y(w)-Y(t))\,|\,P(t),Z(t)\right] = 0, \quad (16.29)$$

where the linear operator $\mathscr{L}$ is defined on $C^{1,1,2}(\mathscr{S})$ as

$$\mathscr{L}F(t,z,p) = F_t(t,z,p) + \mu(t,p)F_p(t,z,p) + \frac{1}{2}\sigma^2(t,p)F_{pp}(t,z,p), \quad (16.30)$$

and the (local) martingale Y is defined as

$$Y(t) = \int_0^t \mathrm{e}^{-rs}\sigma(s,P(s))V_p(s,Z(s),P(s))dW(s)\,.$$

Under appropriate growth conditions on σ and V, the process Y will be a true martingale. Assuming this, we obtain

$$\mathbb{E}\left[\frac{1}{w-t}\int_t^w \mathrm{e}^{-rs}((P(s)-h)u(s)+(\mathscr{L}-r)V(s,Z(s),P(s)) \right.$$
$$\left. +u(s)V_z(s,Z(s),P(s)))\,ds\,|\,P(t),Z(t)\right] \le 0\,,$$

for each admissible $u \in \tilde{\mathscr{A}}(t,T)$. Furthermore, if there exists an optimal admissible control policy $u^* \in \tilde{\mathscr{A}}(t,T)$, then

$$\mathbb{E}\left[\frac{1}{w-t}\int_t^w \mathrm{e}^{-rs}((P(s)-h)u^*(s)+(\mathscr{L}-r)V(s,Z(s),P(s)) \right.$$
$$\left. +u^*(s)V_z(s,Z(s),P(s)))\,ds\,|\,P(t),Z(t)\right] = 0.$$

Now, under appropriate regularity conditions on V (see, e.g., [3]), we can pass to the limit $w \to t$ in (16.29) and, consequently, end up with the HJB equation

$$V_t(t,z,p)+\frac{1}{2}\sigma^2(t,p)V_{pp}(t,z,p)+\mu(t,p)V_p(t,z,p)-rV(t,z,p)$$
$$+\sup_u\{u(t)(p-h+V_z(t,z,p))\} = 0\,, \tag{16.31}$$

where u varies over the set of functions defined on $[0,T]$ satisfying the conditions $0 \le u(t) \le \bar{u}$, $\int_0^t u(s)ds = z$, and $\int_0^T u(t)dt \le M$. We remark that if the value V does not satisfy the required smoothness conditions, it may still be possible to identify it as a viscosity solution of (16.31)—for details, see [9].

Intuitively, we can interpret the quantity V_z in (16.31) as the (marginal) lost option value. From (16.31) it seems clear that the boundary $p-h=-V_z(t,z,p)$ plays a key role when determining the optimal control rate. Indeed, if $p-h>-V_z(t,z,p)$ for $t\in[0,T]$, that is, payoff dominates the lost option value, then

$$\sup_u\{u(t)(p-h+V_z(t,z,p))\} = \bar{u}(p-h+V_z(t,z,p)).$$

On the other hand, if $p-h \le -V_z(t,z,p)$, then

$$\sup_u\{u(p-h+V_z(t,z,p))\} = 0\,.$$

Define now the admissible Markovian exercise policy $\hat{u}(t,Z(t),P(t))$ by

$$\hat{u}(t,z,p) = \begin{cases} \bar{u}, & p-h>-V_z(t,z,p), \\ 0, & p-h\le -V_z(t,z,p)\,, \end{cases} \tag{16.32}$$

for all $t\in[0,T]$. We denote by $d\hat{Z}(t)=\hat{u}(t)\,dt$ the corresponding volume produced from this policy.

We note that if $M \geq \bar{u}T$, then, by Lemma 2, $V_z(t,z,p) = 0$ and the control $\hat{u}$, defined in (16.32), coincide with the optimal control in Proposition 3. The presence of an effective volume constraint appears in the restricted set $\tilde{\mathscr{A}}(t,T)$ of admissible controls.

16.4.1.3 Sufficient Conditions

In the following theorem we present a set of sufficient conditions for the value function in (16.21) to be the solution of the HJB equation (16.31) along with the proposed control $\hat{u}$ to be the optimal one:

Theorem 1. *Assume that a function $F : \mathscr{S} \to \mathbb{R}$ satisfies the following conditions:*

1. *$F(T,\cdot,\cdot) \equiv 0$, $F(\cdot,M,\cdot) = 0$ and $F \in C^{1,1,2}(\mathscr{S})$,*
2. *$(\mathscr{L} - r)F(t,z,p) + u(t)\left(F_z(t,z,p) + (p-h)\right) \leq 0$ for all $(t,z,p) \in \mathscr{S}$ and $u(t) = u(t,z,p)$ defining an admissible Markovian control $u \in \tilde{\mathscr{A}}(t,T)$, where the operator $\mathscr{L}$ is defined in (16.30), and*
3. *the process $Y : t \mapsto \int_0^t \mathrm{e}^{-rs}\sigma(s,P(s))F_p(s,Z(s),P(s))dW(s)$ is a martingale.*

Then $F(t,Z(t),P(t)) \geq V(t,Z(t),P(t))$ for all t.

In addition, if there exists an admissible Markovian control u^o such that

$$\begin{aligned}(\mathscr{L} - r)F(t,z,p) + \sup_u(u(t)\left(F_z(t,z,p) + (p-h)\right)) = \\ (\mathscr{L} - r)F(t,z,p) + u^o(t)\left(F_z(t,z,p) + (p-h)\right) = 0,\end{aligned} \tag{16.33}$$

for all $(t,z,p) \in \mathscr{S}$, then $u^o = u^*$ and the function F coincides with the value function V.

Proof. Let $u \in \tilde{\mathscr{A}}(t,T)$ and $t \in [0,T]$. First, we find using the Itô formula to the process $t \mapsto \mathrm{e}^{-rt}F(t,Z(t),P(t))$ that

$$\begin{aligned}d(\mathrm{e}^{-rt}F(t,Z(t),P(t))) = \mathrm{e}^{-rt}\left[(\mathscr{L} - r)F(t,Z(t),P(t)) + u(t)F_z(t,Z(t),P(t))\right. \\ \left. + \sigma(t,P(t))F_p(t,Z(t),P(t))dW(t)\right].\end{aligned}$$

Using assumption (i) and conditioning up to time t, we find that

$$\begin{aligned}0 = \mathrm{e}^{-rt}F(t,Z(t),P(t)) + \mathbb{E}\left[\int_t^T \mathrm{e}^{-rs}(\mathscr{L} - r)F(s,Z(s),P(s))\,ds \,|\, P(t),Z(t)\right] \\ + \mathbb{E}\left[\int_t^T \mathrm{e}^{-rs}F_z(s,Z(s),P(s))u(s)\,ds \,|\, \mathscr{F}_t\right] \\ + \mathbb{E}\left[(Y(T) - Y(t)) \,|\, P(t),Z(t)\right].\end{aligned}$$

Now, assumptions (ii) and (iii) imply that

$$0 \leq \mathrm{e}^{-rt}F(t,Z(t),P(t)) - \mathbb{E}\left[\int_t^T \mathrm{e}^{-rs}(P(s) - h)u(s)\,ds \,|\, P(t),Z(t)\right], \tag{16.34}$$

which is equivalent to the first claim.

Assume now that there exists an admissible u^o such that the condition (16.33) holds. Then with exactly the same calculation as above we find that

$$0 = \mathrm{e}^{-rt} F(t, Z(t), P(t)) - \mathbb{E}\left[\int_t^T \mathrm{e}^{-rs} (P(s) - h) u^o(s)\, ds \,|\, P(t), Z(t) \right], \qquad (16.35)$$

which concludes the proof. □

From this theorem we see that $\hat{u}$ is an optimal control if we can prove sufficient regularity of the value function. We see that the optimal production policy still has a bang-bang structure, that is, we produce at maximal or minimal rates. However, the threshold for switching from one to the other is not determined by whether one has a positive income stream or not, but by having a *sufficiently* high positive income stream. As V_z is non-positive, we may produce at minimal rate although $P(t)$, the power price, is bigger than the strike h. The maximal production constraint will make the holder of the contract to wait for prices being sufficiently higher than the strike.

We could add the constraint that the holder of the flexible load contract has to produce a minimal volume, that is, that $Z(T) \geq m$ for a constant $m < M$. This would force the holder to produce at a higher rate than zero, even possibly when he can lose money. It is to be expected that the simple bang-bang structure gets lost in this situation. A minimal total volume constraint is a subject for further research.

16.4.2 Multidimensional Model

We now consider the multidimensional case, interpreted as a derivative written on several price processes like the tolling agreement. We outline the general formulation before we relate this to the power and fuel prices and the tolling agreement (16.1). The presentation in this subsection is heavily influenced by the analysis in [2].

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuous function of at least linear growth. Define a stochastic process $\mathbf{X}(t) \in \mathbb{R}^n$ to be the unique strong solution of the stochastic differential equations:

$$d\mathbf{X}(t) = \mathbf{a}(t, \mathbf{X}(t)) + \Sigma(t, \mathbf{X}(t))\, d\mathbf{W}(t), \qquad (16.36)$$

where Σ is an $n \times m$-matrix-valued function on $[0,T] \times \mathbb{R}^n$ and $\mathbf{W}(t)$ an m-dimensional Brownian motion. We view the components $\mathbf{X}_i$ as prices for different commodities which may be correlated via Σ and related via the function f.

Define the value function on $\mathscr{S}^n := [0,T] \times [0,M] \times \mathbb{R}^n$ as

$$V(t, z, \mathbf{x}) = \sup_{u \in \tilde{\mathscr{A}}(t,T)} \mathbb{E}\left[\int_t^T \mathrm{e}^{-r(s-t)} f(\mathbf{X}(s)) u(s)\, ds \,|\, \mathbf{X}(t) = \mathbf{x} \right]. \qquad (16.37)$$

Note that with $n = 1$, $\mathbf{X}(t) = P(t)$ and $f(x) = x - h$, we are back to the case of the previous subsection. In the general model, the function f plays the role as mixing the various factors of $\mathbf{X}$ into one or more price processes, as well as describing the payoff structure. One can modify the results in Proposition 3, Lemma 2 and Proposition 4 to cover this more general situation. Moreover, we can also cover the tolling agreement considered in Sect. 16.3. We restrict our attention to this contract type in the following.

Let

$$d\mathbf{X}(t) = [dP_t, dC_t]', \tag{16.38}$$

with $\mathbf{a}(t,\mathbf{X}(t))$ being a vector in $\mathbb{R}^2$ having coordinates $\mu_i(t,\mathbf{X}(t))$, $i = 1,2$, and $\Sigma(t,\mathbf{X}(t))$ a 2×3 matrix with elements $\sigma_{ij}(t,\mathbf{X}(t))$, $i = 1,2, \mathrm{j} = 1,2,3$. We denote by $\mathbf{B}$ a three-dimensional Brownian motion. The coordinates μ_i and $\sigma_i j$ are supposed to be Lipschitz function with at most linear growth in $\mathbf{x} \in \mathbb{R}^2$, uniformly in $t \in [0,T]$. Define, for $\mathbf{x} = (p,c) \in \mathbb{R}^2$, $f(\mathbf{x}) := p - hc$. The value of the contract is now given by the value function, defined on $\mathscr{S}^2 := [0,T] \times [0,M] \times \mathbb{R}^2$

$$V(t,z,p,c) = \sup_{u \in \tilde{\mathscr{A}}(t,T)} \mathbb{E}\left[\int_t^T \mathrm{e}^{-r(s-t)} (P(s) - hC(s))u(s)\, ds \,|\, P(t) = p, Z(t) = z \right]. \tag{16.39}$$

Note that the control u scales the spread between the electricity and fuel price. Furthermore, note that Lemma 1 is just a special case of Proposition 3 with $P(s) - h$ replaced by $P(s) - hC(s)$.

The necessary and sufficient conditions for an optimal value are derived, following the same procedure as in previous section. The associated HJB equation to the problem (16.39) reads as

$$(\tilde{\mathscr{L}} - r)V(t,z,p,c) + \sup_u \{u(t)(f(p,c) + V_z(t,z,p,c))\} = 0, \tag{16.40}$$

where the operator $\tilde{\mathscr{L}}$ is defined as

$$\begin{aligned} \tilde{\mathscr{L}}F(t,z,p,c) = {} & F_t(t,z,p,c) + \mu_1(t,p,c)F_p(t,z,p,c) + \mu_2(t,p,c)F_c(t,z,p,c) \\ & + \frac{1}{2}(\Sigma\Sigma')_{11}(t,p,c)F_{pp}(t,z,p,c) \\ & + (\Sigma\Sigma')_{12}F_{pc}(t,z,p,c) + \frac{1}{2}(\Sigma\Sigma')_{22}F_{cc}(t,z,p,c). \end{aligned} \tag{16.41}$$

We have the following verification theorem:

Theorem 2. *Assume that a function $F : \mathscr{S}^2 \to \mathbb{R}$ satisfies the following conditions:*

1. *$F(T,\cdot,\cdot) \equiv 0$, $F(\cdot,M,\cdot) = 0$, and $F \in C^{1,1,2}(\mathscr{S}^2)$,*
2. *$(\tilde{\mathscr{L}} - r)F(t,z,p,c) + u(t)\,(F_z(t,z,p,c) + f(p,c))) \le 0$ for all $(t,z,p,c) \in \mathscr{S}^2$ and $u \in \tilde{\mathscr{A}}(t,T)$, where the operator $\tilde{\mathscr{L}}$ is defined in (16.41), and*
3. *the processes*

$$Y : t \mapsto \int_0^t \mathrm{e}^{-rs}\{F_p(s,Z(s),P(s),C(s))\boldsymbol{\Sigma}_{[1]}(s,P(s),C(s)) + F_c(s,Z(s),P(s),C(s))\boldsymbol{\Sigma}_{[2]}(s,P(s),C(s)\}\, d\mathbf{B}(s)$$

are a martingale, where $\boldsymbol{\Sigma}_{[i]}, i = 1,2$ *are the rows of* $\boldsymbol{\Sigma}$.

Then $F(t,z,p,c) \geq V(t,z,p,c)$ *for all* $t \in [0,T]$. *In addition, if there exists an admissible* u^o *such that*

$$(\tilde{\mathscr{L}} - r)F(t,z,p,c) + u^o(t)\,(F_z(t,z,p,c) + f(p,c)) = 0\,, \tag{16.42}$$

for all $(t,z,p,c) \in \mathscr{S}^2$, *then* $u^o = u^*$ *and the function* F *coincides with the value function* V.

The proof is similar to the proof of Theorem 1. For rigorous proofs when the process $\mathbf{X}$ is a general $\mathbb{R}^n$-valued Lévy process and the control space is a subset of $\mathbb{R}^n$, see [2].

The tolling agreement in Sect. 16.2 is studied without a maximal volume constraint; therefore we obtain the optimal control by Lemma 1 directly. If there is a maximal constraint (and possibly additional restrictions on the controls), we obtain the value of the tolling agreement by solving the HJB equation (16.40). The verification Theorem 2 shows that this solution is the correct one if it satisfies certain additional regularity conditions.

16.4.2.1 Boundary Conditions

In order to solve the HJB equations derived in this section numerically we need to find the appropriate boundary conditions. Such conditions are related to the admissible controls, the state-space dynamics and asymptotic properties of the value function.

In general, we defined the value function V in (16.37) on $\mathscr{S}^n$, where n indicates that we have an n-dimensional underlying state-space process $\mathbf{X}$. Clearly, we have the terminal condition

$$V(T,z,\mathbf{x}) = 0 \quad \forall (z,\mathbf{x}) \in [0,M] \times \mathbb{R}^n\,. \tag{16.43}$$

This follows from the definition of the value function and the fact that Z cannot have any jumps since u is bounded. For $z = M$ we have used the entire volume permitted in the contract and can only have $u = 0$ from that time on. Hence, we get the boundary condition

$$V(t,M,\mathbf{x}) = 0 \quad \forall (t,\mathbf{x}) \in [0,T] \times \mathbb{R}^n\,. \tag{16.44}$$

We observe that these conditions are independent of the state $\mathbf{x}$.

To find the boundary conditions in the $\mathbf{x}$-direction(s) is a very delicate issue and has to be done in a case by case manner. It requires a good knowledge about the

underlying price process $\mathbf{X}$, e.g. in a way such that one can find an optimal control explicitly. Theoretically, we can hope for asymptotic results in the state-space variables, but these must be applied to pin down boundary conditions in a numerical scheme for solving the HJB equation on a finite domain. Dirichlet conditions are found in [1, 2] when $\mathbf{X}$ is a one- and two-dimensional Ornstein–Uhlenbeck process, respectively. In [1] they also find Neumann conditions and discuss second-order boundary conditions. Furthermore, due to the higher dimensionality in [2], the difficulty of the problem increases and certain assumptions have to be made about the (existence and) form of the optimal control in order to find appropriate boundary conditions.

Acknowledgements We greatly acknowledge financial support from the project "Energy Markets: Modelling, Optimization and Simulation" (EMMOS), funded by the Norwegian Research Council under grant evita-205328.

References

1. Benth FE, Lempa J, Nilssen TK (2011) On the optimal exercise of swing options in electricity markets. J Energy Mark 4(4):3–28
2. Eriksson M, Lempa J, Nilssen TK (2013) Swing options in commodity markets: a multidimensional Lévy diffusion model. Submitted. Available on http://arxiv.org/abs/1302.6399
3. Fleming WH, Soner HM (2006) Controlled Markov processes and viscosity solutions, 2nd edn. Springer Science+Business Media, Inc, New York
4. Karatzas I, Shreve SE (1991) Brownian motion and stochastic calculus, 2nd edn. Springer, New York
5. Lucia J, Schwartz ES (2002) Electricity prices and power derivatives: evidences from the Nordic power exchange. Rev Derivatives Res 5(1):5–50
6. Margrabe W (1978) The value of an option to exchange one asset for another. J Financ 33(1):177–186
7. Øksendal B (2007) Stochastic differential equations, 6th edn. Springer, New York
8. Schwartz ES, Smith JE (2000) Short-term variations and long-term dynamics in commodity prices. Manag Sci 46(7):893–911
9. Wallin O (2008) Perpetuals, Malliavin calculus and stochastic control of jump diffusions with applications to finance. PhD Thesis, Department of Mathematics, University of Oslo

Part IV
Long-Term and Political Risks

Chapter 17
Risk Hedging Strategies Under Energy System and Climate Policy Uncertainties

Volker Krey and Keywan Riahi

Abstract The future development of the energy sector is rife with uncertainties. They concern virtually the entire energy chain, from resource extraction to conversion technologies, energy demand, and the stringency of future environmental policies. Investment decisions today need thus not only to be cost-effective from the present perspective, but have to take into account also the imputed future risks of above uncertainties. This chapter introduces a newly developed modeling decision framework with endogenous representation of above uncertainties. We employ modeling techniques from finance and in particular modern portfolio theory to a systems engineering model of the global energy system and implement several alternative representations of risk. We aim to identify salient characteristics of least-cost risk hedging strategies that are adapted to considerably reduce future risks and are hence robust against a wide range of future uncertainties. These lead to significant changes in response to energy system and carbon price uncertainties, in particular (i) higher short- to medium-term investments into advanced technologies, (ii) pronounced emissions reductions, and (iii) diversification of the technology portfolio. From a methodological perspective, we find that there are strong interactions and synergies between different types of uncertainties. Cost-effective risk hedging strategies thus need to take a holistic view and comprehensively account for all uncertainties jointly. With respect to costs, relatively modest risk premiums (or hedging investments) can significantly reduce the vulnerability of the energy system against the associated uncertainties. The extent of early investments, diversification, and emissions reductions, however, depends on the risk premium that decision makers are willing to pay to respond to prevailing uncertainties and remains thus one of the key policy variables.

V. Krey (✉) • K. Riahi
IIASA, Schlossplatz 1, 2361 Laxenburg, Austria
e-mail: krey@iiasa.ac.at; riahi@iiasa.ac.at

R.M. Kovacevic et al. (eds.), *Handbook of Risk Management in Energy Production and Trading*, International Series in Operations Research & Management Science 199,
DOI 10.1007/978-1-4614-9035-7_17,

17.1 Introduction

The future development of the energy system is rife with uncertainties that concern virtually the entire energy chain, from resource extraction to conversion technologies, energy demand, and the stringency of future environmental policies, in particular those addressing climate change. Ignorance with respect to the multitude of uncertainties can be very costly, due to a high share of long-lived capital stock in the energy system and the resulting long time spans that transitions require. Investment decisions today thus need not only to be cost-effective from the present perspective, but have to take into account also the imputed future risks of uncertainties. For energy models and scenario analysis this means that uncertainty needs to be represented endogenously in order to include trade-offs between "optimal" decisions based on most likely developments and risks resulting from uncertainties being resolved in one direction of the other.

Although optimization techniques incorporating uncertainty, such as stochastic programming [3, 7], have been developed several decades ago, but their application to realistic problems has only come into reach in recent years with the evolution of computational resources. Therefore, in the majority of energy studies and models, uncertainties are typically treated—if at all—by performing sensitivity analysis for a set of parameters. While it is possible through sensitivity analysis to better understand the uncertainty space and broader ranges of future developments, this method is generally not appropriate for identifying robust "hedging" strategies, including response measures and their economic implications to minimize or at least reduce exposure to unwanted risks. More advanced approaches for performing uncertainty analysis include scenario analysis, in particular if performed by multiple models (see [19]), as well as robust decision making [25] which aims at deriving robust strategies through iterative multi-scenario simulations.

Different approaches to uncertainty analysis in energy-economic models have been described and systematically categorized by Kann and Weyant [19], and more recently by Ascough et al. [1] and Peterson [40]. With a few exceptions most of the described modeling approaches fall into the category of more stylized models, which lack explicit representation of individual energy technologies. In addition, the majority of the models focuses on uncertainties related to the climate system and climate change-related damages with the climate sensitivity being the most popular parameter that is treated as uncertain (e.g., [27, 28, 54]). There are a few technology-explicit model applications, e.g., [2, 20, 23, 24, 26, 27, 34]; however, they concentrate either on a very limited number of uncertain parameters in comparison to the total number of parameters included in the models or consider very few realizations of these parameters.

This chapter does not aim at providing a comprehensive overview of existing modeling approaches documented in the literature, but instead introduces a newly developed modeling framework of the global energy system that features an endogenous representation of uncertainties. The basic structure of the model builds upon the deterministic energy engineering model MESSAGE [33, 34, 44]. We employ risk management techniques, developed for portfolio management applications

in finance (cf. [30, 41]), and incorporate a variety of different representations to measure risks into MESSAGE. Following the school started by Markowitz [30], our model in addition to the expected value of energy system costs takes into account a risk measure in the optimization procedure. On top of a variety of different risk measures we provide a set of alternative (equivalent) problem formulations (e.g., risk-constrained cost minimization, cost-constrained risk minimization). The different model formulations increase the flexibility of the modeling approach and permit us to put more or less emphasis on the tails of the uncertainty distribution or to consider different risk attitudes of decision makers (e.g., towards limiting risks below critical thresholds in contrast to exploring the effect of different levels of risk premiums[1]). The modeling approach is related to, for example, chance constraint programming [6] and robust optimization [4] in the sense that no recourse is considered. Given this close relationship of our modeling approach with robust optimization we will refer to the model version including uncertainty as robust decision making framework throughout this chapter.

In our model application all relevant cost parameters are treated as uncertain, i.e., costs concerning the entire chain of energy technologies including resource extraction , energy conversion technologies, and energy-saving measures. In addition, to account for the uncertainty of the policy intensity to climate change, the carbon price is also modeled as an uncertain parameter.

Through a series of sensitivity analysis we aim at identifying salient characteristics of least-cost hedging strategies that are able to considerably reduce future risks and are hence robust against a wide range of future uncertainties. In particular, we explore the effect of uncertainties on (i) investment decisions in the energy sector, (ii) technology deployment and diversification of the technology portfolio, and (iii) associated greenhouse gas emissions. From a methodological perspective, we are also interested in synergies of hedging strategies against technology- and policy-related uncertainties. Finally, the question of how much risk can be avoided at which cost or, alternatively, whether it is possible to come up with more robust strategies at affordable costs is central to our modeling approach.

The sequel of this chapter is structured as follows: In Sect. 17.2 we lay out the methodological basis of our modeling framework, starting with various types of problem formulations and risk measures. In the later part of the section we also address the question of how joint distributions of uncertain input parameters are generated. Section 17.3 illustrates how the methodology operates on the basis of a very simple static three-technology model. The following Sect. 17.4 introduces the central application of this study, the full global energy systems model and the main scenario input assumptions. Results of the modeling exercise including uncertainty are extensively discussed in Sect. 17.5, ranging from implications of hedging strategies for primary energy supply and emissions to diversification of the technology portfolio and reallocation of investments within the energy system. The chapter concludes with a summary of the main findings in Sect. 17.6.

[1] In this chapter the term *risk premium* refers to additional expenditures to limit exposure to unwanted risks.

17.2 Model Formulation and Solution Method

Energy systems models are frequently used to aid scenario analysis and to provide quantitative information about possible future development pathways in the energy sector. In the process of constructing scenarios many assumptions about future developments of socioeconomic, demographic, and technological change have to be made. In particular, more detailed energy systems models include a large number of technologies, which need to be represented in a parametric way. How these parameters evolve over time is subject to large uncertainties. In our model we assume all relevant cost parameters (investments, operation and maintenance costs, carbon price) to be uncertain.

The objective of this chapter goes beyond the documentation of our robust energy systems model but to provide a modeling framework that is generic enough to be used within (or together with) other similar energy-economic optimization models.[2] To capture different risk attitudes of decision makers, a number of alternative ways of measuring risks (e.g., upper mean absolute deviation, downside risk, Conditional Value-at-Risk) have been implemented. In addition, a variety of alternative (equivalent) problem formulations is provided to increase flexibility of the modeling framework. What all formulations have in common is that they describe economic trade-offs between decisions based on expected (most likely) future trends and the associated economic risks of the underlying uncertainty. The three alternative problem formulations that we consider are:

1. Minimization of a weighted sum of deterministic total system costs and a so-called risk measure as suggested by Messner et al. [34]
2. Minimization of total system costs under constrained risk
3. Cost-constrained minimization of risk (which considers a maximum risk premium that may be paid for the risk reduction)

As mentioned above, an earlier version of the MESSAGE model [34] used a similar robust modeling approach as under point (1), but concentrated on a very limited number of technologies as well as uncertain parameters. Further applications of this type of approach include uncertain import prices of fossil energy carriers [23] as well as uncertain increasing returns to scale [12, 13]. A more general discussion of the methodology can be found in [10, 31].

We shall next describe the three different problem formulations (Sect. 17.2.1), followed by a discussion of the employed risk measures (Sect. 17.2.2).

[2] The model's source code is made available at `http://www.iiasa.ac.at/web/home/research/researchPrograms/Energy/Robust-MESSAGE.en.html`.

17.2.1 Generic Problem Formulations

To start with, let us consider the system of equations of a generic intertemporal deterministic linear programming (LP) minimization problem:

$$\begin{aligned} &\min \sum_t \mathbf{c}_t^T \mathbf{x}_t \cdot \delta(t)\,, \\ &\mathbf{A}\mathbf{x} = \mathbf{b}\,, \\ &\mathbf{x} \geq 0\,. \end{aligned} \tag{17.1}$$

Here $t = 1 \ldots T$ is the time-period index, $\mathbf{c}_t$ is the cost coefficient vector of the objective function in period t, and $\mathbf{x}_t$ are the corresponding decision variables in period t, with $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_T)$ referring to the vector of decision variables for all periods $t = 1 \ldots T$. The set of constraints is defined by the matrix $\mathbf{A}$ and the vector of the right hand sides $\mathbf{b}$. The last term in the objective function, $\delta(t)$, is the discount factor. In the following we will refer to the deterministic objective function in the first line of (17.1) as $F^{\text{det}}(\mathbf{x})$ and to the cost-optimal solution of the deterministic problem as $\mathbf{x}^{\star}_{\text{det}}$.

Based on the above defined deterministic model, we now describe a set of model formulations of our robust decision-making framework that include an endogenous representation of risk that result from future uncertainties. For this purpose a risk measure (or risk functional), denoted by $R(\mathbf{x})$, is introduced (see also Sect. 17.2.2). Note in particular that the risk measure is an endogenous function of the decision variables $\mathbf{x}$, thus depending on, e.g., investment decisions driving technology deployment.

We implemented three alternative problem formulations. From a methodological perspective, the three formulations represent different ways to combine the deterministic objective function, i.e., total system costs $F^{\text{det}}(\mathbf{x})$ and the risk measure $R(\mathbf{x})$. Conceptually, the formulations allow for different policy perspectives, where depending on the context, it may be preferable either to control costs or to control risks or to define the risk aversion of the decision maker. The formulations are equivalent to each other and can even be combined in a synergistic way as will be discussed later in this section. The three formulations are discussed in turn:

1. Minimization of a weighted sum of deterministic total system costs $F^{det}(\mathbf{x})$ and the risk measure $R(\mathbf{x})$:

$$F^{sto}(\mathbf{x}) = F^{det}(\mathbf{x}) + \rho \cdot R(\mathbf{x})\,.$$

   A simple linear combination of deterministic total system costs and risk measure allows to explore the impact of risk on the optimal solution. In this formulation there is no clear focus on either total system costs or risk measure, but the relative weight of the two can be adjusted with the help of the factor ρ, an indicator for the risk aversion of the decision maker (cf. [34]).

2. Minimization of deterministic total system costs $F^{\text{det}}(\mathbf{x})$ under constrained risk measure $R(\mathbf{x})$:

$$\begin{aligned} &\min F^{\text{det}}(\mathbf{x}) \\ &\text{s.t. } R(\mathbf{x}) \leq R^{\max}. \end{aligned}$$

   In this formulation the emphasis is on risk reduction. From the perspective of a decision maker the focus is to reduce the risk by constraining the risk measure $R(\mathbf{x})$ to a maximal permissible value. The cost minimization then guarantees that a solution is on the efficient frontier, i.e., the risk level $R^{\max}$ is achieved at lowest possible costs. The formulation is particularly important if critical thresholds for risks can be identified. A practical example would be local water supply management, which typically focuses on the optimal allocation of resources in order to reduce the risk of supply shortages below certain levels [9].
3. Minimization of risk measure $R(\mathbf{x})$ under constrained expected total system costs $F^{det}(\mathbf{x})$:

$$\begin{aligned} &\min R(\mathbf{x}) \\ &\text{s.t. } F^{\text{det}}(\mathbf{x}) \leq (1+f) \cdot F^{\text{det}}(\mathbf{x}^{\star}_{\text{det}}). \end{aligned}$$

   In contrast to the previously presented formulation here the focus is to control the risk premium, i.e., the "additional" costs for reducing risk. These "hedging expenditures" are limited to a fraction f of the total costs in absence of uncertainty [the costs of the deterministic solution $F^{\text{det}\star} = F^{\text{det}}(\mathbf{x}^{\star}_{\text{det}})$]. The objective of this problem formulation is then to minimize the resulting risk (given budgetary constraints for the risk premium). This formulation gains importance if risk thresholds cannot be identified or when the problem is characterized by large complexity. Particularly in the latter case a sensitivity analysis for different risk premiums can provide important policy insights, enhancing the understanding of the magnitude of risk that can be reduced at specific costs. Future energy projections, as explored here, are a typical example of a complex problem with often counteracting uncertainties (e.g., the increasing likelihood of high carbon prices and the uncertainty of future cost improvements for new zero-emission technologies are pulling the solution into opposite directions). We will come back to this and primarily use this formulation for illustrating our results.[3]

The three formulations are—with appropriately chosen parameters $R^{\max}$, f, and ρ—equivalent to each other and their combined use can help to understand different aspects of risk hedging strategies. Cost-constrained risk minimization (3) has the

[3] From a technical perspective, the latter two formulations have the advantage that the relative orders of magnitude of total system costs $F^{\text{det}}(\mathbf{x})$ and risk measure $R(\mathbf{x})$ can be allowed to be very different. This happens, e.g., in the case of a linear cost function and a quadratic risk measure (e.g., semi-variance, see Sect. 17.2.2). Apparently this situation causes problems in the third case, because one of the two terms then dominates the other one for $\rho \sim 1$; thus the magnitude of ρ has to be chosen individually for each risk measure used.

advantage of being comparatively easy to interpret and to communicate, since the hedging cost f denotes in essence a risk or insurance premium. However, a drawback is that the risk measure defines the objective function, leading to shadow prices of, e.g., energy carriers or carbon emissions which are not comparable with those of the deterministic model. Therefore, combining cost-constrained risk minimization and risk-constrained cost minimization allows a complementary view on the problem by determining the risk level corresponding to a certain risk premium f first and then running the model again in the risk-constrained formulation that allows obtaining "ordinary" shadow prices, like marginal carbon abatement costs. Another interesting aspect is that the linear combination $F^{\text{sto}}(\mathbf{x})$ makes the trade-off between expected costs and risk explicit in the objective function. With an appropriately chosen ρ this allows deriving an "optimal" level of risk—an admittedly hypothetical concept, but with clear methodological advantages for exploring cost-benefit analysis of, e.g., climate change.

17.2.2 Risk Measures

As stated above, we are interested in a generic framework for risk assessment rather than a particular type of formulation. Therefore, in addition to the different problem formulations presented above, we have implemented a number of—partly well-known—risk measures. Similar to the alternative problem formulations the choice of an appropriate risk measure depends on the risk attitude of a decision maker as well as on the specific characteristics of the problem under consideration. For instance, if *low-probability high-impact* events are of particular interest, a risk measure that focuses on the tail of the distribution (e.g., semi-variance, β-CVaR) is more appropriate than one that equally weights all positive deviations from the expected value (e.g., upper mean absolute deviation).

In the following, $\mathbf{c}_t(\omega)$ and $\bar{\mathbf{c}}_t$ are the uncertain cost parameters in period t and their expected values, respectively. For practical reasons, i.e., to ensure solvability of the problem, we restrict ourselves to measures that are implementable in linear and quadratic programming:

1. The *upper mean absolute deviation* is a linear risk measure originally used by Messner et al. [34] and is defined in the following way:

$$R(\mathbf{x},\omega) = \sum_t \max\left\{0, [\mathbf{c}_t(\omega) - \bar{\mathbf{c}}_t]^T \mathbf{x}\right\} \cdot \delta(t). \tag{17.2}$$

   Its expected value $R(\mathbf{x}) = \mathbf{E}_\omega R(\mathbf{x},\omega)$ corresponds the expected underestimation of total system costs which is used to measure the economic risk associated with a strategy $\mathbf{x}$.
2. *Conditional Value-at-Risk (CVaR)*, also referred to as *expected shortfall*, is related to the above defined upper mean absolute deviation but only takes the worst $(1-\beta)$ fraction of outcomes into consideration, where β typically is

chosen to be 0.90, 0.95, or 0.99. Our model implementation is based on the work of [47] which, in addition to the risk measure's definition in (17.2), requires the following two equations:

$$\mathrm{CVaR}(\mathbf{x}) = \mathbf{E}_\omega \sum_t \left\{ \alpha_t (1-\beta)^{-1} \cdot z_t(\omega) \cdot \delta_t \right\},$$
$$R(\mathbf{x}, \omega, t) = \alpha_t - z_t(\omega).$$

Here α_t serves as a proxy for the so-called Value-at-Risk in period t and $z_t(\omega)$ is an auxiliary variable. More background information and details on the implementation can be found in [39, 47].

3. *Semi-variance* or *downside risk*:

$$R(\mathbf{x}, \omega) = \sum_t \left(\max \left\{ 0, [\mathbf{c}_t(\omega) - \bar{\mathbf{c}}_t]^T \mathbf{x} \right\} \right)^2 \cdot \delta(t). \tag{17.3}$$

The expected value of this quadratic risk measure corresponds to a semi-variance, i.e., only positive deviations contribute to it, but in contrast to (17.2) quadratically.

4. Linear-Quadratic Risk-Benefit Function

$$R(\mathbf{x}, \omega) = \sum_t \left\{ \gamma \cdot \left(\max \left\{ 0, [\mathbf{c}_t(\omega) - \bar{\mathbf{c}}_t]^T \right\} \mathbf{x} \right)^2 - \min \left\{ 0, [\mathbf{c}_t(\omega) - \bar{\mathbf{c}}_t]^T \right\} \mathbf{x} \right\} \cdot \delta(t). \tag{17.4}$$

Following the arguments by Grübler and Gritsevskyi [13], positive deviations from the expected costs $\bar{\mathbf{c}}_t$ contribute quadratically to the risk-benefit function, whereas negative contributions, i.e., opportunities or benefits, contribute linearly. This formulation reflects that underestimating costs is penalized more heavily in competitive markets than overestimation. From a company's perspective the latter might result in lower profits whereas the former can result in bankruptcy.

The first three risk measures are well known and widely used in the finance and risk management literature (cf. [41]), even though they have rarely been applied to energy-economic problems. To a large extent the alternative risk measures represent different risk attitudes towards either the average risk or extreme tail events. For example, the quadratic risk measures (17.3) and (17.4) put much more emphasis on the "low-probability high-cost" events in the tails of the distributions compared to the upper mean absolute deviation in (17.2), which averages risks over the whole range of excess costs. Similarly, the case of CVaR represents also risk aversion towards the tails, since per definition only the worst $(1-\beta)$ fraction of realizations is taken into account. A specific characteristic of the linear-quadratic risk measure in (17.4) is that it takes opportunities into consideration. However, if distributions are sufficiently wide, the quadratic part typically dominates the linear one such that results turn out to be very close to that of the semi-variance (17.3).

17.2.3 Numerical Computation

We assume distribution functions for the uncertain cost parameters (see Sect. 17.4.2). For practical implementation purposes we draw a finite number of N samples from these joint distributions. By doing so we obtain a numerical estimate of the risk measure $R(\mathbf{x})$ as defined in (17.5) below

$$R(\mathbf{x}) \to R^N(\mathbf{x}) = \frac{1}{N} \sum_{s=1}^{N} R(\mathbf{x}, \omega^s) \,. \tag{17.5}$$

Given the sampling procedure the approach can be referred to as distribution-free. The quality of the solution critically depends on the sample size. In other words, N needs to be selected large enough, so that it approximates a solution with $N \to \infty$. We assess the minimum sample size through experiments, where N is increased until solutions converge and outcomes do not show any qualitative difference. A detailed description of the convergence criterion as well as the sampling techniques can be found in Appendix 17.8.

We also account for correlations between uncertain parameters (see Sect. 17.4.2). Depending on the sampling algorithm, we use either the so-called copulae (in case of random sampling) or the algorithm suggested by Iman and Conover [18] [in case of latin hypercube sampling (LHS)] to induce correlation among uncertain parameters.[4]

17.3 Three-Technology Model

This section presents results from a very simple and idealized model with the aim to illustrate, from a conceptual point of view, how in our modeling approach uncertainties and risks affect the decision-making process.

The simple model consists of just three variables x, y, and z. In the energy context these three variables can be thought of as different types of power plants (e.g., natural gas, coal, nuclear) with expected values for electricity generation costs $\bar{c}_x$, $\bar{c}_y$, and $\bar{c}_z$ respectively. Uncertainties with respect to costs, characterized by the variance σ_i^2, are assumed to differ across the three technologies (see Table 17.1). The three power plants need to supply an electricity demand d. To further the complication from inter-temporal effects, the model is chosen to be static.

Assuming that there is no uncertainty, the problem formulation is reduced to the simple deterministic objective function $F^{det}(x, y, z) = \bar{c}_x x + \bar{c}_y y + \bar{c}_z z$. Due to the employed cost minimization the model suggests to deploy only the cheapest technology. Even if cost differences between technologies would be very minimal, the

[4] For k parameters, as a result of both procedures, we obtain samples on the k-dimensional unit cube $[0, 1]^k$ that can subsequently be transformed into arbitrary distributions with the corresponding quantile functions. Samples are generated with the help of the graphics and statistics software R [43], making use of several add-on packages [5, 53].

winner always takes it all. In our example this is the natural gas power plant in the upper left corner of Fig. 17.1a which illustrates the total system costs of the deterministic objective function. The lines in Fig. 17.1a denote cost meridians of identical system costs for a combination of nuclear and natural gas shares in total demand. Naturally the lower the contribution of natural gas becomes, the more the other technologies need to be deployed, leading consequently also to an increase in the system costs. The third variable z (coal) is not shown explicitly, because its contribution corresponds just to the gap between the other two technologies' supplies and the demand, that is, $z = d - x - y$.[5] A major drawback of the deterministic solution is that it always features the least-cost option supplying the total demand. Small changes in cost assumptions may thus dominate the results, with switching between one extreme to another. This behavior is known as penny switching or knife edge effect of linear programming.

Table 17.1 Technology specifications in the 3-technology model

Variable	Technology	$\bar{c}_i$ [ct/kWh]	σ_i [ct/kWh]
x	Nuclear power plant	4.5	1.0
y	Natural gas power plant	3.5	1.5
z	Coal power plant	4.0	1.25

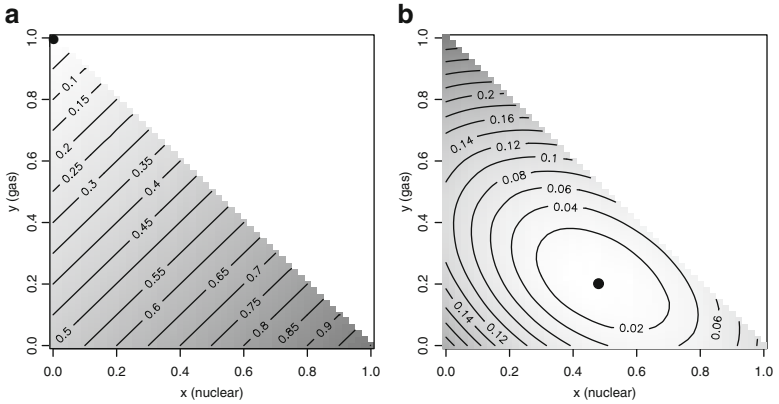

Fig. 17.1 Illustration of (**a**) total system cost $F^{det}(x,y,z)$ and (**b**) risk measure $R(x,y,z)$ as a function of technology activities

Panel (b) in Fig. 17.1 shows the values of the risk measure $R(x,y,z)$ if we assume that the costs are uncertain.[6] As illustrated by the graph, also $R(x,y,z)$ is dependent on the share of the individual technologies. From a pure risk perspective, however, natural gas has for example become much less attractive due to its high-cost

[5] In the numerical example without loss of generality the demand is set to $d = 1$.

[6] We implemented the risk measure as defined by the upper mean absolute deviation as defined in (17.2).

uncertainty (see standard deviations σ_i in Table 17.1). Most importantly, the risk measure in Fig. 17.1b features a minimum—again indicated by the black dot—that corresponds to a diversified technology portfolio with contributions of all the three power plant types. In that sense, the model confirms the well-known rule that in the case of uncertainty it is advisable "not to put all the eggs into one basket." To which extent diversification or hedging can help reduce the imputed risk of uncertainty depends on the problem at hand and the model formulations presented further above (Sect. 17.2.1). Regardless of the problem formulation, robust response strategies need to consider both panels of Fig. 17.1 and the economic trade-off between the expected costs F^{det} as well as the imputed risk R.

It needs to be emphasized that diversification can only help as a hedging strategy, if the costs of the technologies are not perfectly correlated. In the case of highly correlated electricity generation costs it is for example preferable to mostly choose the technology with less volatile costs to minimize the risk measure. This behavior is illustrated in Fig. 17.2 where the correlation coefficient ρ_{yz} between gas- and coal-fired electricity generation is varied between 0 and 1. With increasing correlation the share of natural gas at minimal risk is reduced to zero in comparison to more than 20 % in the case without correlation. Note also that the share of the uncorrelated nuclear plant is increasing in response to the cost dependency between coal and gas. This has important practical implications for the bigger global energy systems model, presented in the next section.

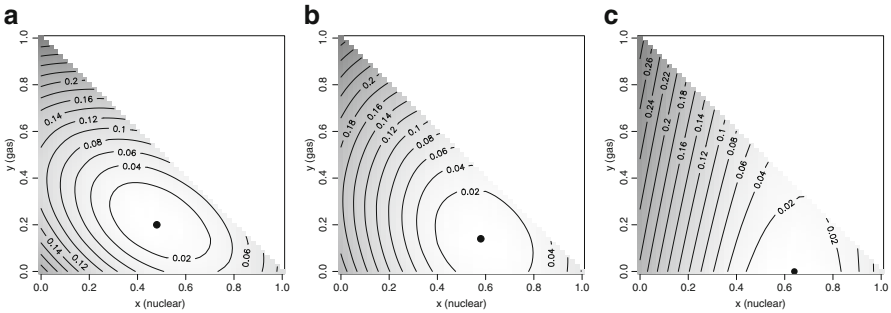

Fig. 17.2 Influence of correlation on diversification of the technology portfolio for correlations (**a**) $\rho_{yz} = 0$, (**b**) $\rho_{yz} = 0.5$, and (**c**) $\rho_{yz} = 1$

The introduction of an uncertain carbon price would influence the cost distributions of the three technologies in our example very differently. The cost distribution of nuclear as a carbon-free technology would not change at all whereas gas- and in particular coal-fired power generation would be penalized by shifting the distributions towards higher costs. As a result the optimal share of nuclear in a risk-minimal portfolio would increase whereas coal's contribution would decrease.

The combination of expected costs F^{det} and risk measure R in our modeling framework is different for the three formulations described in Sect. 17.2.1. The mechanism of the three formulations is illustrated graphically in Fig. 17.3. As in

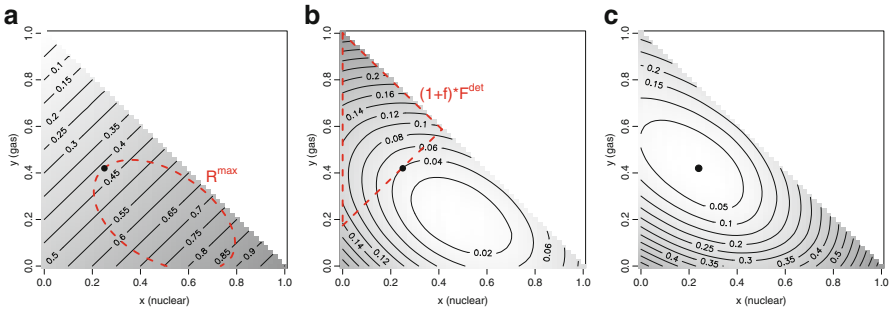

Fig. 17.3 Illustration of alternative model formulations: (**a**) minimization of a linear combination of total system costs and risk measure, (**b**) minimization of total system costs under a risk constraint, (**c**) minimization of the risk measure under a cost constraint

the previous figures the corresponding minima are again indicated by the black dots in the graphs.

The left-hand panel of Fig. 17.3 represents risk-constrained cost minimization, i.e., the objective function is identical to that of the deterministic model. In addition, the risk constraint $R^{\max}$—indicated by the red dashed line—is projected onto the surface of the cost function. The resulting cost minimum corresponds to a technology portfolio on the edge of the $R^{\max}$ surface with the lowest possible objective function value. The risk constraint $R^{\max}$ excludes the optimal deterministic solution featuring the extreme of 100 % gas electricity generation. Consequently, all three power plants contribute to electricity generation.

Figure 17.3b denotes the result of a cost-constraint risk minimization. Panel (b) thus shows the risk measure R with the cost constraint $(1+f)\cdot F^{\det}$ projected onto its surface as the red dashed line. The resulting risk minimum corresponds to a technology portfolio on the edge of the surface denoted by the cost constraint. Synonymous to the above implementation, the cost constraint $(1+f)\cdot F^{\det}$ excludes the minimum risk solution; thus, e.g., natural gas shares are higher than in a case with pure risk minimization.

The third graph on the right side of Fig. 17.3 corresponds to a linear combination of expected costs $F^{\det\star}$ and risk measure R, also featuring an optimum with a diversified technology portfolio. Moreover, the three panels of Fig. 17.3a–c illustrate the equivalence between the three alternative problem formulations. As a result the optimum is identical in the three formulations if the parameters ρ, R^{max}, and f are chosen accordingly. Although mathematically equivalent, from a conceptual perspective, the three formulations allow for different policy perspectives, where depending on the context it may be preferable either to control risk (left panel) or costs (middle panel) or to define specific risk aversion of the decision maker (right panel).

We move next to the more complex global energy systems model and the implementation of uncertainty into a "real world problem."

17.4 Global Energy Systems Model

In order to explore the impact of cost uncertainties on optimal transitions within the global energy system and the resulting development pathways, we employ the above methodologies within a systems engineering model. On the one hand a relatively simple or stylized model structure is a precondition to keep the model transparent and the results interpretable. On the other hand, modeling approaches considering uncertainties are computationally more demanding than deterministic models and therefore put limitations on the number of parameters treated as uncertain. Compared to other energy-economic models, our framework is thus of intermediate complexity, characterized by a relatively comprehensive representation of energy technologies within a single world-region.

17.4.1 Model Structure and Scenario Assumptions

Figure 17.4 provides a schematic illustration of the model's reference energy system (RES). The RES is designed to cover a large number of possible energy supply chains, from primary energy extraction to a range of energy conversion technologies, and the transmission and distribution of final energy carries to three aggregated demand sectors.

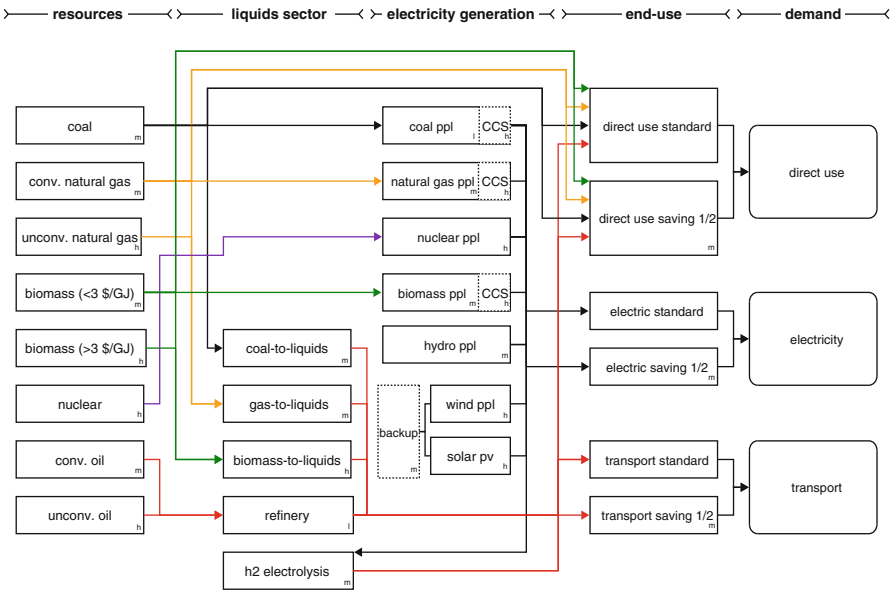

Fig. 17.4 Schematic illustration of the one-regional global model. The *small letters* in the boxes indicate the uncertainty level assigned to the cost development of individual technologies (l=low, m=medium, h=high); see Sect. 17.4.2 for details

The model includes various fossil, nuclear, and renewable energy resources along with estimates of the associated potentials and extraction costs (see left-hand side of Fig. 17.4). The conversion sector consists of nine electricity generation technologies with the possibility of carbon capture and storage (CCS) from fossil- and biomass-based power generation. In addition the model considers five alternative technologies to supply liquid fuels, including refineries for oil products as well as coal, gas, and biomass to liquid options and hydrogen.[7] Transportation and distribution costs of particularly grid-bound energy carriers, such as electricity and natural gas, are considered as well; however, they are not shown explicitly in Fig. 17.4.

The demand side is more aggregated than the supply side and distinguishes three demand categories for electric, non-electric (direct use), and transportation fuel demand. In the demand sectors we currently do not model individual appliances, but use energy conservation cost curves. The reference demand is based on the B2 scenario [45] from the IPCC Special Report on Emissions Scenarios (SRES) [37]. The parameterization of the conservation cost curves and the corresponding energy conservation potentials is derived from an updated version of the B2 scenario using the IIASA Integrated Assessment modeling framework [16, 46] including the detailed 11-regional MESSAGE-MACRO model [32, 46].

Large-scale energy systems models typically include a number of additional restrictions or constraints in order to (i) avoid typical penny switching effects of linear programming approaches and (ii) guide the model into a "realistic" direction. Such external *model guidance* often mimics market penetration limitations of specific technologies and remains however to some extent arbitrary with limited empirical basis. In contrast to this practice, we do only include restrictions that have a physical or technical motivation, like, e.g., resource availability, renewable potentials, maximum share of intermittent electricity generation (25 % of final demand, otherwise additional backup capacity needed), or baseload constraints (60 % of final electricity demand). An exception is the use of solid fuels in the end-use sectors which is limited to the level of the B2 baseline [16] in order to mimic noneconomic considerations and inconvenience of solid fuel consumption at the consumer level.

The main underlying assumptions with respect to scenario drivers, such as economic growth, population, or technological change, build upon the B2 SRES storyline and the most recent quantitative update summarized in [46]. The B2 scenario is a middle of the road "dynamics as usual" scenario, which combines intermediate population and economic growth with modest but balanced technology improvements. The balanced and intermediate characteristic of the scenario makes it ideal for defining the expected values in our analysis. Hence, the parameterization of the technologies, including the evolution of expected costs over time, and energy demand stem mainly from the B2 scenario. In addition, we build upon the review of technological change in the scenario literature performed by Nakicenovic and Riahi [36], which analyzes future distributions of costs of three ensembles of scenarios for

[7] Hydrogen production is limited to electrolysis. While this permits that hydrogen is produced from all primary fuels, we did not consider other technologies such as natural gas steam reforming in order to keep the number of technologies as small as possible for the computationally expensive optimization procedure.

the World Energy Council [38], the IPCC SRES, and the IPCC Third Assessment Report [35]. We use their analysis in order to define broader uncertainty ranges of future technology costs (see Appendix 17.7 for further details on cost and demand assumptions).

As noted earlier, all our scenarios consider a modest price for carbon, approximating that there will be some (but not drastic) efforts to limit climate change over the long term. The expected carbon price [and its probability density function (PDF)] was derived from a subset of 58 stabilization scenarios from the IPCC-AR4 scenario database [15] with CO_2-equivalent concentration targets of 650 ppmv and higher. A lognormal distribution was fitted to this sample of carbon prices in the year 2100 (see Appendix 17.7.3) and the resulting values were then propagated backwards to 2010 with the discount rate. This procedure results in a moderate expected carbon price of 4.6 US$/tC in 2010 which grows with the model's discount rate of 5 % over time. The carbon price in later periods is, e.g., 12.2 US$/tC in 2030, 32.5 US$/tC in 2050, and 372 US$/tC in 2100.

17.4.2 Uncertain Parameters

We need to define uncertainties with appropriate assumptions about the shape, variance, and correlation between different uncertain cost parameters.

Unlike in natural sciences, controlled experiments are unfortunately not available to define the shape of probability distribution functions of future technology costs. There is though some limited empirical evidence from time-series analysis of historical technology data (e.g., nuclear power generation [21]), which suggest the use of lognormal or similar distributions (e.g., Gamma), characterized by a tail on the upper side and a cutoff on the lower part of the costs. Similar to earlier uncertainty analysis by for example [12] we thus apply lognormal distributions[8] to all uncertain cost parameters where the expected values correspond to the deterministic costs.[9]

Depending on which part of the technology costs are dominant, we either model (capacity-related) investment or (activity-related) variable operation and maintenance costs as uncertain parameters. For example, the variable costs of fossil fuel extraction as well as biomass production and nuclear fuel costs are modeled as uncertain parameters. For power generation, including CCS and liquid fuel production, investment costs typically dominate levelized production costs (excl. fuel costs) and are therefore modeled as uncertain. In contrast, for energy-saving options which operate on the level of conservation cost curves the cost of the activity is assumed to be uncertain.

We assume also that the cost uncertainty of technologies is increasing over time. For this purpose we use the future cost distributions from [36] as a proxy to

[8] The choice of lognormal distributions for the costs corresponds to normally distributed growth rates of these.

[9] In total 32 cost parameters are treated as uncertain; 31 of which are technology-related and the 32nd, the carbon price, is policy-related.

define the cost variance for individual technologies at the end of the time horizon.[10] Figure 17.5 gives a schematic illustration of the increasing uncertainty over time (technical details of the implementation are given in Appendix 17.7.2). Apart from the evolution of the expected value, Fig. 17.5 also shows the 25th and 75th percentile (shaded area) as well as the 1st and 99th percentile of the distribution function. Given our approach, costs generally change more rapidly in the first decades and then converge towards their long-term value in the second half of the century. Perhaps, most importantly our implementation of uncertainty considers not only the possibility of dropping costs, but also a long tail with small likelihoods of increasing costs as observed during the recent years.

We distinguish three broader uncertainty categories for individual technologies: low uncertainty ($\sigma_{low} = 0.15$), medium uncertainty ($\sigma_{med} = 0.3$), and high uncertainty ($\sigma_{high} = 0.6$). Following [36], mature technologies with only small cost reduction potentials (e.g., coal power plant, oil refinery) exhibit low variance and are thus assigned to the lowest uncertainty category. Readily available technologies that have been deployed on a large scale, but are still expected to have significant cost reduction potential (e.g., gas combined cycle power plant) are grouped in the medium uncertainty category. In addition, also mature technologies where heterogeneity and local context add to the cost uncertainty (e.g., hydro-power, fossil fuel extraction, energy-saving measures) are assigned to the intermediate category. Finally, advanced technologies with potential for high-cost reductions typically show a wide spread of cost assumptions across different scenarios (e.g., solar photovoltaics). These technologies are grouped with those that are affected by additional risks (e.g., acceptance problems of nuclear power) in the category with the

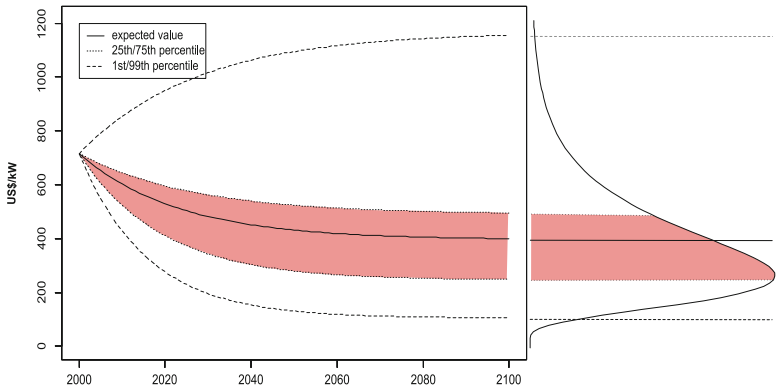

Fig. 17.5 Illustration of cost interpolation procedure (*left*) and cost distribution in 2100 (*right*) for investment costs of a natural gas combined cycle power plant

[10] The cost distributions are defined for the final year of the model's time horizon, i.e., 2100. To derive cost paths for the model's full time horizon (2000–2100) we exponentially interpolate between the base year value in 2000 and the random parameter values in 2100.

highest uncertainty. The resulting classification of individual technologies is denoted in Fig. 17.4 as small letters in the technology boxes (l = low, m = medium, and h = high).

Many energy technologies share similar components (e.g., gas turbines in natural gas combined cycle and integrated coal gasification combined cycle power plants) or rely on identical technologies (e.g., exploration and drilling techniques for conventional oil and gas reserves). Hence, the future development of their costs and the associated diffusion process are not independent of each other [19, p. 36]. As illustrated earlier, this has major implications for hedging or diversification strategies. We thus explicitly include correlations among different uncertain parameters in our modeling framework. For some of the technologies data was available from specific technology component analysis (electricity generation technologies [22]) to derive the correlation coefficients. For others we rely on expert opinions. Similar to the uncertainty categories, we distinguish several levels of correlation, i.e., perfect ($\rho_{ij} = 1$), high ($\rho_{ij} = 0.7$), medium ($\rho_{ij} = 0.35$), and uncorrelated ($\rho_{ij} = 0$). A more detailed description of this procedure can be found in Appendix 17.7.2.

17.5 Results

This section presents results of the robust decision-making framework. We analyze a series of model runs and compare scenario outcomes with and without consideration of uncertainties. By doing so we explore the main characteristics of least-cost risk hedging strategies and the extent to which the imputed risk of future uncertainties can be reduced and at what costs. We are hence particularly interested in the relationship between the "risk premium" and avoided risk, including implications for the tail of the cost distribution. In addition, we analyze responses of the energy system with respect to the technology portfolio and investment patterns, as well as consequences for carbon emissions under uncertainty.

17.5.1 Energy System Costs

Our modeling approach considers the uncertainty of future technology costs as well as of the carbon price. On an aggregated level these uncertainties translate into distinct future distributions for the total energy system costs, which critically depend on investment decisions and the type of technologies that become adopted. Comparisons of probability distributions of different energy deployment pathways are thus critically important for understanding the implied risk of different strategies. It is worthwhile noting that parameters other than costs cannot be included into this framework as these would require the introduction of recourse conditional on the actual realization of these parameters.

A comparison of the PDF of the total energy system costs of two alternative development pathways, with and without considering uncertainty, is illustrated in Fig. 17.6a–d. The individual panels show discounted system costs over the century as well as for individual points in time (2030, 2050, and 2100).[11] The deterministic solution's PDFs are shown in black whereas the PDFs of our robust decision-making framework, assuming a risk premium of 1 %, are shown in red.

In the deterministic model uncertainty is ignored in the decision process and system costs are minimized based on expected values of all input parameters. The resulting distribution of energy system costs is relatively wide, including high-cost tails with comparatively higher probability of very costly outcomes. In contrast, the PDFs resulting from the robust solution with a risk premium of 1 % are more centered around their expected value than the deterministic ones. While this tendency increases over time as uncertainties grow towards the end of the century (see Sect. 17.4.2), the distributions clearly show increasing confidence that future system costs will be closer to the expected values.

By construction the expected value of system costs, indicated by the red vertical lines in all subfigures, is shifted by 1 % to the right indicating the additional costs (or hedging investments) that were spent to reduce uncertainties and their imputed risk. Consequences of this investment are visible in the tails of the cost distributions, represented by the 99th percentiles in Fig. 17.6 which are shifted towards the expected value, implying significantly reduced risk of unfavorable outcomes with extreme costs. Remarkably, the shaving of the tales has occurred even though we employ *upper mean absolute deviation* as our default risk measure which puts uniform emphasis on all parts of the distribution exceeding the expected value and not only the tails (see also Sect. 17.2.2). This development needs thus to be seen as an endogenous response driven by the characteristics of energy system uncertainties.

This effect is most pronounced in the long term. For example, by 2100 total costs in the deterministic case are twice as high as the expected value at the 99th percentile of the distribution (see Fig. 17.6d). With additional hedging investments of just 1 % the 99th percentile's value is reduced by about 60 % in relation to the expected value. Or in other words, hedging investments of just about 100 billion US$ have acted as a leverage to reduce the risk at the 99th percentile by more than 6 trillion US$. This behavior nicely illustrates the trade-off between expected value costs and risk of severely underestimating future costs.

As a result of discounting with 5 % the PDF of total discounted system costs (Fig. 17.6a) is dominated by the relatively narrow near-term distributions and exhibits a shape which is similar to the PDFs of 2030 and 2050 but much narrower than the 2100 PDF which is suppressed by a factor of ~ 80 in comparison with the 2010 PDF due to discounting.

[11] The energy system cost PDFs are generated by propagating the joint input distributions through the model, given a fixed solution. Technically speaking, the $N = 20,000$ realizations of uncertain cost parameters are multiplied with the deterministic and robust solution vectors, respectively, thus obtaining 20,000 objective function values. A kernel density estimate is then used to generate the PDFs in Fig. 17.6.

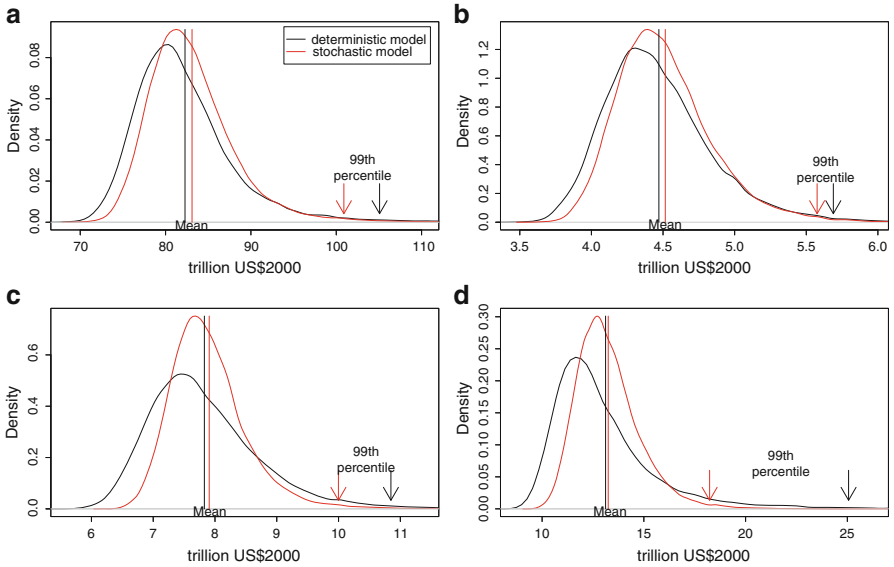

Fig. 17.6 Probability density functions of future energy systems costs. Deterministic model without uncertainty (*black*) and robust decision-making framework results assuming a risk premium of 1 % (*red*): (**a**) total discounted system costs (2010–2100), (**b**) system costs in 2030, (**c**) in 2050, and (**d**) in 2100

The above calculations assume a risk or insurance premium of 1 %, hence limiting the additional hedging investments to 1 % of total systems costs of the deterministic case. It needs to be emphasized, however, that in the "real world" the risk premium is dependent on the risk attitude of the decision maker and is therefore a policy variable. A quantitative analysis of the trade-off between the costs of hedging (i.e., the risk premium f) and the resulting benefits in terms of reduced risk is nevertheless central for providing guidelines and to understand the order of magnitude of this trade-off. Figure 17.7 thus shows the relationship between increasing risk premium and the resulting benefits in terms of reduced risks through changes in the distribution of future system costs. We specifically focus on the 80th to the 99th percentile of the cost distribution, with the solid lines showing how these quantiles change as a function of the risk premium f in the case of *upper mean absolute deviation* and the dashed lines showing the relationship for the case of *semi-variance*. As clearly illustrated by Fig. 17.7, increasing willingness to invest into the risk premium is generally resulting in reduced risk of high energy system costs. The marginal benefits of hedging investments, however, decrease with increasing risk premium f—a clear indication of decreasing returns of scale at high premiums. How the different quantiles perform as a function of the risk premium also depends on the employed risk measure. The linear risk measure (upper mean absolute deviation) reduces the 80th and 90th percentiles stronger than the quadratic one (semi-variance) whereas the situation is the opposite for the 95th and 99th percentiles. As expected, the quadratic

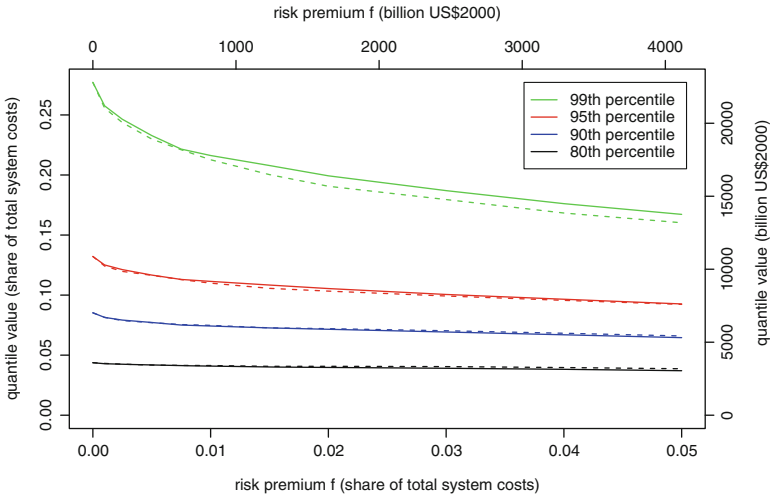

Fig. 17.7 Trade-off between expected total system costs and risk for linear (*solid lines*) and quadratic risk measure (*dashed lines*)

risk measure puts a higher emphasise on reducing the impact of the extreme tail events and therefore the 99th percentile is reduced strongest, followed by the 95th, 90th, and the 80th percentiles in relative terms.

Regarding the choice of the risk premium, Fig. 17.7 clearly shows that 1 % is in the range where the marginal returns of the hedging investments become relatively saturated—as indicated by the flattening curves for most quantiles in the figure. We therefore select in the sequel a risk premium of $f = 1\,\%$ as our central case, but will continue to show the sensitivity of the results for alternative risk premiums if necessary.

17.5.2 Primary Energy Supply

The development of total primary energy supply (TPES), resulting from different assumptions about uncertainty, is shown in Fig. 17.8. Panel (a) displays the development in the deterministic case without any uncertainty, while panel (b)–(d) illustrate the impact of considering either only technology-related uncertainties (b), only carbon price uncertainties (c), and finally taking both carbon and technology uncertainties into account simultaneously (d).

Comparing panels (a) and (b) of Fig. 17.8 reveals the main responses of the energy system due to technology uncertainty.[12] An important characteristic of the deterministic energy system is that the lack of uncertainty results in the sequential deployment of first the cheap options until they are exhausted, followed later

[12] It is important to recall that both panels consider a modest expected value carbon price.

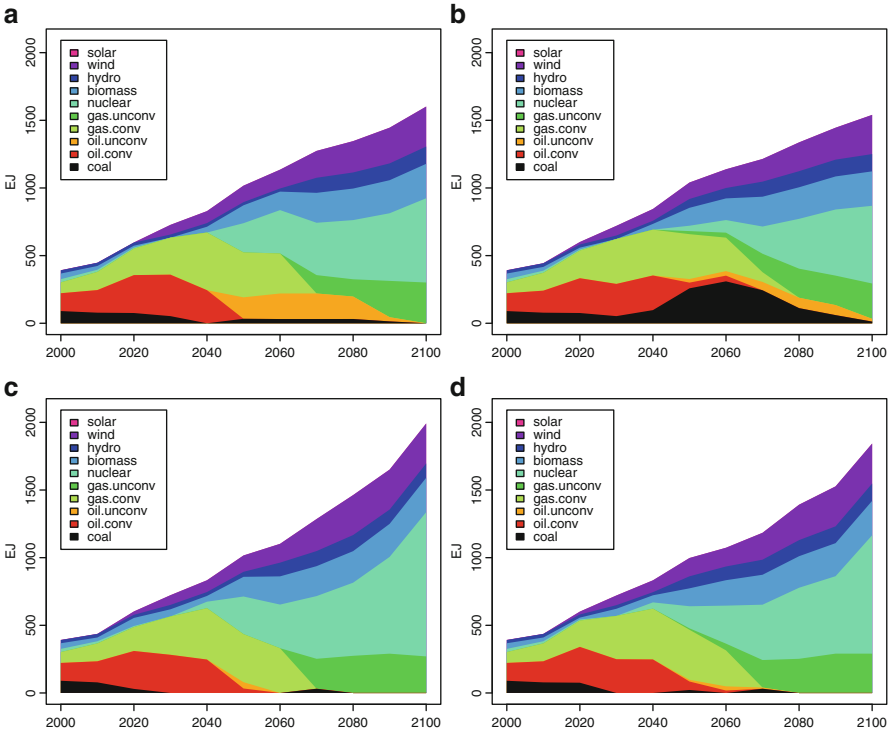

Fig. 17.8 TPES in the (**a**) deterministic case, (**b**) technology uncertainties only case, (**c**) carbon price uncertainty only case, and (**d**) technology and carbon price uncertainties case

by an almost instant switch to new technologies or resources. By contrast hedging against technology uncertainty results in the comparatively early introduction of new and advanced alternatives. This is, e.g., visible in the deployment schedules of oil resources, where under technology uncertainty depletion of conventional oil (characterized by relatively lower uncertainty) is delayed, while unconventional oil is introduced earlier in time (compared to the deterministic case). The result is a mixture of conventional and unconventional extraction in order to increase the resilience of the system against the possibility that unconventional oil might not become available at the expected price. The same holds not only for unconventional gas but also for other advanced technologies, which deploy comparatively earlier in the case of considering uncertainty. In the medium to long term this results in a more diverse technology portfolio, which we will discuss in some detail in Sect. 17.5.4 on diversification. Perhaps, worth noting is also the revival of coal under technology uncertainty, which is a direct result of the relatively lower uncertainty of coal occurrences and electricity generation technologies, which push coal tentatively back into the system around mid century. In the latter half of the century, coal becomes less deployed due to the increasing carbon price over time.

Comparing panels (a) and (c) illustrate the impact of carbon price uncertainty. It needs to be noted that the case with carbon price only uncertainty corresponds in theory to an alternative deterministic model run with relatively higher carbon price, since the risk term penalizes carbon emissions only, and all other energy system aspects are deterministic. There is thus no trade-off between different types of uncertainties as in the case of technology uncertainty shown in panel (b). It is therefore not surprising that the scenario with uncertain carbon prices features (compared to the deterministic case) primarily a further reduction of the deployment of carbon-intensive technologies (and emissions). Diversification as observed under technology uncertainty is thus lacking in this setup.

Note also the increase in total primary energy use, which is primarily a result of decarbonization of the end-use sectors by a fuel switch to hydrogen whose production is quite energy-intensive, in particular because we have adopted fossil fuel-equivalent TPES accounting method for renewable and nuclear energy.[13]

Comparing all three panels with uncertainty treatment (b)–(d) with the deterministic panel (a) we recognize quite different changes in quantity and structure of TPES. Whereas in the *technology uncertainties only* case (b), an increased use of coal as well as a diversification of TPES is observed, and the *carbon price uncertainty only* case (c) mainly shows more pronounced decarbonization and an increase of TPES towards the end of the century with no revival of coal. The results of the combined *technology and carbon price uncertainty* run in panel (d) combine some of these effects and are characterized by not only both diversification and decarbonization but also increased TPES towards the end of the century.

In summary, we thus conclude that it is important to take a holistic view and consider both technology- and policy-related uncertainties simultaneously. Only including one of them leads to either diversification or hedging against possibly high carbon prices with distinctly different technology portfolios. Incomprehensive account of uncertainty may thus lead to biased policy recommendations.

17.5.3 CO_2 Emissions

As mentioned earlier, our scenarios assume an expected value for the carbon price of about 4.6 US$/tC in 2010 which increases with the discount rate to about 370 US$/tC in 2100. This corresponds to the mean across all scenarios in the IPCC scenario database [15] with CO_2-equivalent concentration stabilization targets of 650 ppmv and above. The deterministic implementation of this carbon price trajectory results in our modeling framework in cumulative CO_2 emissions of $\sim$ 880 GtC over the course of the century, corresponding to CO_2 concentrations around 530 ppm towards 2100. Considering also non-CO_2 emissions, based on the scenario classification from Chap. 3 of the IPCC Fourth Assessment Report [11], this would corre-

[13] Each unit of electricity generated from renewable and nuclear energy contributes with 2.56 units to TPES corresponding to a conversion efficiency of 39 %.

spond to about 650 ppmv CO_2-equivalent concentrations. This result, although not surprising, illustrates that our modeling approach leads to very similar results than other deterministic models assessed by the IPCC.

Considering the uncertainty of the carbon price, however, we observe significant changes of the emissions pathway towards more stringent mitigation. This is particularly due to the lognormal distribution of the carbon price, including low-probability events in the tail with much higher carbon prices of several thousand US$/tC. Therefore, in the robust cases hedging against the tail of high carbon prices becomes a major motivation to reduce carbon emissions, even if additional technology uncertainty is considered. This response to uncertainty has also been observed in previous studies, such as [29, 42, 54]. While we derive the same conclusion, the reason for the response is different. Both [29, 54] conclude that relatively lower emissions would be rectified due to the uncertainty of climate change damages (i.e., uncertainties in the response of the physical climate system to an increase in GHG emissions), whereas our analysis suggests lower emissions because of the economic risk of uncertain carbon prices.

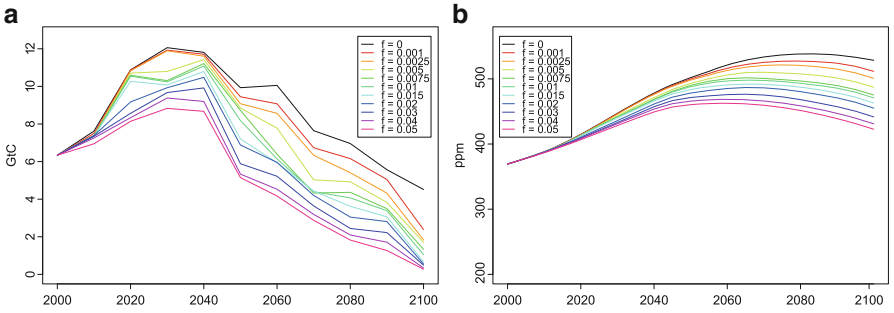

Fig. 17.9 (**a**) Annual energy-related carbon emissions in GtC and (**b**) atmospheric CO_2 concentrations in ppm as a function of the risk premium f

We find further that the stringency of mitigation is critically dependent on the risk premium. The relationship between the risk premium and annual carbon emissions and resulting atmospheric CO_2 concentrations is summarized in Fig. 17.9.[14] We observe that already very small additional hedging investments of only 0.1 % result in a reduction of cumulative emissions by 50 GtC or 5.5 % in comparison to the deterministic case. This corresponds to a CO_2 stabilization level of about 515 ppm. In our standard case with a risk premium of 1 %, cumulative emissions are approximately reduced by an additional 22 % ($\sim$ 690 GtC) for the linear risk measure (upper mean absolute deviation), corresponding to a further reduction of the CO_2 stabiliza-

[14] The annual carbon emissions in Fig. 17.9a are a direct model output whereas the CO_2 concentrations in the atmosphere in Fig. 17.9b are calculated with the climate model MagiCC 4.1 [52]. As our model calculates only energy-related emissions, we added for this purpose non-energy-related CO_2 emissions (e.g., land-use change, cement production, gas flaring) from a 670 ppmv stabilization scenario developed at IIASA [46].

tion level to ~ 480 ppm. The impact of the distribution's tail is more pronounced by the quadratic and CVaR risk measures. Thus, in the quadratic case, emissions are reduced by another 2 % points to about 670 GtC in comparison with the linear risk measure. The 95 %-CVaR risk measure even results in cumulative emissions of ~ 620 GtC reaching a CO_2 concentration level of close to 460 ppm CO_2 by the end of the century. For higher risk premiums even more effort is put into carbon abatement to limit the impacts of eventually high carbon prices, e.g., at $f = 3\,\%$ we find a reduction in excess of one third (580 GtC, 450 ppm) in comparison with the deterministic optimization and at $f = 5\,\%$ cumulative emission reductions constitute even more than 40 % (510 GtC, 430 ppm).

The peaking year of energy-related CO_2 emissions is only marginally affected by the risk premium and varies just between 2030 and 2040. However, the magnitude of the emission peak changes considerably from 12 GtC in the deterministic case ($f = 0$) to 8.8 GtC at a risk premium of $f = 5\,\%$. The impact on near-term emissions is relatively smaller in our standard case with a risk premium of 1 %, where emissions are about 10 GtC around 2030 and stay relatively unaffected until 2020, because of the energy system's inertia (see Fig. 17.9).

17.5.4 Diversification

As discussed in the context of the 3-technology model in Sect. 17.3, but also in the previous paragraphs, diversification may serve as a possible hedging strategy to increase resilience of a system. While we are focusing in this section particularly on the diversification within the electricity sector, it needs to be stressed that the energy system is more complex, and diversification may also occur as a result of shifting investments between different sectors. These aspects are discussed later in Sect. 17.5.5, which is focusing on the investment patterns under uncertainty.

In order to measure diversity, we employ an integrated multi-criteria diversity index developed by Stirling [51], which is based on distance metrics and will be referred to in the following as the Stirling index.[15] We are in particular interested in the relationship between the Stirling index and the risk premium and to which extent increasing risk aversion is triggering diversity as a response to uncertainty.

For this purpose, Fig. 17.10a, b display the electricity generation portfolio's dependence on the risk premium f in 2030 and 2050 respectively. In addition to the technology shares in electricity generation, the relationship between the Stirling index and the risk premium is shown on the right axes of the graphs. As can be seen from Fig. 17.10a, b the Stirling index is generally increasing at higher risk premiums, but the behavior is quite different in 2030 and 2050.

[15] The index is defined as $\mathbf{M} = \sum_{ij} d_{ij} p_i p_j$, where p_i is technology i's share of electricity generation and d_{ij} the distance in Euclidean disparity space between technology i and j [51, Chap. 3.2]. For the graphs a distance of 0.5 between fossil energy technologies (coal and gas) is used whereas for all other technologies we assume a distance of 1.

Fossil power generation from natural gas dominates electricity generation in the short to medium term as is evident from its high share in 2030 (Fig. 17.10a). In the deterministic case hydro and wind are the only other two technologies contributing to electricity generation. At risk premiums below 2 % this situation only marginally changes towards a larger share of hydropower plants. Above 2 % nuclear power comes in as a fourth option, resulting in a significant increase of the Stirling index. This situation indicates that the gap in levelized electricity generation costs between natural gas and nuclear is substantial (0.96 ct/kWh which corresponds to $\sim 26\,\%$ higher costs for nuclear) which requires a relatively high risk premium of 2 % to bring in this alternative. Coal power generation is phased out until 2030 as a result of a moderate carbon price and the uncertainties that come along with it.[16]

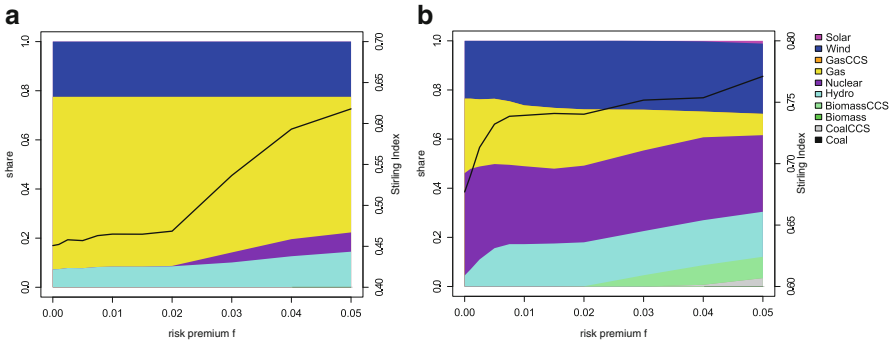

Fig. 17.10 Technology shares in electricity generation and corresponding Stirling index in (**a**) 2030 and (**b**) 2050 as a function of the risk premium f

In 2050 the deterministic electricity generation portfolio features four technologies, namely gas, nuclear, hydro, and wind that contribute to electricity generation. Already at relatively low risk premiums of less than 1 % the share of hydro more than triples at the cost of gas and nuclear power generation, resulting in a noticeably higher Stirling index. The reason for this early diversification is that levelized electricity generation costs are very close for these three technologies with hydro only being some 8 % and 6 % more expensive than gas and nuclear respectively. With further increasing risk premium the technology portfolio grows to seven technologies with biomass CCS power plant, coal CCS power plant, and solar PV joining in. In addition, the shares are much more evenly distributed, such that at $f = 5\,\%$ no technology supplies more than 31 % of total electricity in comparison to almost 42 % in the deterministic model run. This is an illustration of the previously cited *Don't put all your eggs in one basket* rule. The observed diversification is though significantly stronger by 2050 compared to 2030, due to the short-term inertia of the system against rapid structural changes.

[16] We assume on average a lifetime of 30 years for fossil power plants. Therefore by 2030 all power plants that were built in the base year 2000 reach the end of their lifetime.

It has to be emphasized that in contrast to modeling frameworks that explicitly aim at diversification as an objective (e.g., [50]) in our modeling framework diversification is an endogenous result driven by the aim to reduce risk. The extent of diversification is, however, critically dependent on the nature of the system and the dependence structure of joint input distributions which we have assumed (see Appendix 17.7.2). Our sensitivity analysis of the same scenarios indicates that in absence of any correlation between the costs of power generation technologies diversification would be significantly more pronounced.

17.5.5 Energy-Related Investments

We finally review the implications of the risk-hedging strategies for energy-related investments. Our systems engineering perspective permits us to explore shifts of investment between various technology clusters in fuel extraction, electric, non-electric (liquid fuels), and the energy end-use sectors.

The cumulative energy system investments between 2010 and 2050 are summarized in Fig. 17.11. Although we are dealing with a moderate stabilization scenario ($\sim$ 690 ppmv CO_2-equivalent concentration in the deterministic case), investments into fossil fuel technologies still dominate the first half of the century. In particular not only upstream investments but also electricity generation and liquid fuel production are characterized by high shares of fossil fuels, particularly in absence of uncertainty. This situation changes in the robust cases with increasing shares of investments into low-carbon options such as biomass, nuclear, and renewable electricity generation and synthetic fuels. In addition, increased efforts to improve energy efficiency in the end-use sectors become a more important factor in the robust cases, where the strongest increase occurs at risk premiums higher than 1 %. Most of these efficiency improvements take place in the non-electric and transport end-use sectors, because decarbonization is typically more costly in these sectors than, e.g., for electricity.

Total energy-related investments in the deterministic case are estimated to be around 49 trillion US$2,000 between 2010 and 2050. Additional investments into risk-hedging range between 7 % and 30 %, corresponding to a total of 52 and 64 trillion US$ in the cases with 1 % and 5 % risk premium respectively. Despite the comparatively modest increase of total costs, which is determined by the risk premiums, a significant increase in investments is required. Along with the increase of the total energy investments, we observe a considerable reallocation of investments among the different sectors of the energy system, most notably from the supply-side sectors to the end-use sectors, but also from fossil to renewable technology clusters. For example, the reallocation of investments between the four major sectors indicated in Fig. 17.11 (i.e., resource extraction, electricity generation, non-electric sector, and end-use) comprises 4.5 % and 15 % of total energy-related investments between 2010 and 2050 under the 1 % and the 5 % risk premium, respectively. These numbers increase further if reallocation of investment within the four major sectors,

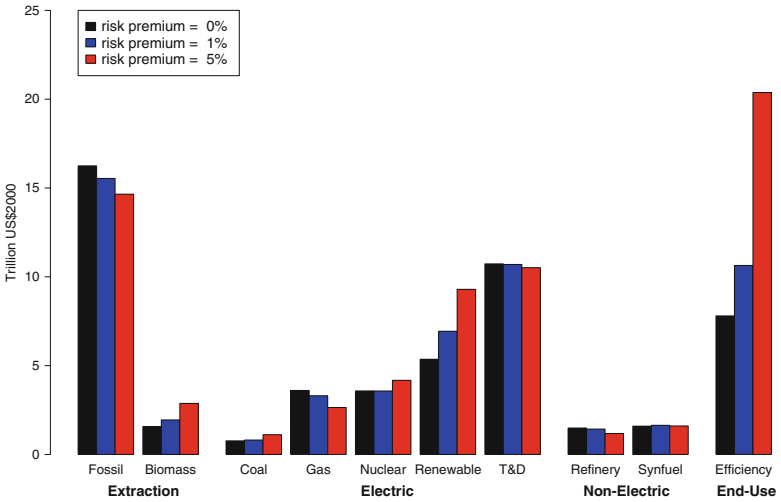

Fig. 17.11 Cumulative energy system investments for the period 2010–2050 in different sectors as a function of the risk premium

e.g., from fossil electricity generation towards renewables, is taken into account. It is interesting to note that the reallocation of investments becomes increasingly important for the lowest risk premiums of 1 % and below, simply because the total increase of energy system expenditures, and therefore also investments, is tightly constrained by the risk premium. For these cases, the reallocation effect is comparable to the absolute increase of energy investments.

In terms of energy expenditures, i.e., in addition to investments also including operation and maintenance costs, the reallocation effect is much more drastic. We find that the reallocation of energy-related expenditures is up to a factor of 10 higher than the total increase in expenditures in the case of very low risk premiums. In Sect. 17.5.1 it was shown that more robust solutions can be obtained even at very low hedging costs. However, the dominance of redistribution of sectoral investments and expenditures over the actual increase in costs, in particular at low risk premiums, illustrates that hedging results in significantly different investment patterns in comparison with the deterministic expected least cost solution. Therefore, it is not so much the total costs of hedging against technology and carbon price uncertainties that need to be in the focus of attention, but rather how investments are allocated within the energy system,[17] with major implications also for the appropriate portfolio of up-front R&D expenditures.

[17] From a more technical perspective this illustrates that quasi-degeneration is an important problem in modeling, i.e., quite different solutions can be accommodated within very small variations of the objective function value of optimization models.

17.6 Summary and Conclusions

Traditional deterministic energy models without an endogenous representation of uncertainty favor cost-optimal investments into a limited set of technologies that are expected to perform best in the future. Exploring the uncertainty of future energy systems costs, however, we find that such strategies can be very costly. This is in particular due to the nature of imputed energy systems uncertainties, characterized by long tails and the possibility of very high costs in case future uncertainties are resolved in an unfavorable direction.

In this chapter we thus presented a new modeling framework of the global energy system, which combines traditional elements of systems engineering modeling approaches with salient features of a risk management perspective. Employing optimization techniques from modern portfolio theory with a representation of uncertain costs and associated risks along the energy chain, including extraction and conversion technologies as well as demand-side management costs, permitted us to identify future development pathways that are cost-effective not only from today's perspective and expectations, but factor in also the imputed risk of uncertainty.

Through a series of sensitivity analysis we identify characteristics of risk hedging strategies that are adapted to considerably reduce future risks and are hence robust against a wide range of future uncertainties. We observe significant changes in response to energy system and carbon price uncertainties with major implications for the expected energy system costs, timing of investments, the choice of technology, as well as resulting emission levels.

Firstly, we find that hedging strategies under uncertainty are characterized by higher short- to medium-term investments into advanced technologies, including not only earlier deployment of renewables but also exploration of unconventional natural gas resources. Our results illustrate that while in the absence of uncertainty it seems to be cost-effective to postpone investments into new alternatives in order to maximize profits from available low-cost options early on; a more comprehensive view of the future including the uncertainty that new options might not become available at the expected costs imposes long-term deployment risks and thus triggers early up-front investments into niche markets and demonstration plants.

Secondly, we find that CO_2 emission reductions to be much more pronounced under uncertainty. This response to uncertainty has been observed in previous studies, such as [29, 54]. While we derive the same conclusion, the reason for the response is different. Yohe et al. [54] conclude that relatively lower emissions would be rectified due to the uncertainty of climate change damages (i.e., uncertainties in the response of the physical climate system to an increase in GHG emissions); our analysis suggests lower emissions because of the economic risk of uncertain carbon prices and technology costs.

Thirdly, our analysis suggests a considerable diversification of the technology portfolio under uncertainty. Diversification not only helps to reduce the "average risk," but it also results in significant reduction of the risk of high impact tail events. In our analysis, for example, a modest risk premium of about 1 % of total energy expenditures reduces the value of the 99th percentile by up to a factor of two relative to

the expected value expenditures, thus reducing the risk of large losses significantly. This conclusion has important implications for energy and climate policy, emphasizing the risk of unbalanced R&D portfolios or picking winners at a premature stage and thus focusing on a too narrow policy portfolio.

With respect to costs, we find that modest risk premiums (or hedging investments) can significantly reduce the vulnerability of the energy system against the associated uncertainties. The extent of early investments, diversification, and emissions reductions, however, depends on the risk premium that decision makers are willing to pay to respond to prevailing uncertainties. In other words, our modeling framework helps to understand how much risk can be avoided through which mechanisms and at what costs. How much risk needs to be avoided is though dependent on the risk aversion of the whole society or the decision makers in the respective sectors—and remains thus one of the key policy variables.

References

1. Ascough II JC, Maier HR, Ravalico JK, Strudley MW (2008) Future research challenges for incorporation of uncertainty in environmental and ecological decision-making. Ecol Model 219(3–4):383–399
2. Babonneau F, Kanudia A, Labriet M, Loulou R, Vial J-P (2012) Energy security: a robust optimization approach to design a robust european energy supply via tiam-world. Environ Model Assess 17(1-2):19–37
3. Beale EML (1955) On minimizing a convex function subject to linear inequalities. J R Stat Soc Series B (Methodological) 17(2):173–184
4. Ben-Tal A, El Ghaoui L, Nemirovski A (2009) Robust optimization. Princeton University Press, Princeton
5. Carnell R (2006) lhs: Latin hypercube samples. R package version 0.3
6. Charnes A, Cooper W (1959) Chance constrained programming. Manag Sci 6:73–89
7. Dantzig GB (1955) Linear programming under uncertainty. Manag Sci 1: 197–206
8. de Vries BJM, van Vuuren DP, Hoogwijk MM (2007) Renewable energy sources: their global potential for the first-half of the 21st century at a global level: an integrated approach. Energy Policy 35(4):2590–2610
9. Dessai S, Hulme M (2007) Assessing the robustness of adaptation decisions to climate change uncertainties: a case study on water resources management in the east of england. Glob Environ Change 17(1):59–72
10. Ermoliev Y, Wets RJ-B (eds) (1988) Numerical techniques for stochastic optimization, vol 10 of Springer series in computational mathematics. Springer, Berlin
11. Fisher B, Nakicenovic N, Alfsen K, Morlot JC, de la Chesnaye F, Hourcade J-C, Jiang K, Kainuma M, Rovere EL, Matysek A, Rana A, Riahi K, Richels R, Rose S, Vuuren Dv, Warren R (2007) Issues related to mitigation in the long

term context. In: Metz B, Davidson O, Bosch P, Dave R, Meyer L (eds) Climate change 2007: mitigation. Contribution of working group III to the fourth assessment report of the inter-governmental panel on climate change. Cambridge University Press, Cambridge, pp 169–250
12. Gritsevskyi A, Nakićenović N (2000) Modeling uncertainty of induced technological change. Energy Policy 28:907–921
13. Grübler A, Gritsevskyi A (2002) A model of endogenous technological change through uncertain returns on innovation. In: Grübler A, Nakićenović N, Nordhaus W (eds) Technological change and the environment. Resources for the Future Press, Washington
14. Grübler A, Nakicenovic N, Riahi K, Wagner F, Fischer G, Keppo I, Obersteiner M, O'Neill B, Rao S, Tubiello F (2007) Integrated assessment of uncertainties in greenhouse gas emissions and their mitigation: introduction and overview. Technol Forecast Soc Change 74(7):873–886
15. Hanaoka T, Kawase R, Kainuma M, Matsuoka Y, Ishii H, Oka K (2006) Greenhouse gas emissions scenarios database and regional mitigation analysis. Technical report, National Institute of Environmental Studies
16. IIASA GGI (2007) IIASA GGI scenario database. http://www.iiasa.ac.at/Research/ENE/GGIDB_index.html
17. Iman RL, Conover WJ (1980) Small sample sensitivity analysis techniques for computer models with an application to risk assessment. Commun Stat Theor Methods 9(17):1749–1842
18. Iman R, Conover W (1982) A distribution-free approach to including rank correlation among input variables. Commun Stat B 11:311–334
19. Kann A, Weyant J (2000) Approaches for performing uncertainty analysis in large-scale energy/economic policy models. Environ Model Assess 5(1):29–46
20. Kanudia A, Loulou R (1998) Robust responses to climate change via stochastic MARKAL: the case of Québec. Eur J Oper Res 106:15–30
21. Koomey J, Hultman NE (2007) A reactor-level analysis of busbar costs for us nuclear plants, 1970–2005. Energy Policy 35(11):5630–5642
22. Kouvaritakis N, Panos V (2005) Modelling the two factor learning curve equations - guidelines on implementation. Technical report, E3M Lab, School of Electrical and Computer Engineering, National Technical University of Athens
23. Krey V, Martinsen D, Wagner H-J (2007) Effects of stochastic energy prices on long-term energy-economic scenarios. Energy 32(12):2340–2349
24. Labriet M, Kanudia A, Loulou R (2012) Climate mitigation under an uncertain technology future: a tiam-world analysis. Energy Econ 34, Suppl 3(0): S366–S377
25. Lempert RJ, Groves DG, Popper SW, Bankes SC (2006) A general, analytic method for generating robust strategies and narrative scenarios. Manag Sci 52(4):514–528
26. Loulou R, Kanudia A (1999) Minimax regret strategies for greenhouse gas abatement: methodology and application. Oper Res Lett 25(5):219–230

27. Loulou R, Labriet M, Kanudia A (2009) Deterministic and stochastic analysis of alternative climate targets under differentiated cooperation regimes. Energy Econ 31(Supplement 2):S131–S143. doi: 10.1016/j.eneco.2009.06.012
28. Manne A (1996) Hedging strategies for global carbon dioxide abatement: a summary of poll results emf 14 subgroup – analysis for decision making under uncertainty. In: Nakicenovic N, Nordhaus WD, Richels R, Toth FL (eds) Climate change: integrating science, economics and policy, vol CP-96-1 of Conference proceedings. International Institute for Applied Systems Analysis, Laxenburg, pp 207–228
29. Manne AS, Richels RG (1992) Buying greenhouse insurance: the economic costs of carbon dioxide emission limits. The MIT Press, Cambridge
30. Markowitz HM (1952) Portfolio selection. J Financ 7(1):77–91
31. Marti K, Ermoliev Y, Pflug G (eds) (2004) Dynamic stochastic optimization, vol 532 of Lecture notes in economics and mathematical systems. Springer, Berlin
32. Messner S, Schrattenholzer L (2000) Message-macro: linking an energy supply model with a macroeconomic module and solving it iteratively. Energy 25(3):267–282
33. Messner S, Strubegger M (1995) User's guide for MESSAGE III. IIASA working paper WP-95-69, International Institute for Applied Systems Analysis (IIASA), Laxenburg
34. Messner S, Golodnikov A, Gritsevskii A (1996) A stochastic version of the dynamic linear programming model MESSAGE III. Energy 21(9):775–784
35. Metz B, Davidson O, Swart R, Pan J (eds) (2001) Climate Change 2001: Mitigation. Third assessment report of the IPCC. Cambridge University Press, Cambridge. http://www.grida.no/climate/ipcc_tar/wg3/
36. Nakicenovic N, Riahi K (2001) An assessment of technological change across selected energy scenarios. Report, World Energy Council (WEC), London
37. Nakićenović N, Swart R (eds) (2000) IPCC Special Report on Emissions Scenarios. IPCC special reports. Cambridge University Press, Cambridge. http://www.grida.no/climate/ipcc/emission/
38. Nakićenović N, Grübler A, McDonald A (eds) (1998) IIASA-WEC global energy perspectives. Cambridge University Press, Cambridge. http://www.iiasa.ac.at/cgi-bin/ecs/book_dyn/bookcnt.py
39. Palmquist J, Rockafellar RT, Uryasev S (1999) Portfolio optimization with conditional value-at-risk objective and constraints. Technical report, Center for Applied Optimization, Department of Industrial and Systems Engineering, University of Florida
40. Peterson S (2006) Uncertainty and economic analysis of climate change: a survey of approaches and findings. Environ Model Assess 11(1):1–17
41. Pflug G, Roemisch W (2007) Modeling, measuring and managing risk. World Scientific Publishing Company, London
42. Pizer WA (1999) The optimal choice of climate change policy in the presence of uncertainty. Resour Energy Econ 21(3–4):255–287

43. R Development Core Team (2008) R: a language and environment for statistical computing. R foundation for statistical computing. Vienna, Austria. ISBN 3-900051-07-0. http://www.R-project.org
44. Rao S, Riahi K (2006) The role of non-CO_2 greenhouse gases in climate change mitigation: long-term scenarios for the 21st century. Energy J 27(Special Issue Nov):177–200
45. Riahi K, Roehrl RA (2000) Greenhouse gas emissions in a dynamics-as-usual scenario of economic and energy development. Technol Forecast Soc Change 63(2–3):175–205
46. Riahi K, Grübler A, Nakicenovic N (2007) Scenarios of long-term socio-economic and environmental development under climate stabilization. Technol Forecast Soc Change 74(7):887–935
47. Rockafellar RT, Uryasev S (2000) Optimization of conditional value-at-risk. J. Risk 2(3):21–41
48. Rogner HH (1997) An assessment of world hydrocarbon resources. Ann Rev Energy Environ 22(1):217–262
49. Rokityanskiy D, Benítez PC, Kraxner F, McCallum I, Obersteiner M, Rametsteiner E, Yamagata Y (2007) Geographically explicit global modeling of land-use change, carbon sequestration, and biomass supply. Technol Forecast Soc Change 74(7):1057–1082
50. Stirling A (1994) Diversity and ignorance in electricity supply investment – addressing the solution rather than the problem. Energy Policy 22(3):195–216
51. Stirling A (1998) On the economics and analysis of diversity. Electronic working paper 28. University of Sussex, Brighton
52. Wigley T (2003) MAGICC/SCENGEN 4.1: Technical manual. Technical report, National Center for Atmospheric Research
53. Yan J (2007) Enjoy the joy of copulas: with a package copula. J Stat Softw 21(4):1–21
54. Yohe G, Andronova N, Schlesinger M (2004) Climate: to hedge or not against an uncertain climate future? Science 306(5695):416–417
55. Zhang Y, Pinder G (2004) Latin hypercube lattice sample selection strategy for correlated random hydraulic conductivity fields. Water Resour Res 39(8):1226

17.7 Appendix: Model Input Assumptions

This appendix provides additional information on numerical assumptions that have been used in the model runs as well as technical details about the model implementation.

17.7.1 Demands

As described in Sect. 17.4 we distinguish the three demand categories electric, non-electric (direct use) and transportation fuel demand. Demands are defined in terms of baseline final energy consumption derived from the B2 scenario that was developed with the IIASA GGI Integrated Assessment modeling framework [14, 16, 46]. Table 17.2 presents the numerical demand values for the periods 2000, 2050 and 2100.

Table 17.2 Final energy demands in the three demand categories for the periods 2000, 2050 and 2100

Demand [EJ]	2000	2050	2100
Electric	45.9	178.2	335.7
Non-electric	168.9	348.3	362.1
Transport	74.7	190.3	246.7

These demands can be lowered by additional investments into sectoral energy-saving measures which are modeled on the level of discretized two-step conservation cost curves. More details on the energy saving potentials and costs will be provided in the following section on technology data.

17.7.2 Technologies

Descriptions of individual technologies in the model include economic (e.g. investment costs, fixed and variable operation and maintenance costs) as well as technical parameters (e.g. efficiency, emissions, load factor). In our modeling framework only economic parameters, i.e. either investment or variable operation and maintenance costs, are assumed to be uncertain whereas all other parameters are treated deterministically and in addition do not include any time-dependence to simplify interpretation of results.

Due to the large potential for wind power, we distinguish two categories with different wind conditions which were derived from the analysis by de Vries et al. [8]. The cost data for the two categories are identical, just the load factor is assumed to be different to reflect the difference in wind conditions. Potentials for the two wind categories as well as for other renewable energy carriers are summarized in Table 17.5 of the following section.

17.7.2.1 Deterministic Parameters

Table 17.3 summarizes the assumptions made for all parameters of individual technologies. As mentioned above, the technical parameters in the first three columns of the table (pll = plant lifetime, plf = plant load factor, eff = net conversion efficiency)

Table 17.3 Deterministic technology-specific parameters (pll in years, inv and fom in US$/kW and vom in US$/kWyr)

Technology	pll	plf	eff	inv	fom	vom	inv	fom	vom
				– 2000 –			– 2100 –		
Coal extraction	10	1.00	1.00	130		35	130		35
Oil conv. extraction	10	1.00	0.96	150		80	150		80
Oil unconv. extraction	10	1.00	0.82	225		140	225		140
Gas conv. extraction	10	1.00	0.97	100		75	100		75
Gas unconv. extraction	10	1.00	0.92	180		160	180		160
Biomass $<$ 3US$/GJ	10	1.00	1.00			65			65
Biomass $\geq$ 3US$/GJ	10	1.00	1.00			125			125
Nuclear fuel	10	1.00	0.50			30			30
Coal power plant	30	0.75	0.38	1,300	74		1,100	62.6	
Gas combined cycle	30	0.75	0.50	716	51		400	32	
Nuclear power plant	30	0.75	0.38	2,500	108		1,800	99	
Biomass power plant	30	0.75	0.33	1,567	82		1,200	68	
Hydro power plant	50	0.42	0.38	2,500	40		2,500	40	
Wind turbine (cat. 1)	30	0.34	0.38	1,344	56		600	37	
Wind turbine (cat. 2)		0.22							
Solar PV	30	0.25	0.38	4,756	111		1,000	48	
Backup (e.g. CAS)	30	0.50	0.80	500	20		500	20	
Coal CCS module	30	0.75	−0.25	705		55.9	705		55.9
Gas CCS module	30	0.75	−0.13	503		19.8	503		19.8
Biomass CCS module	30	0.75	−0.25	846		109	846		109
Hydrogen electrolysis	30	0.95	0.80	452	20	4	380	15	4
Coal methanol	30	0.90	0.63	1,350	76	10	1,150	76	10
Gas methanol	30	0.90	0.68	630	46	5.4	480	35	5.4
Bioethanol	30	0.90	0.87	1,400	74	8	507	55	8
Refinery	30	0.90	0.93	66	7.5		66	7.5	
Electricity t/d	30	0.55	0.90	800	55	18	800	55	18
Gas t/d	30	0.70	0.95	200	24	3.5	200	24	3.5

are assumed to be time-independent. In contrast to that, economic parameters (inv = investment costs, fom = fixed operation and maintenance costs, vom = variable operation and maintenance costs) typically vary over time. Therefore, Table 17.3 contains two sets of these parameters, one showing the value in the base year 2000 and one for the year 2100. The interpolation procedure between these two values is described in Sect. 17.4.2 and is illustrated in Fig. 17.5 where additional technical details are provided in the following section on uncertain parameters.

As mentioned in Sect. 17.4 we do not model individual technologies in the end-use sectors, but have chosen to use discretized conservation cost curves instead. These cost curves were derived from a set of model runs (B2 baseline, 670 ppmv and 480 ppmv stabilization scenarios) with the 11-regional MESSAGE-MACRO model [32, 33, 46] which are documented in the corresponding scenario database [16]. Original scenario data were aggregated to the global level, after which an ex-

ponential trend in time was fitted to the data to obtain smooth curves. The resulting two-step discretizations of these conservation cost curves are summarized in Table 17.4, for the first year they become available to the model, i.e. in 2010, as well as for 2050 and for the end of the model's time horizon 2100. This procedure allows us to roughly reproduce the demand response in stabilization scenarios of the much more detailed MESSAGE-MACRO model which is part of IIASA's Integrated Assessment modeling framework [46] without adding the same degree of technological detail.

Table 17.4 Conservation cost curve parameters for the three demand categories

Conservation category	Potential [EJ]			Costs [US$/GJ]		
	2010	2050	2100	2010	2050	2100
Electricity 1	0.57	3.31	30.84	16.62	17.82	19.47
Electricity 2	1.51	2.62	5.17	19.15	20.36	21.97
Non-electric 1	2.18	11.07	85.43	4.12	6.18	10.18
Non-electric 2	9.05	25.01	89.25	4.31	8.18	18.20
Transport 1	0.82	5.49	58.78	4.00	6.63	12.49
Transport 2	4.23	13.50	57.68	4.28	8.24	18.74

The resource base of fossil energy carriers has its foundations in [48] and is adjusted to the B2 storyline [16, 46]. For coal all grades A–E are included, conventional oil includes categories I–III and unconventional oil categories IV–V. Conventional gas is an aggregate of categories I–III and unconventional gas covers categories IV–VI. Potentials of renewable energy carriers with the exception of wind are based on the estimates used in [16, 46]. Because of the large potential for wind we distinguish two categories with different wind conditions which are based on the B2 potentials published by de Vries et al. [8]. The assumptions used in the model are documented in Table 17.5. The values provided for fossil energy carriers correspond to the resource base available in the base year 2000. The quantities available in later years are then a result of the optimization procedure. For renewable energies the potentials are provided on an annual basis.

To calculate CO_2 emissions that originate from burning fossil energy carriers we uniformly apply the following emission factors on the basis of the lower heating value.

- Coal: 25.8 MtC/EJ
- Oil: 20.0 MtC/EJ
- Natural gas: 15.3 MtC/EJ

Biomass is predominantly provided by the forest sector and therefore assumed to be carbon neutral as a result of a sustainable production approach (see [49] for details).

Table 17.5 Fossil fuel resource base in 2000 and renewable energy potentials in 2000, 2050 and 2100 [EJ]

Resource category	2000	2050	2100
Coal	260,450	Endogenous	
Oil conventional	11,770	Endogenous	
Oil unconventional	8,890	Endogenous	
Gas conventional	17,920	Endogenous	
Gas unconventional	23,020	Endogenous	
Biomass $<$ 3US$/GJ	107	132	149
Biomass $\geq$ 3US$/GJ	22	62	106
Hydro	13	35	50
Wind (cat. 1)	20	54	54
Wind (cat. 2)	23	61	61
Solar PV	2.3	33	73

17.7.2.2 Uncertain Parameters

The PDFs of uncertain technology investment costs have been determined in the following way: Based on a review of technological change across selected energy scenarios [36] we have determined the expected value of costs in the year 2100 by calculating the median across the scenario assumptions. Also the variability of costs has been estimated from this source as described in Sect. 17.4.2. However, we only take the numerical values as a first indicator and assign the next highest uncertainty category (standard deviation σ: low uncertainty $\sigma_{low} = 0.15$, medium uncertainty $\sigma_{med} = 0.3$ and high uncertainty $\sigma_{high} = 0.6$) to the technologies, because of the limited set of scenarios that was included in the analysis. This procedure was applied to most electricity generation technologies, except the backup technology, and all liquid fuel technologies with the exception of the oil refinery. All other technologies were assigned to one of the three uncertainty categories based on experts' opinion, the result of which is shown in Fig. 17.4 as mentioned in Sect. 17.4.2.

Figures 17.12 and 17.13 show histograms and pair-wise scatter-plots for the nine electricity generation technologies based on the samples that result from the above described procedure. In Fig. 17.13 only scatter-plots for pairs of technologies with a correlation coefficient larger than 0.05 are shown. The investment costs correspond to the period 2100 where random sampling was used to generate samples of size $N = 20,000$.

To derive random cost paths for the model's full time horizon (2000–2100) we proceed in the following way: Based on one of the two sampling procedures described in Sect. 17.2.3 we generate N sets of random cost parameters for the period 2100. We then exponentially interpolate between the base year value in 2000 and the random parameter values in 2100. To fix the third parameter in the exponential function $y(x) = y_0 + A \cdot \exp(-\lambda x)$ we assume the asymptote y_0 to be 1 % higher (lower) than the 2100 value depending on whether the 2100 is higher (lower) than the 2000 value. A graphical illustration of this procedure for the investment costs

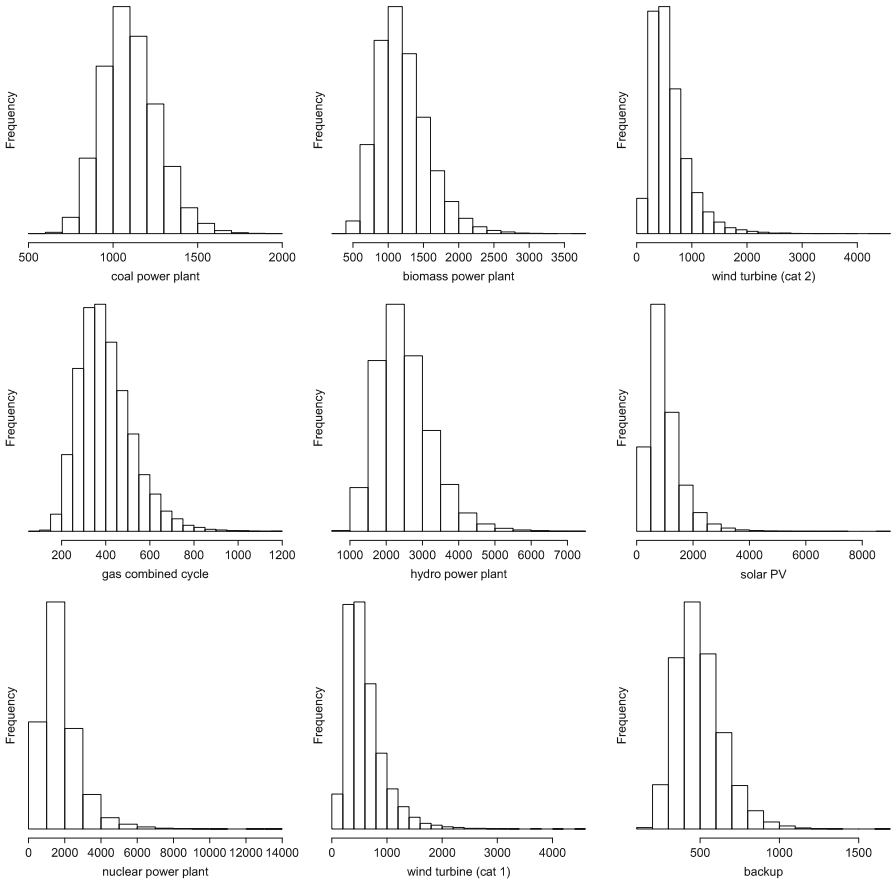

Fig. 17.12 Histograms of investment costs [US$/kW] of electricity generation technologies in the period 2100 for a sample size of $N = 20,000$ (random sampling)

of a natural gas combined cycle power plant is given in Fig. 17.5 of Sect. 17.4.2. In contrast to linear interpolation this procedure approaches the final level of costs relatively quickly—typically within the first half of the century—whereas cost levels stay relatively constant in the second half of the century. Therefore, envelopes of costs "widen" quickly, thus creating some overlap of input cost distributions early in the century.

17.7.3 Carbon Price

As briefly described in Sect. 17.4.2 the carbon price and its distribution in 2100 was derived from a sample of stabilization scenarios from the IPCC scenario database

[15]. A lognormal distribution fits the entire set of carbon prices from a total of 134 stabilization scenarios quite well (see Fig. 17.14). In addition, a Shapiro–Wilk normality test was conducted with the logarithmized data set which provides a p-value of 0.14 and therefore does not allow to reject the hypothesis of the logarithmized data being normally distributed.

The actual fit of the lognormal distribution used in the one-regional global energy systems model was obtained from a subsample of stabilization scenarios with CO_2-equivalent concentration targets of 650 ppmv and above. Parameters were determined by taking mean and standard deviation of the logarithmized dataset. This procedure resulted in numerical values for the expected value of $\bar{p}_c = 372$US$/tC and of $\sigma_c = 520$US$/tC for the standard deviation. The resulting values (mean as well as realizations of uncertain parameters) for the year 2100 were subsequently propagated backwards to 2010 with the model's 5 % discount rate to obtain carbon price trajectories.

17.8 Appendix: Sample Function Approximation

As described in Sect. 17.2.3, the risk measures $R(\mathbf{x}, \omega)$ are in practical model applications estimated by N independent realizations, so-called sample functions, where $N \to \infty$.

$$R(\mathbf{x}) \to R^N(\mathbf{x}) = \frac{1}{N} \sum_{s=1}^{N} R(\mathbf{x}, \omega^s).$$

The sample size N is evaluated by experiment, i.e. N is increased as long as different draws of the same sample size still produce noticeably varying solutions. Although it is problematic to proof convergence from a theoretical point of view, we find that in practice such approximations work better than expected theoretically (see also [10, Chap. 1.8]). For this purpose we introduce a so-called Taxicab- or 1-norm as a quantitative measure

$$\|\mathbf{x}\|_1 = \sum_i |x_i| \,, \tag{17.6}$$

where $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_T)$ is the vector of the model's decision variables for all periods $t = 1 \ldots T$.

Based on the Taxicab-norm we define a convergence criterion to measure (relative) deviations between two solutions i and j with identical sample size N

$$\Delta_{ij} = \frac{\|\mathbf{x}_i - \mathbf{x}_j\|_1}{\|\mathbf{x}_i + \mathbf{x}_j\|_1} \,, \tag{17.7}$$

where $\mathbf{x}_i$ and $\mathbf{x}_j$ are the solution vectors of the two solutions respectively. We require the maximum of all pair-wise distances Δ_{ij} to be less than a $\varepsilon > 0$

$$\max_{i,j} \Delta_{ij} \leq \varepsilon \,. \tag{17.8}$$

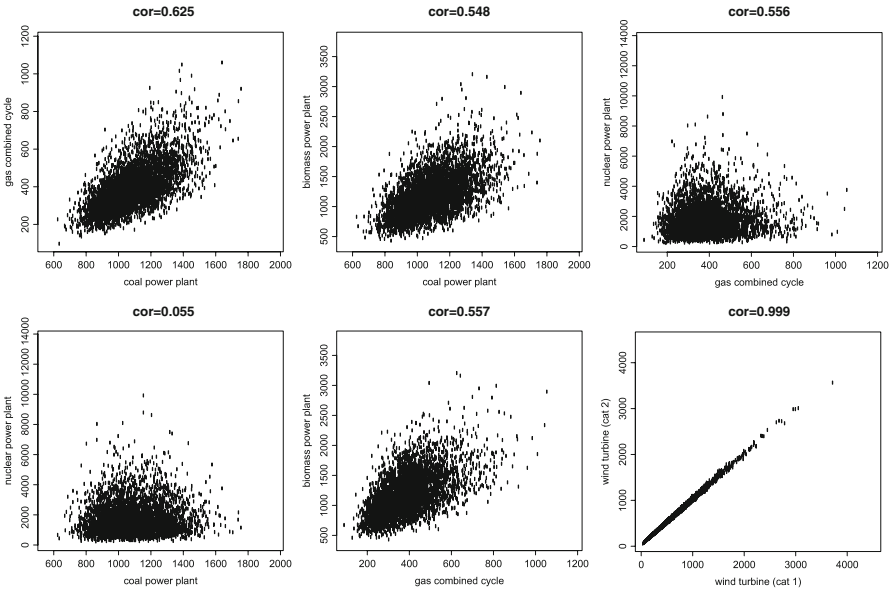

Fig. 17.13 Pair-wise scatter-plots of investment costs [US$/kW] of electricity generation technologies in the period 2100 with non-zero correlation ($\rho_{ij} > 0.05$) for a sample size of $N = 20,000$ (random sampling)

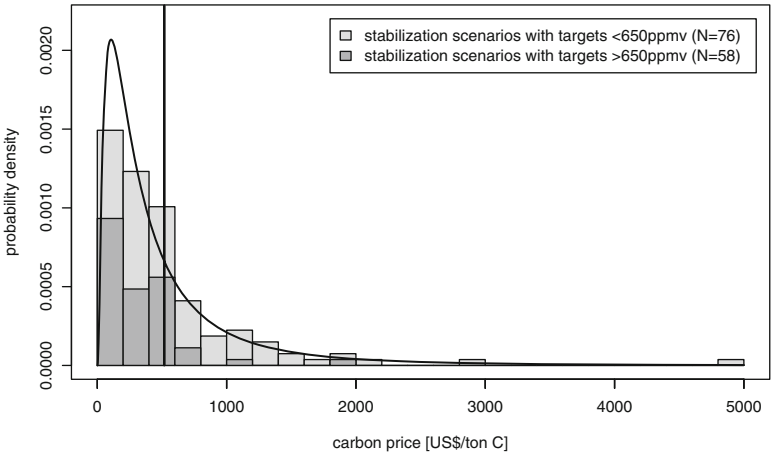

Fig. 17.14 Carbon price distribution of 134 stabilization scenarios in the IPCC scenario database [15] and lognormal fit to data

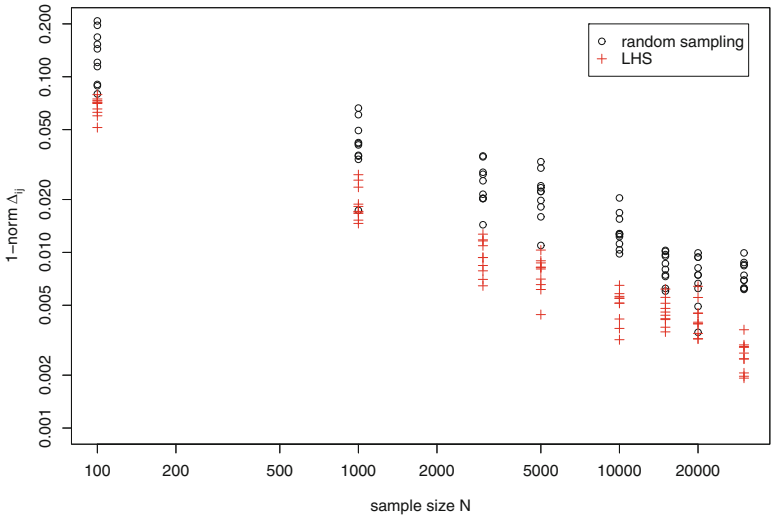

Fig. 17.15 Pairwise convergence measure Δ_{ij} as a function of sample size N for random sampling and LHS (5 model runs resulting in 10 combinations)—double-logarithmic scale

In realistic applications of energy systems models, the sample size N imposes restrictions on the number of uncertain parameters. Therefore, improved sampling methods to decrease the sample size N, but at the same time satisfying a given convergence criterion ε, are a possibility to reduce computational effort or alternatively increase the number of uncertain parameters. For this purpose we employ LHS (e.g. [17]) in contrast to random sampling. Correlations among latin hypercube sampled parameters are introduced with the algorithm suggested by Iman and Conover [18] (see also [55]).

The convergence behavior of LHS in comparison with random sampling can be seen in Fig. 17.15. This figure was generated with the model presented in Sect. 17.4 on the basis of 5 model runs for each sample size N with random and LHS. LHS gives significantly better convergence across all sample sizes N that have been analyzed. To obtain the same convergence level ε with LHS the sample size can be chosen almost one order of magnitude smaller than with random sampling. Therefore, the additional effort for employing LHS is well justified, in particular for larger models where memory limitations can become a constraint. As a result of these numerical experiments we used LHS for all model runs presented in this chapter. The sample size was chosen to be $N = 20,000$ corresponding to a convergence criterion of $\varepsilon = 0.75\,\%$.

Chapter 18
Comparative Assessment of Accident Risks in the Energy Sector

Peter Burgherr, Stefan Hirschberg, and Matteo Spada

Abstract This chapter is structured in five parts. The introduction discusses the relevance of accidents in the energy sector and puts them into the broader perspective of sustainability, energy security, and critical infrastructure protection. Furthermore, an overview of various risk assessment concepts is given. The second part provides a detailed overview of the comprehensive framework for comparative risk assessment developed by the Paul Scherrer Institut (PSI), at the core of which is the energy-related severe accident database (ENSAD). Third, a broad range of risk indicators and other measures are described and calculated allow for an objective, fair, and quantitative comparison of accident risks across a broad range of fossil, nuclear, and renewable technologies. This evaluation is complemented by a compilation of additional risk factors that can play a key role in decision processes and stakeholder interaction. However, for the time being they are often not amenable to full quantification because they cannot be described and analyzed by traditional risk metrics mainly focusing on consequences or because only limited historical experience is available. The chapter ends with a summary of the main findings and conclusions that can be drawn from comparative risk assessment as well as their potential implications for policy making.

18.1 Introduction

18.1.1 Definitions of Risk

Risk assessment is a well-established discipline with numerous important achievements in the past three decades [46]. Depending on the field of application and the object under study a large variety of definitions of the term risk exists [5, 49].

P. Burgherr (✉) • S. Hirschberg • M. Spada
Laboratory for Energy Systems Analysis Paul Scherrer Institut, Villigen PSI, Switzerland
e-mail: peter.burgherr@psi.ch

R.M. Kovacevic et al. (eds.), *Handbook of Risk Management in Energy Production and Trading*, International Series in Operations Research & Management Science 199,
DOI 10.1007/978-1-4614-9035-7_18,

In engineering and natural sciences, risk is frequently defined in a quantitative way: risk (R) = probability (p) · consequence (C). Another approach looks at risk as the triplet of threat (T) · vulnerability (V) · consequences (C) [28, 92]. Aven [4] proposes a slightly different approach using knowledge-based probabilities to express risk, vulnerability, and resilience. Additionally, subjective factors of risk perception can influence stakeholders' acceptance or aversion towards a specific technology, involving trade-offs between quantitative and qualitative risk factors [47, 97]. For example, energy choices may be influenced by potential climate change impacts and risk beliefs [45]. Furthermore, the consideration of low-frequency–high consequence events and the adequate treatment of uncertainty are key elements of every risk analysis [6, 80]. Finally, it is important that decision makers are offered information based on probable exposure scenarios, so that they can make risk-informed decisions [56].

18.1.2 Relevance of Accidents in the Energy Sector

In the past 40 years a number of catastrophic accidents (e.g., Santa Barbara oil spill from drilling platform in 1969, Three Mile Island nuclear power plant accident in 1979, Exxon Valdez tanker spill in 1989, Deepwater Horizon oil spill in 2010, or Fukushima nuclear accident in 2011) have affected the entire energy-related business and industry, not only due to the human health effects and environmental impacts but also due to the lessons learned from them [98]. This class of risks is of particular importance for two main reasons. On the one hand there is often a strong societal aversion towards such low-probability high-consequence events and on the other hand a perceived lack of urgency about these risks can be observed among the public and decision makers, despite the potential of such threats to dramatically affect human health, property, environment, society, and economy [44]. Therefore, comparative assessment of accident risks is a central aspect in a comprehensive evaluation of the performance of energy technologies, which has been recognized and emphasized repeatedly since the 1980s [41, 59, 85]. Although, it has been shown that accidents attributable to the energy sector form the second largest group of all manmade accidents worldwide, their treatment in terms of completeness and data quality was not satisfactory [42]. Therefore, the Paul Scherrer Institut (PSI) initiated in the early 1990s a long-term research activity, building upon extensive historical experience complemented by probabilistic safety assessment (PSA), at the core of which is the energy-related severe accident database (ENSAD) [52].

18.1.3 Accidents in the Context of Sustainability and Energy Security

Conceptually, comparative risk assessment is also closely intertwined with the broader concepts of sustainability, energy security, and critical infrastructure protection (e.g., [62, 96, 110]). However, energy security is a complex concept covering

many disciplines including (1) engineering responsible for technical safety and sufficient capacity, (2) economy concerned about functioning energy markets, and (3) political sciences analyzing geopolitical security threats. Thus, it is not surprising that there is no unique definition that grasps all aspects [25, 72, 77, 109]. However, a variety of concepts and approaches have recently been developed to comprehensively evaluate energy security and its implications [73, 103, 107, 108]. Furthermore, several holistic perspectives have been proposed and developed to address and evaluate vulnerabilities, interdependencies, and resilience of critical infrastructures [7, 40, 55, 71]. In recent years the aspect of risk governance has also received a prominent position on the international agenda [37, 60, 87]. Finally, there is a vivid discussion about rare and unimaginable events with large consequences, for which often the metaphors black swans, perfect storms, unknown unknowns, here be dragons, and dragon kings are used [36, 81, 94].

18.1.4 Comparative Risk Assessment of Energy Accidents

In spite of these methodological advancements and application developments, it is paramount that comparative risk assessment of severe accidents strongly relies on a consistent and comprehensive basis upon which a variety of risk indicators can then be systematically quantified to address the impacts of different energy technologies on human health, environment, economy, and society [13, 20, 54]. Risk indicators can also provide essential inputs for energy scenario analysis and policy formulation with regard to energy security aspects and sustainability performance associated with our current and future energy systems [10, 17, 34, 88]. Finally, several approaches have been proposed to combine risk assessment, life cycle assessment, and multi-criteria decision analysis (MCDA) in the past decade [58, 63, 74, 76]. While the comparative assessment of severe accident risks for major centralized technologies such as fossil (coal, oil, natural gas), hydro, and nuclear energy chains is well established [13, 17, 20, 52], corresponding evaluations for new renewables like solar photovoltaics (PV), wind, biomass, and geothermal electricity generation have received substantial interest in recent years due to their advances in technological development and market penetration [10]. A key challenge of broadening the analytical scope by including hazards and risks from new renewables is that technology comparisons are based on a set of common and consistent risk indicators, which can be expressed in quantitative terms and are applicable to the various electricity generation technologies under consideration. Since its first release in 1998 [52], the methodological framework based on ENSAD has been refined and extended by adding numerous new elements and broadening the analytical scope and coverage, including:

- Consideration of new information sources for specific energy chains and countries, e.g., coal China [12, 53], natural gas [11], and oil spills [8, 21, 35]
- Estimation of external costs [13, 16]
- Application of a simplified level-3 PSA for nuclear energy [16, 17]
- Georeferencing of accident data and use of geographic information systems (GIS) [9, 20]
- Evaluation of new renewable technologies [10, 17]
- Calculation of specific risk indicators to be used within MCDA [34, 88, 90]
- Development of advanced analytical approaches such as Bayesian hierarchical networks and fat tail distribution fitting [33]
- Consideration of intentional attacks within the broader context of energy security and critical infrastructure protection [15, 18]

In the past decade accidents in the energy sector have received a high degree of attention by policy makers and the general public. On the one hand this is attributable to several exceptionally disastrous events with correspondingly high media coverage. On the other hand consideration of accident risks has become an integral part of sustainability and energy security evaluations of energy technologies as well as energy scenario analysis. Therefore, a growing number of studies on accident risks in the energy sectors have been published in recent years, covering a broad range of aspects, such as different energy technologies, geographic regions, consequence categories, safety issues, financial and market aspects, and strategy and policy development. A concise overview of selected examples is given in Table 18.1.

Table 18.1 Selection of recent studies on accident risks in the energy sector

Thematic area	References
Coal mine accidents	China [23, 75, 102]
	India [79]
	USA [2, 89]
Oil refineries	[66, 99]
Downstream oil industry	[38]
Oil spills from tankers	[51, 61, 67, 68, 84]
Offshore oil facilities	[3, 31, 83, 86, 93, 104, 105]
Carbon capture and storage (CCS)	[48, 69, 111]
Solar photovoltaics	[43]
Wind power	[22, 78]
Biofuel	[32]
Accidents triggered by natural hazards (Natech)	[29, 30, 65, 70]
Comparative studies	[27, 39, 91, 95]

18.2 Methodological Approach

18.2.1 Framework for Comparative Risk Assessment

Figure 18.1 shows a graphical overview of PSI's overarching methodological framework for comparative risk assessment, which is an essential component in the context of comprehensive sustainability and energy security assessments, and can also provide inputs to decision-making processes. The ENSAD is the central and connecting element of this framework. In its initial implementation the analysis was limited to technological accidents of major centralized technologies, namely fossil energy carriers and hydro and nuclear power [52]. However, in the past years the scope has been significantly extended to specifically include (1) accidents that are triggered by natural hazards (Natech accidents) and (2) intentional malicious attacks (e.g., vandalism, sabotage, sabotage, terrorist threat). Additionally, evaluations for a broad portfolio of low-carbon technology options were developed and established, ranging from new renewables (e.g., photovoltaics, wind, geothermal) to fossil with CCS and advanced nuclear designs (generations III and IV). For fossil energy chains (coal, oil, natural gas) and hydropower, extensive historical experience is contained in ENSAD for the period 1970–2008 [19]. However, for hydropower some complementary, site-specific analyses are also carried out. In the case of nuclear energy a simplified level-3 PSA is employed to address hypothetical accidents [17, 19]. In contrast, consideration of new renewable technologies is based on a "hybrid" approach that combines available accident data with chain-specific modeling and expert judgment [10, 19]. Depending on the actual study objectives, a tailored data set can be compiled, which then can be analyzed (1) for individual energy chains, (2) for comparison between energy chains, and (3) for addressing spatial patterns and temporal trends. Furthermore, evaluations can be performed at the level of technologies or supply mix scenarios, also allowing extrapolating into the future.

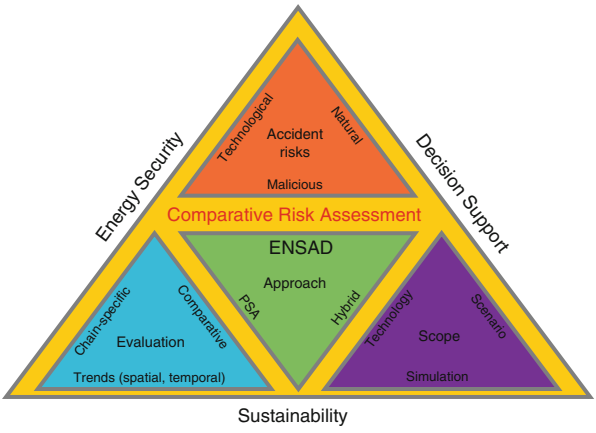

Fig. 18.1 Comprehensive framework for comparative risk assessment in the energy sector

18.2.2 Severe Accident Database

ENSAD is a relational database in MS Access format that since its initial release [52] has been continuously updated and extended, with regard to content, scope, and analytical capabilities. An overview of the main structural elements of ENSAD is given in Fig. 18.2. Every accident record in ENSAD is assigned a unique identification number (ID) that will only be used once, even if a record is deleted later. Accident information can be divided in two main blocks. The first block is directly (1:1) linked, containing information on the incident date, location, event categorization and description, and meta and change log information. The second block is linked through a 1:n relationship because information on damages can vary among different primary information sources, and in this way it is possible to take into account all available information.

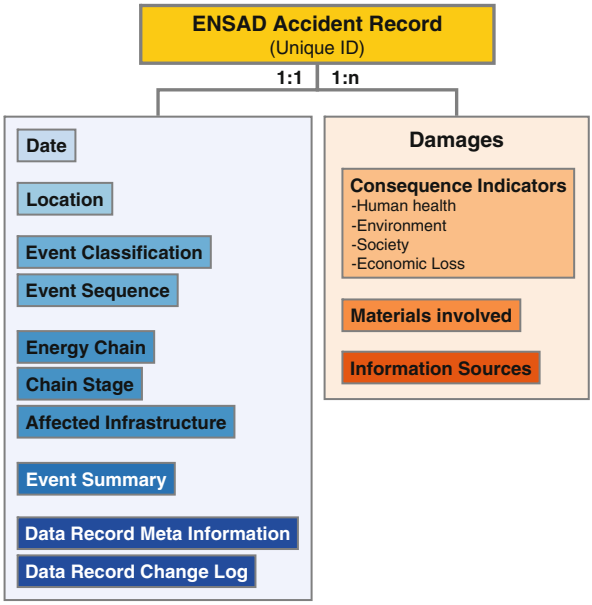

Fig. 18.2 Schematic representation of the accident record structure in ENSAD

ENSAD utilizes merged and harmonized historical data from a large variety of information sources. Therefore, ENSAD can be considered superior compared to single database approaches that are often limited concerning geographic area, time period, and energy chains included. Table 18.2 gives an overview of the main information sources that have been considered within ENSAD; however, in total, about 80 different sources have been used to date. The reason for this large amount of databases is twofold. First some of the databases used in the initial phase [52] have been discontinued (e.g., MHIDAS), but could be replaced by new ones (e.g., OSH

Update). Second, databases can be classified as primary (ca. 30) and secondary (ca. 50) information sources, the latter is important for purposes of cross-checking and complementing gaps in retrieved data. Finally, one should note that both freely available sources and commercial databases are regularly surveyed because the latter may contain proprietary information not available at all or documented in a less detailed manner in noncommercial sources.

Table 18.2 Selection of main information sources used within ENSAD. Abbreviations: *C* = commercial database, *F* = freely available database

Database	Geographic area	Accident types
Hint (C)	Worldwide	Industry
OSH update (C)	Worldwide	Industry
Swiss Re (C/F)	Worldwide	Natural and man-made disasters
EM-DAT (F)	Worldwide	Natural and man-made disasters
Industrial Fire World Log (F)	Worldwide	Industry
FACTS online (C)	Worldwide	Industry
MHIDAS (C)	Worldwide	Industry
eMARS (F)	EU, OECD, UNECE	Industry
Aria/Barpi (F)	Worldwide	Industry
Lloyd's casualty service (C)	Worldwide	Industry
China coal industry yearbook (C)	China	Coal
WOAD (C)	Worldwide	Offshore oil and gas
Centre de documentation, de Recherche et d'Expérimentations sur les Pollutions Accidentelles des Eaux Cedre (CEDRE) (F)	Worldwide	Tanker oil spills
International tanker owners pollution federation ltd. (ITOPF) (C/F)	Worldwide	Tanker oil spills
International oil pollution compensation funds (IOPCF) (F)	Worldwide	Tanker oil spills
The Center for tankship excellence (CTX) (F)	Worldwide	Tanker oil spills
Regional marine pollution emergency response centre for the Mediterranean Sea (REMPEC) (F)	Mediterranean	Tanker oil spills
National oceanic and atmospheric administration (NOAA), NOAA incident news (F)	Mainly USA	Oil spills
Bibliography of the History of Dam Failures (BHDF)	Worldwide	Dams
ICOLD Catalogue of Dam Disasters	Worldwide	Dams

ENSAD is clearly focused on so-called severe accidents. Accidents resulting in more severe consequences are of greatest concern to industry, authorities and regulators, insurance companies, policy makers, and the general public. Furthermore, accidents with disastrous impacts on the human health, the environment, the society, and the economy receive a broad and often controversial coverage in the media. Finally, a much higher level of completeness and accuracy of reported information are generally available for severe accidents, which is also important with regard

to comparisons among individual countries and country groups and globally. In the literature no commonly accepted definition can be found of what constitutes a severe accident. Differences concern the actual damage types considered (e.g., fatalities, injured persons, evacuees, or economic costs), use of loose categories such as "people affected," and differences in damage thresholds to distinguish severe from smaller accidents [19, 52]. The severe accident definition used in ENSAD comprises seven criteria. An accident is considered severe if it fulfils at least one severity threshold (Table 18.3). Generally, fatality data is the most reliable, accurate, and complete, whereas for injured or evacuated persons, details on the severity of an injury or the duration of an evacuation are frequently not clearly indicated. The estimation of precise values for economic loss is also often difficult as different sources of information report various types of economic damages (e.g., insured vs. total loss), depending on their specific scope (e.g., insurance company vs. disaster recovery organizations). The other consequence indicators are either only relevant for specific energy chains (i.e., hydrocarbon release and land/water contamination) or ENSAD contains very few entries (ban on food consumption). Therefore, ENSAD-based results presented here are focused on the number of fatalities. A detailed discussion of the various consequence indicators including their accuracy, robustness, and completeness is for example given in [17].

Table 18.3 Severity thresholds

Risk description	Impact category	ENSAD severity threshold	Consequence indicator
Human health	Fatalities	≥ 5	Fatalities per GWeyr
	Injured	≥ 10	Injured per GWeyr
Societal	Evacuees	≥ 200	Evacuees per GWeyr
	Ban on food consumption	yes	Nominal scale
Environmental	Release of hydrocarbons	$\geq 10{,}000$ t	Tonne per GWeyr
	Land/water contamination	$\geq 25\,km^2$	km2 per GWeyr
Economic	Economic loss	≥ 5 Mio $USD_{(2000)}$	USD per GWeyr

18.2.3 Scope and Assumptions for Analysis

Full-Energy Chains: When comparing accident risks across different energy chains it is important to consider complete chains because accidents can occur at all stages, and also the contributions from different stages can differ between chains [17, 19, 52]. In general, an energy chain comprises the following stages: exploration, extraction, processing and storage, long distance transport, regional and local distribution, power and/or heat generation, waste treatment, and disposal. However, not all of these stages are applicable to every energy chain.

Evaluation Period: ENSAD currently holds a total of 32,705 accident records, of which 83.2% are classified as man-made, 16.3% as natural disasters, and 0.5% as

conflicts. Within the category of man-made accidents, 20,245 are energy-related, the vast majority of these (93.8%) fall in the period 1970–2008, 1.2% occurred in the years 1960–1969, 4.1% after 2008, and 0.9% before 1960. Accidents prior to the 1970s are normally not included in the analysis of existing technologies because they are too far back to be relevant for current conditions. For a more detailed discussion see [17, 19, 52].

Data Normalization: To enable a fair and unbiased comparison of accident risks among energy chains, it is important to use data normalized to the unit of energy production. For fossil energy chains the thermal energy is converted to an equivalent electrical output using either a generic efficiency factor (used in the examples of this chapter) of 0.35 [13, 52] or chain-specific efficiency factors [17]. For nuclear power, hydropower, and new renewables the normalization is straightforward because the generated product is electrical energy. Although not a SI unit, the Giga-Watt-electric-year (GWeyr) is used because large individual plants have capacities in the order of 1 GW electrical output (GWe), and thus the GWeyr is a natural unit to compare accident risks in technology assessment.

Regional Aggregation: Risk indicators can be calculated for different spatial scales, i.e., for individual countries and country groups or globally. The chosen resolution depends on the specific study objectives, but in most cases the level of country groups is adequate. Furthermore, a global analysis naturally does not provide any regional differentiation, and looking at individual countries is only possible if sufficient data are available for the countries of interest. Therefore, risk indicators are commonly calculated for three major country groups, i.e., OECD and non-OECD countries as well as European Union (EU 27). This distinction is meaningful because of significant differences in safety management and regulatory frameworks between highly developed countries (OECD and EU 27) and the mostly less developed non-OECD countries [14]. The Chinese coal chain is treated separately because it has been shown that its performance is clearly different from other non-OECD countries [12, 53]. In the case of nuclear site-specific calculations are performed, using a simplified level-3 PSA, with results available for Switzerland, Germany, France, Italy, Finland, the USA, and China [17, 19, 52, 88]. For hydropower, detailed dam failure modeling and consequence calculations for a large Swiss dam have been undertaken to complement historical experience ([11] and references therein). For new renewables, risk indicators can be considered representative for average conditions in OECD and EU 27 countries [10].

Evaluation of Accident Risks: Classical risk indicators include aggregated indicators (e.g., fatalities per GWeyr) and frequency consequence (F-N) curves, both of which have been extensively described in [52] and are briefly summarized in the section "fossil energy chains." However, various new methods and techniques have recently been adapted for use in comparative risk assessment:

- Estimation of economic losses and external costs to monetize impacts of accidents and to provide a common basis for aggregation of different types of consequences (e.g., fatalities, injured, evacuees) (e.g., [13, 16])
- Geo-referencing of accident data to allow for geostatistical analysis and risk mapping using GIS (e.g., [9])

- Better representation of low-frequency–high-consequence events through extreme value theory and distribution fitting (e.g., Generalized Pareto (GP), Value at Risk (VaR) (e.g., [19, 33]).

18.3 Evaluation of Accident Risks in the Energy Sector

This section provides a selection of risk assessment results. First, an overview of the accident data contained in ENSAD is given. Second, fossil energy chains are analyzed because for them extensive historical experience in ENSAD is available. Third, aggregated risk indicators are calculated for a broad portfolio of large centralized technologies (coal, oil, natural gas, hydropower, and nuclear) as well as decentralized new renewables (PV, wind, biogas, solarthermal, and geothermal). Fourth, external costs of accidents in fossil energy chains are presented.

18.3.1 Overview and Contents of ENSAD

The selection of results presented in this chapter is focused on severe accidents that resulted in at least five fatalities. Summary statistics for fossil, hydropower, and nuclear accident data contained in ENSAD for the period 1970–2008 are shown in Table 18.4, amounting to a total of 3,222 severe ($\geq$ 5 fatalities) accidents with a cumulated number of 96,630 fatalities.

Table 18.4 Summary of severe ($\geq$5 fatalities) accidents contained in ENSAD for fossil, hydropower, and nuclear accidents in the period 1970–2008. *Acc* = accidents, *Fat* = fatalities

	OECD Acc/Fat	EU 27 Acc/Fat	Non-OECD Acc/Fat	
Coal	87/2259	45/989	2394/38672	(all non-OECD)
			162/5788	(non-OECD w/o China)
			818/11302	(China 1994–1999)
			1214/15750	(China 2000–2008)
			2032/27052	(China (1994–2008)
Oil	187/3495	65/1243	358/19516	
Natural gas	109/1258	37/367	78/1556	
Hydro	1/14[a]	1 /116[b]	22/30083[c]	(all non-OECD)
			21/4083	(w/o Banqiao/Shimantan)
			12/26108	(China alone)
			11/108	(China w/o Banq./Shim.)
Nuclear	-/-	-/-	1/31[d]	

[a] Teton dam failure (USA, 1976)
[b] Belci dam failure (Romania, 1991)
[c] Banqiao/Shimantan dam failures (China, 1975) together caused 26000 fatalities
[d] Only immediate fatalities of the Chernobyl accident are shown here. See section "comparison of fossil, nuclear and renewable technologies" for a more detailed discussion of the nuclear chain

18.3.2 Fossil Energy Chains

A comprehensive evaluation of how severe accident risks should take into account a variety of factors because no single aspect can provide the full picture. Therefore, a set of six risk indicators was defined:

- The number of accidents per GWeyr (accident rate)
- The number of fatalities per GWeyr (fatality rate)
- Mean accident severity, i.e., the average number of fatalities per accident
- The Value at Risk for the 99th percentile (VaR99) of the severity distribution
- The probability of exceeding VaR99
- The probability of at least one severe accident per year

While fatality rates provide a measure of the expected fatalities per unit of energy produced, maximum credible consequences of a single accident can be seen as a measure of risk aversion. Based on historical accident data, maximum consequences for a specific energy chain and country group can be simply determined as the most deadly accident in a given time period, which is straightforward but disregards the specific distribution properties of accident data. Alternatively, one could use the VaR that is commonly used in financial risk management and measures the consequences for low-frequency extreme events at a given percentile (e.g. 99%). Here we estimated VaR99 using the Epanechnikov kernel density estimator, which does not assume that the data follow a normal or any other distribution and thus does not make VaR estimation normal or even parametric [1]. Additionally, the probability of exceeding VaR99 is also calculated, using the formula p = m/n+1, where m is the accident rank and n is the total number of accidents. The probability that at least one severe accident per year occurs is calculated using the equation p = 1 - exp(-R*T), where R denotes the consequences (fatality rate) per year and T is equal to 1 year. Figure 18.3 shows the above-described six indicators for fossil energy and different country groups in the period 1970–2008. Generally, OECD and EU 27 countries exhibit a better performance than non-OECD.

In a second step, the resulting matrix of 6 risk indicators × 11 samples (i.e., energy chain-country group combinations) was then analyzed by means of principal component analysis (PCA), which was first described by [82] and [57], but see [64] for a detailed overview. Generally, PCA is a mathematical procedure that transforms a set of possibly correlated variables (here indicators) into a smaller number of uncorrelated variables, the so-called principal components (PC). Thus, it provides a reduction of multidimensional data onto two or three independent (orthogonal) axes as well as graphic representation of inter-sample and inter-variable relationships for exploratory data analysis. Data were normalized to ensure equal weights for all variables and then subjected to correlation matrix PCA using the ADE-4 software [26, 101]. Figure 18.4 shows (a) the correlation circle and (b) the positions of the 11 analyzed country aggregates on the PC1 × PC2 factorial map. The first two PC axes explained 49.7% and 36.1% of the total variation. The probability of exceeding VaR99 was positively related to PC1, whereas the probability of at least one severe accident per year was negatively related to PC1. The axis PC2 was best explained

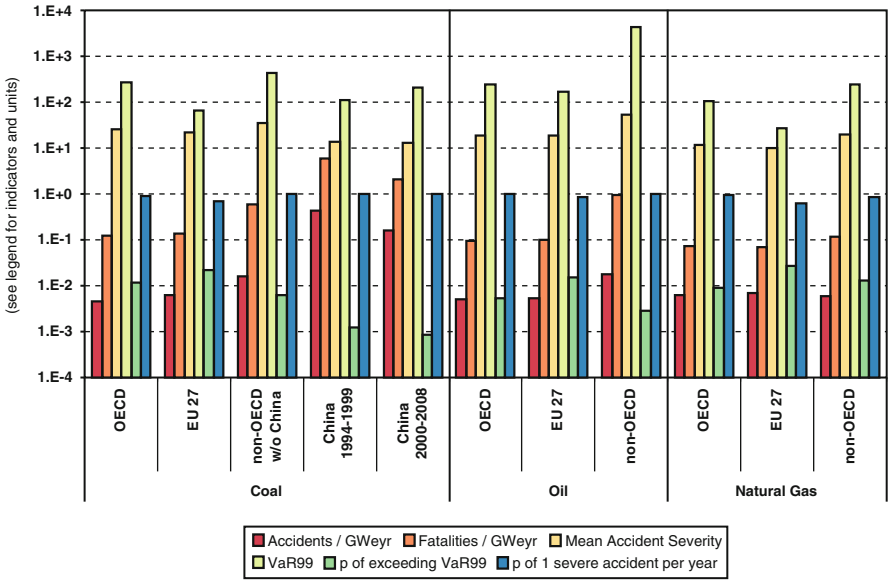

Fig. 18.3 Comparison of six accident risk indicators based on historical experience of severe (≥ 5 fatalities) accidents in fossil energy chains for OECD, EU 27, and non-OECD countries in the period 1970–2008. Note that coal chain was treated separately (compare section on methodological approach), with indicators calculated for two time periods

by VaR99 and mean accident severity. Accident and fatality rates contributed about equally to both axes. Coals China during 1994–1999 and 2000–2008 are most isolated from all other country groups, which is primarily due to the large annual numbers of severe accidents (and thus high accident rates) and by far highest fatality rates. However, it is also apparent that coal China is slowly approaching other non-OECD countries. The non-OECD oil chain is also clearly separated because VaR99 and mean accident severity are highest, which is also in accordance with actual historical experience that the two most deadly oil accidents occurred in the Philippines (4,386 fatalities) and Afghanistan (2,700). OECD and EU 27 countries are both clearly separated along PC1, whereas non-OECD countries (including China) were separated along PC2. Finally, fossil energy chains in OECD and EU 27 showed distinctly lower variation compared to non-OECD and coal China. Third, the comparison of results was expanded beyond aggregated risk indicators by combining results of the frequency and consequence analyses by means of frequency-consequence (F-N) curves. F-N curves are a common way to express collective or societal risks in quantitative risk assessment, that is, the F-N curves provide an estimate of the risk of accidents that affect a large number of people. F-N curves show the relationship of the cumulative frequency (F) of events having consequences $\geq$ N (e.g., fatalities) which is usually presented in a diagram with double logarithmic axes. In this way, they comprise more information than aggregated indicators because they show the probability of accidents with varying degrees of severity of consequences,

including chain-specific maximum damages. Additionally, the 5% and 95% confidence intervals can be calculated for the point estimates of the F-N curve, based on the Chi-square distribution [50]. Figure 18.5 shows F-N curves for coal, oil, and natural gas in OECD, EU 27, and non-OECD countries for the period 1970–2008. Generally, OECD and EU 27 countries had lower frequencies than non-OECD for all fossil chains; however, for natural gas, this difference was less distinct. The Chinese coal chain showed a significantly worse performance than other non-OECD countries, with frequencies even greater than 10-1 at lower death tolls (5–13 fatalities). Concerning maximum consequences, OECD values were between 1.5 and 4.2 times greater than EU 27. Non-OECD values were substantially higher than corresponding OECD values, except for coal China where only 15 out of 2,032 accidents resulted in 100 or more fatalities. Finally, the range of observed maximum consequences among fossil chains was larger in non-OECD, particularly because the oil chain can reach maximum numbers up to one order of magnitude higher than other fossil chains due to two extremely deadly accidents in the Philippines and Afghanistan.

18.3.3 Comparison of Fossil, Nuclear, and Renewable Technologies

Figure 18.6 shows the two indicators fatality rate and maximum consequences for a broad portfolio of fossil, nuclear, and renewable technologies and different country groupings. For fossil energy chains and hydropower, fatality rates and maximum consequences are generally lower in OECD and EU 27 countries compared to non-OECD. Among fossil chains, natural gas performs best with respect to both indicators. The fatality rate for coal China (1994–2008) is significantly higher than for the other of non-OECD [12, 53]; however, a comparison between 1994–1999 and 2000–2008 (see Fig. 18.3) indicates that China could slowly approach the non-OECD level. Among large centralized technologies, hydro OECD and Western-style nuclear power plants exhibit the lowest fatality rates, but at the same time the consequences of catastrophic accidents can be very large. Experience with hydro in OECD countries points to very low fatality rates, although based on a theoretical model it has been shown that maximum consequences for the total failure of a large Swiss dam could range between 7,125 and 11,050 fatalities without pre-warning, but could be reduced to 2–27 fatalities given sufficient prewarning time ([11] and references therein). In contrast, several dam failures in non-OECD have claimed large numbers of victims. Concerning nuclear energy three core-melt events have occurred at Three Mile Island 2 (TMI-2, USA, 1979), Chernobyl (Ukraine, 1986), and Fukushima Daiichi (Japan, 2011). The TMI-2 accident did not result in any immediate fatalities, and due to the small release of the radioactivity the estimated collective effective dose to the public was about 40 person-Sv. Based on this one extra cancer fatality was estimated. However, 144,000 people were evacuated from the area around the plant. For more information see [52]. The Chernobyl accident

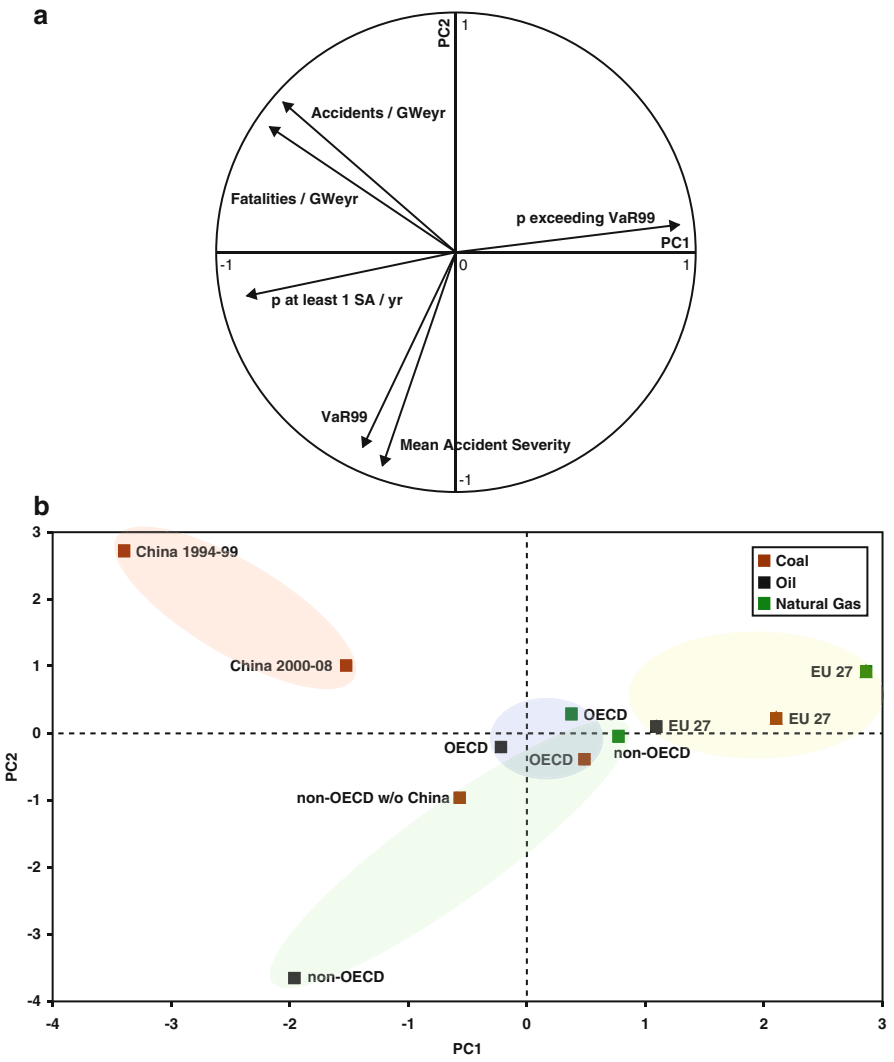

Fig. 18.4 PCA results showing (**a**) the correlation circle of six risk indicators and (**b**) graphical representation of the various energy chains and country groups in the PC1 × PC2 factorial plane

caused 31 immediate fatalities, whereas PSA-based maximum consequences including expected latent fatalities range from about 9,000 for Ukraine, Russia, and Belarus to about 33,000 for the whole northern hemisphere in the next 70 years [52]. According to a recent study by numerous United Nations organizations up to 4,000 people could die due to radiation exposure in the most contaminated areas [24]. This estimate is substantially lower than the upper limit of the PSI interval, which, however, was not restricted to the most contaminated areas. Published health effects of the Fukushima Daiichi nuclear accident show large variations, which is partially

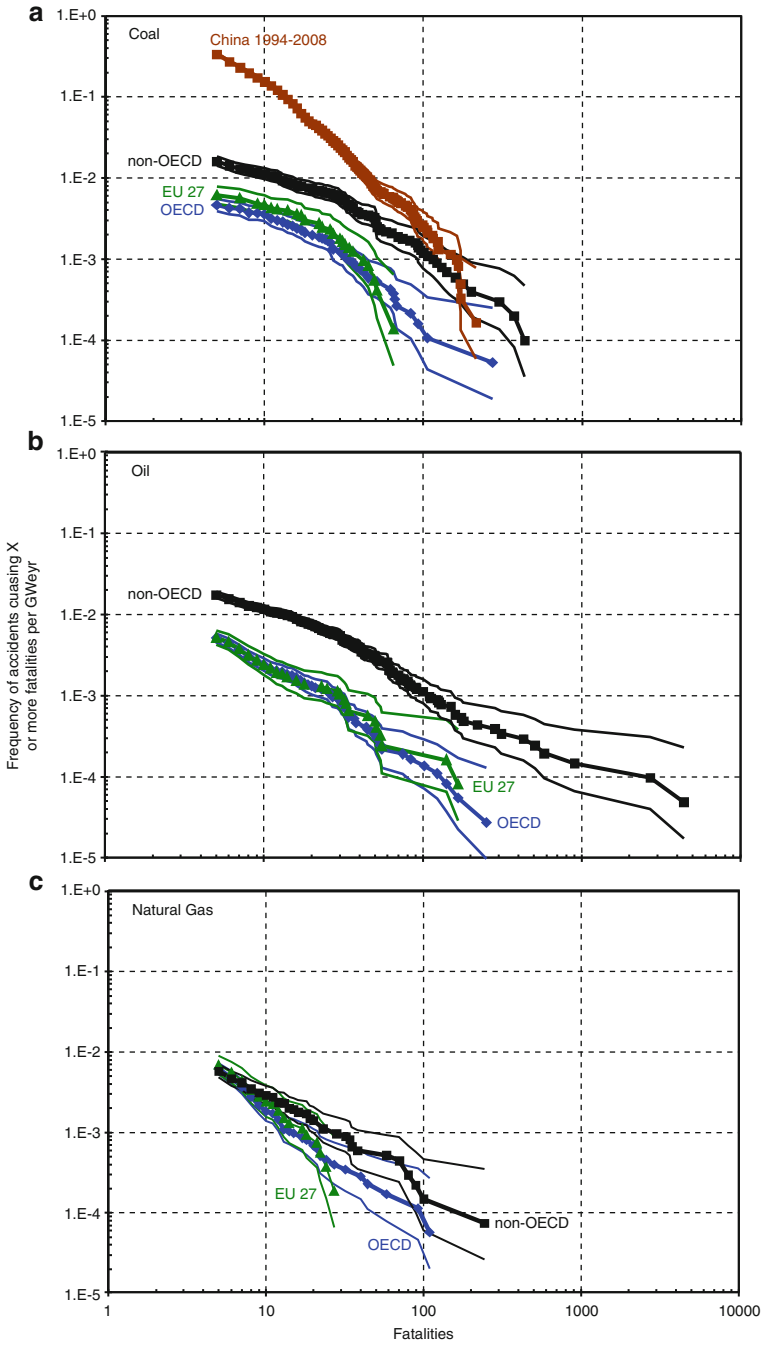

Fig. 18.5 Comparison of F-N curves based on historical experience of severe (≥ 5 fatalities) accidents in (**a**) coal, (**b**) oil, and (**c**) natural gas chains of different country groups for the period 1970–2008, except for coal China 1994–2008

attributable to different assumptions and methods used. For example, [100] estimated an additional 130 (range 15–1,100) worldwide cancer-related latent fatalities, whereas [106] reported a much higher value of 10,000, and some rather extreme sensitivity cases up to 300,000 latent fatalities. According to current knowledge, new generation III reactors are expected to have significantly lower fatality rates than currently operating power plants because of various safety augmenting systems, but maximum consequences could increase due to the much larger radioactive inventory of a 1,600 MW EPR, as indicated by the results of a simplified level-3 PSA [17, 19, 88].

Finally, new renewable technologies exhibit distinctly lower fatality rates than fossil chains, and are fully comparable to hydro and nuclear in highly developed countries. Concerning maximum consequences, renewables clearly outperform all other technologies because their decentralized nature strongly limits their catastrophic potential. However, it should be noted that current analyses of risks associated with new renewables have limited scope and do not include probabilistic modeling of hypothetical accidents. This may have bearing, particularly on results for solar PV.

18.3.4 External Costs of Accidents

In Figs. 18.7 and 18.8 external costs of severe ($\geq$5 fatalities) accidents were estimated for fossil energy chains [period 1970–2008, except coal China (1994–2008)]. Assumptions for external cost calculation were based on the results of the EU project NewExt (for details see [16]). Table 18.5 provides an overview of the central (median), minimum, and maximum values used for the values of statistical life (VSL), as well as the degrees of internalization for occupational and public fatalities in OECD, EU 27, and non-OECD countries.

Table 18.5 Summary of central, minimum (Min), and maximum (Max) values for value of statistical life (VSL) and degree of internalization for occupational and public fatalities used in external cost calculations

	Million $EUR_{(2002)}$			Degree of internalization					
	VSL			Occupational fatalities			Public fatalities		
Region	Central	Min	Max	Central	Min	Max	Central	Min	Max
OECD/EU 27	1.045	0.400	3.310	0.8	0.7	1.0	0.5	0.3	0.7
Non-OECD	1.045	0.400	3.310	0.5	0.0	1.0	0.2	0.0	0.5

Central estimates of external costs for OECD and EU 27 countries were rather similar, whereas non-OECD countries were significantly higher. Non-OECD values for coal and oil were about one order of magnitude higher, and in the case of coal China even two, whereas natural gas non-OECD was roughly 50% higher. When looking at sensitivities, it is evident that changes in VSL have the strongest effect on

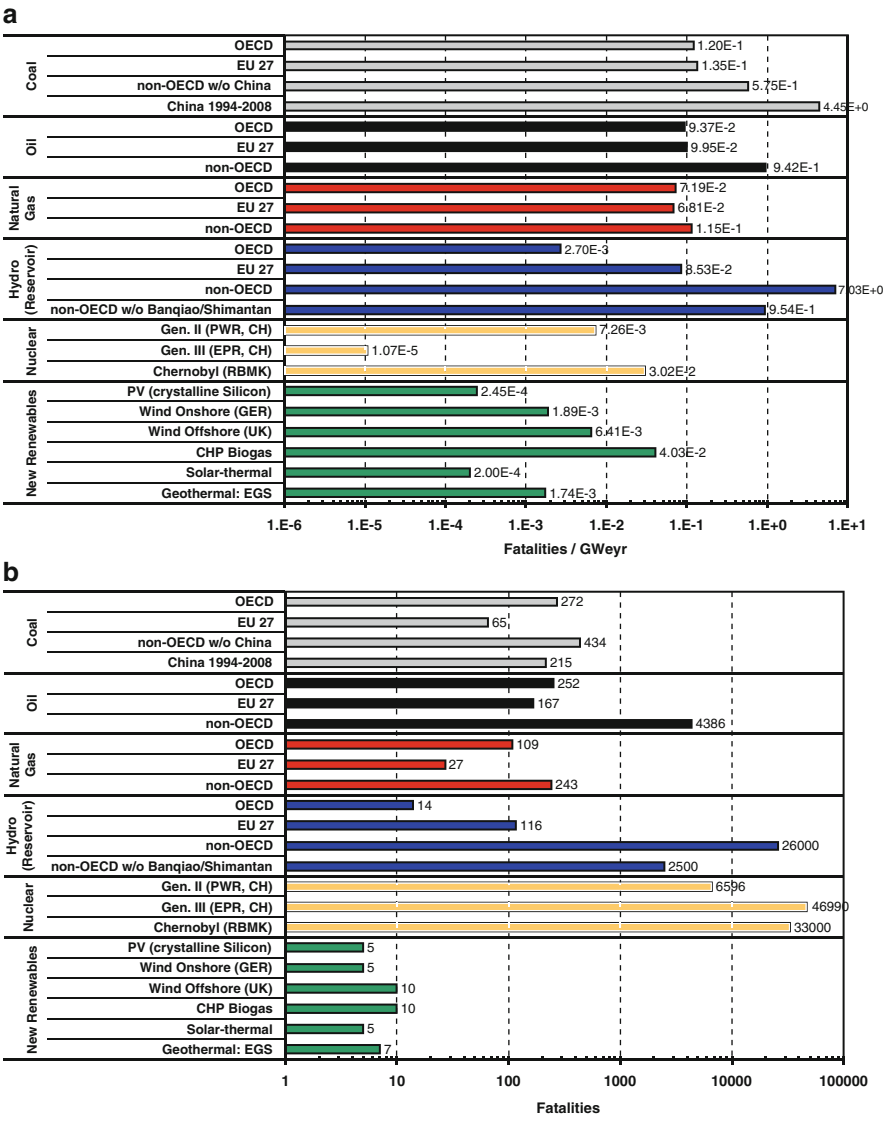

Fig. 18.6 Comparison of (**a**) fatality rates and (**b**) maximum consequences of a broad selection of energy technologies. Fossil and hydropower are based on the ENSAD database [period 1970–2008, except coal China (1994–2008)]; for nuclear a simplified level-3 PSA is applied; and for other renewable sources a hybrid approach using available data, modeling, and expert judgment is used. Abbreviations: *PWR* pressurized water reactor; *EPR* European pressurized reactor; *RBMK* reactor bolshoy moshchnosty kanalny, a water-cooled graphite-moderated reactor; *PV* photovoltaic; *CHP* combined heat and power; *EGS* enhanced geothermal systems

external costs, whereas the degree of internalization for occupational fatalities is intermediate and for public fatalities smallest. This difference between the two fatality

categories was most pronounced for the coal chain, where mining accidents clearly dominate, whereas for oil and natural gas it was less distinct because particularly during transportation stages public fatalities can make a substantial contribution. In the case of hydropower (not shown), external costs are almost exclusively dominated by public fatalities because dam failures primarily affect downstream settlements, i.e., the general public. Therefore, occupational fatalities are practically negligible because of the limited number of staff needed for operation of a hydro dam. However, accidents during the construction phase of a dam can lead to many occupational fatalities as in the case of the Mattmark dam (Switzerland) in 1965 when an ice-avalanche catastrophe caused the death of 88 workers, or the Guavio dam (Colombia) in 1983 when torrential rains led to mudslides, burying and killing 160 workers changing shifts at the dam [13]. PSA-based external costs for a Swiss nuclear power plant (not shown) have been estimated at 1.2E−3 USD-cent/kWhe, with 5% and 95% percentiles at 1.0E−4 and 3.8E−3 USD cent/kWhe, respectively, by [52].

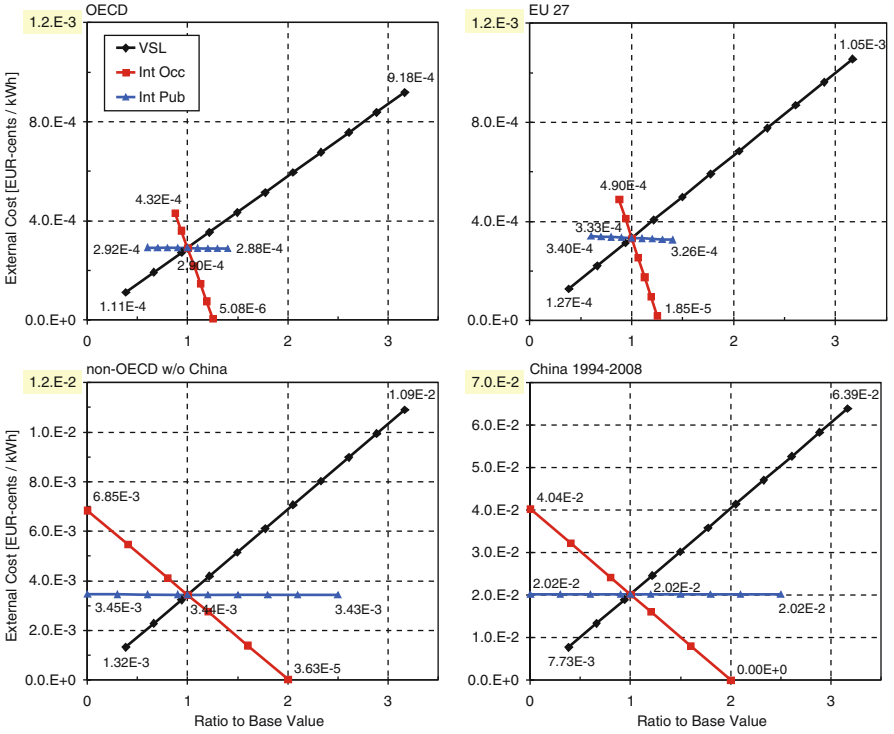

Fig. 18.7 External costs in EUR cents per kWh of severe ($\geq$ 5 fatalities) accidents in the coal chain (1970–2008, except China, 1994–2008) for OECD, EU 27, and non-OECD countries. In addition to central values, sensitivities are also shown for VSL (value of statistical life) and different degrees of internalization (Int) for occupational (Occ) and public (Pub) fatalities. Since maximum y-axis values are partially different, they are shaded in *light yellow*

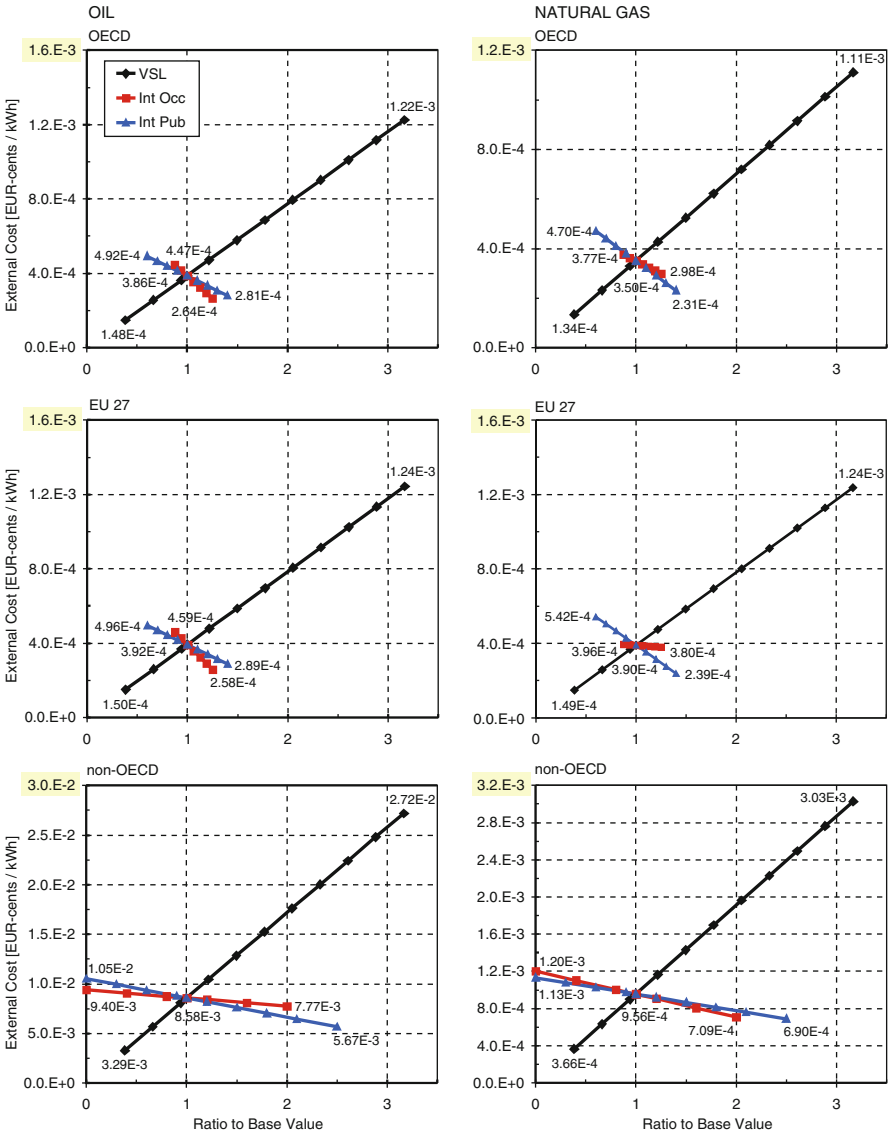

Fig. 18.8 External costs in EUR cents per kWh of severe ($\geq$ 5 fatalities) accidents in the oil and natural gas chains (1970–2008) for OECD, EU 27, and non-OECD countries. In addition to central values, sensitivities are also shown for VSL (value of statistical life) and different degrees of internalization (Int) for occupational (Occ) and public (Pub) fatalities. Since maximum y-axis values are partially different, they are shaded in *light yellow*

18.4 Additional Risk Aspects

While in the previous section the focus was on quantifiable risk indicators, Table 18.6 provides an overview of additional risk aspects that are not amenable to full quantification yet because only limited data and experience are available or they cannot be fully covered by traditional risk indicators that focus mainly on consequences. For a more in-depth discussion and a list of key references we refer to the IPCC special report on renewable energy sources and climate change mitigation, Chapter 9.3.4.7 on accidents and risks [10].

Table 18.6 Overview of selected additional risk aspects for various energy technologies

Risk aspect	Affected technologies
New resource developments	Deep offshore oil resources (e.g., Gulf of Mexico, Brazil)
	Oil resources in extreme environments (e.g., Arctic)
Induced seismicity, subsidence	Oil and gas production, coal mining
	Hydropower reservoirs
	Enhanced geothermal systems (EGS)
	Carbon capture and storage (CCS)
Resource competition	Bioenergy (e.g., food vs. fuel; water resources)
	Hydro reservoir (electricity vs. irrigation vs. supply)
Hazardous substances	Explosive, flammable, toxic, and asphyxiant substances in PV module production
	Spills of chemicals via hydraulic fracturing (shale gas, geothermal) can lead to groundwater contamination
Climate effects	Large wind deployment could locally increase lower atmosphere temperature
Long-term storage (public acceptance)	Disposal of nuclear waste and CCS
Proliferation	Nuclear energy
Geopolitics, terrorist threat, capture/ransom	Large renewable capacities in geopolitically less stable regions
	Intentional attacks on energy infrastructure
	Pirate attacks on oil/gas tankers

18.5 Conclusions and Recommendations

- The comparative risk assessment approach developed by PSI provides a comprehensive framework for the evaluation of severe accidents in the energy sector, with the database ENSAD as the core element.

- The evaluation of a broad set of risk indicators is essential because stakeholders and decisionmakers may not only focus on objective risks (e.g., fatality rates), but also consider subjective risk aversion (e.g., maximum consequences) and/or other risk aspects that are difficult to quantify.
- External costs of severe accidents are rather insignificant when compared to the effects of global warming and air pollution [13].
- Nevertheless, energy-related accidents have the potential for disastrous consequences on human health and the environment as well as society and economy, and thus they should be considered in a holistic evaluation of sustainability and energy security concerns.
- In summary, no technology performs best or worst in all respects, i.e., trade-offs and compromises are necessary to ensure a sustainable and secure energy supply.

References

1. Agarwal R, Ramakrishnan V (2010) Epanechnikov kernel estimation of value at risk. Soc Sci Res Network. SSRN 1537087. doi:10.2139/ssrn.1537087. http://dx.doi.org/10.2139/ssrn.1537087
2. Asfaw A, Mark C, Pana-Cryan R (2013) Profitability and occupational injuries in US underground coal mines. Accid Anal Prev 50:778–786
3. Atlas RM, Hazen TC (2011) Oil biodegradation and bioremediation: a tale of the two worst spills in US history. Environ Sci Tech 45:6709–6715
4. Aven T (2011) On some recent definitions and analysis frameworks for risk, vulnerability, and resilience. Risk Anal 31(4):515–522
5. Aven T (2012) On the link between risk and exposure. Reliab Eng Syst Safety 106:191–199
6. Aven T, Zio E (2011) Some considerations on the treatment of uncertainties in risk assessment for practical decision making. Reliab Eng Syst Safety 96:64–74
7. Bruneau M, Chang SE, Eguchi RT, Lee GC, O'Rourke TD, Reinhorn AM, Shinozuka M, Tierney K, Wallace WA, von Winterfeldt D (2003) A framework to quantitatively assess and enhance the seismic resilience of communities. Earthquake Spectra 19(4):733–752
8. Burgherr P (2007) In-depth analysis of accidental oil spills from tankers in the context of global spill trends from all sources. J Hazard Mater 140:245–256
9. Burgherr P (2009) Using GIS and multivariate analyses to visualize risk levels and spatial patterns of severe accidents in the energy sector. In: Martorell S, Guedes Soares C, Barnett J (eds) Safety, reliability and risk analysis: theory, methods and applications. Taylor & Francis Group, London
10. Burgherr P (2011) Accidents and risks (chapter 9.3.4.7). In: Sathaye J, Lucon O, Rahman A, Christensen J, Denton F, Fujino J, Heath G, Kadner S, Mirza M, Rudnick H, Schlaepfer A, Shmakin A 2011: Renewable energy in the context of sustainable energy (chapter 9). In: Edenhofer O, Pichs-Madruga R,

Sokona Y, Seyboth K, Matschoss P, Kadner S, Zwickel T, Eickemeier P, Hansen G, Schlömer S, von Stechow C (eds) IPCC special report on renewable energy sources and climate change mitigation. Cambridge University Press, Cambridge, United Kingdom and New York

11. Burgherr P, Hirschberg S (2005) Comparative assessment of natural gas accident risks. PSI Report No. 05–01. Paul Scherrer Institut, Villigen PSI
12. Burgherr P, Hirschberg S (2007) Assessment of severe accident risks in the Chinese coal chain. Int J Risk Assess Manag 7(8):1157–1175
13. Burgherr P, Hirschberg S (2008a) A comparative analysis of accident risks in fossil, hydro and nuclear energy chains. Hum Ecol Risk Assess 14(5):947–973
14. Burgherr P, Hirschberg S (2008b) Severe accident risks in fossil energy chains: a comparative analysis. Energy 33:538–553
15. Burgherr P, Hirschberg S (2009) Comparative risk assessment for energy systems: a tool for comprehensive assessment of energy security. In: Voeller JG (ed) Wiley handbook of science and technology for homeland security. John Wiley & Sons Inc., Hoboken
16. Burgherr P, Hirschberg S, Hunt A, Ortiz RA (2004) Severe accidents in the energy sector. In: Friedrich R (ed) Final report to the European Commission of the EU 5th framework programme "new elements for the assessment of external costs from energy technologies" (NewExt). DG research, technological development and demonstration (RTD), Brussels, Belgium
17. Burgherr P, Hirschberg S, Cazzoli E (2008) Final report on quantification of risk indicators for sustainability assessment of future electricity supply options. NEEDS Deliverable nř D7.1: research stream 2b. NEEDS project "new energy externalities developments for sustainability", Brussels, Belgium
18. Burgherr P, Eckle P, Hirschberg S (2010) Severe accidents in the context of energy security and critical infrastructure protection. In: Ale BJM, Papazoglou IA, Zio E (eds) Reliability, risk and safety. Taylor & Francis Group, London, pp 456–464
19. Burgherr P, Eckle P, Hirschberg S, Cazzoli E (2011) Final report on severe accident risks including key indicators. SECURE Deliverable No. D5.7.2a. In: SECURE project: security of energy considering its uncertainty, risk and economic implications, Brussels, Belgium
20. Burgherr P, Eckle P, Hirschberg S (2012a) Comparative assessment of severe accident risks in the coal, oil and natural gas chains. Reliab Eng Syst Safety 105:97–103
21. Burgherr P, Eckle P, Michaux E (2012b) Oil tanker transportation risk: driving factors and consequence assessment. In: 11th international probabilistic safety assessment and management conference and European safety and reliability conference 2012 (PSAM11 & ESREL2012), Helsinki, 25–29 June 2012
22. Carneiro FOM, Rocha HHB, Rocha PAC (2013) Investigation of possible societal risk associated with wind power generation systems. Renew Sust Energy Rev 19:30–36
23. Chen H, Qi H, Long R, Zhang M (2012) Research on 10-year tendency of China coal mine accidents and the characteristics of human factors. Safety Sci 50:745–750

24. Chernobyl Forum (IAEA, W., UNDP, FAO, UNEP, UN-OCHA, UNSCEAR, World Bank, Governments of Belarus, the Russian Federation and Ukraine) (2005) Chernobyl's legacy: health, environmental and socio-economic impacts and recommendations to the governments of Belarus, the Russian Federation and Ukraine. The chernobyl forum: 2003–2005. Second revised version. IAEA, Vienna
25. Cherp A, Jewell J (2011) The three perspectives on energy security: intellectual history, disciplinary roots and the potential for integration. Curr Opin Environ Sust 3:202–212
26. Chessel D, Dolédec S (1996) ADE version 4: hypercard stacks and quick basic Microsoft programme library for the analysis of environmental data, user's manual. In: Ecologie des Eaux Douces et des Grands Fleuves. Université Lyon 1, Lyon
27. Colli A, Serbanescu D, Ale BJM (2009) Indicators to compare risk expressions, grouping, and relative ranking of risk for energy systems: application with some accidental events from fossil fuels. Safety Sci 47:591–607
28. Cox LA Jr (2008) Some limitations of "risk = threat × vulnerability × consequence" for risk analysis of terrorist attacks. Risk Anal 28(6):1749–1761
29. Cozzani V, Campedel M, Renni E, Krausmann E (2010) Industrial accidents triggered by flood events: Analysis of past accidents. J Hazard Mater 175:501–509
30. Cruz AM, Krausmann E (2009) Hazardous-materials releases from offshore oil and gas facilities and emergency response following Hurricanes Katrina and Rita. J Loss Prevent Process Ind 22:59–65
31. Dadashzadeh M, Abbassi R, Khan F, Hawboldt K (2013) Explosion modeling and analysis of BP deepwater horizon accident. Safety Sci 57:150–160
32. Delzeit R, Holm-Müller K (2009) Steps to discern sustainability criteria for a certification scheme of bioethanol in Brazil: approach and difficulties. Energy 34:662–668
33. Eckle P, Burgherr P (2013) Bayesian data analysis of severe fatal accident risk in the oil chain. Risk Anal 33(1):146–160
34. Eckle P, Burgherr P, Hirschberg S (2011) Final report on multi-criteria decision analysis (MCDA). SECURE deliverable No. D6.2. In: SECURE project: security of energy considering its uncertainty, risk and economic implications, Brussels, Belgium
35. Eckle P, Burgherr P, Michaux E (2012) Risk of large oil spills: a statistical analysis in the aftermath of deepwater horizon. Environ Sci Tech 46(23):13002–13008
36. Elahi S (2011) Here be dragons… exploring the 'unknown unknowns'. Futures 43:196–201
37. Ersdal G, Aven T (2008) Risk informed decision-making and its ethical basis. Reliab Eng Syst Safety 93:197–205
38. Fabiano B, Curro F (2012) From a survey on accidents in the downstream oil industry to the development of a detailed near-miss reporting system. Process Safety Environ Prot 90:357–367

39. Felder, F.A.: A critical assessment of energy accident studies. Energy Policy 37, 5744–5751 (2009)
40. Filippini R, Silva A (2012) Resilience analysis of networked systems-of-systems based on structural and dynamic interdependencies. In: PSAM 11 & ESREL 2012, Helsinki, p 10
41. Fritzsche AF (1989) The health risks of energy production. Risk Anal 9(4):565–577
42. Fritzsche AF (1992) Editorial - severe accidents: can they occur only in the nuclear production of electricity? Risk Anal 12:327–329
43. Fthenakis VM, Kim HC (2011) Photovoltaics: life-cycle analyses. Solar Energy 85(8):1609–1628
44. Garrick BJ (2008) Examples of risks having the potential for catastrophic consequences. In: Garrick BJ (ed) Quantifying and controlling catastrophic risks. Academic, Burlington, pp 203–230
45. Greenberg M, Truelove HB (2011) Energy choices and risk beliefs: is it just global warming and fear of a nuclear power plant accident? Risk Anal 31(5):819–831
46. Greenberg M, Haas C, Cox A Jr, Lowrie K, McComas K, North W (2012) Ten most important accomplishments in risk analysis, 1980–2010. Risk Anal 32(5):771–781
47. Gregory R, Lichtenstein S (1994) A hint of risk: tradeoffs between quantitative and qualitative risk factors. Risk Anal 14(2):199–206
48. Ha-Duong M, Loisel R (2011) Actuarial risk assessment of expected fatalities attributable to carbon capture and storage in 2050. Int J Greenhouse Gas Control 5:1346–1358
49. Haimes YY (2009) On the complex definition of risk: a systems-based approach. Risk Anal 29(12):1647–1654
50. Härtler G (1983) Statistische Methoden für die Zuverlässigkeitsanalyse. VEB Verlag Technik, Berlin
51. Hassler B (2011) Accidental versus operational oil spills from shipping in the Baltic Sea: risk governance and management strategies. Ambio 40:170–178
52. Hirschberg S, Spiekerman G, Dones R (1998) Severe accidents in the energy sector, 1st edn. PSI Report No. 98–16. Paul Scherrer Institut, Villigen PSI
53. Hirschberg S, Burgherr P, Spiekerman G, Cazzoli E, Vitazek J, Cheng L (2003) Assessment of severe accident risks. In: Eliasson B, Lee YY (eds) Integrated assessment of sustainable energy systems in China. The China energy technology program - a framework for decision support in the electric sector of Shandong province. Alliance for global sustainability series, vol 4. Kluwer Academic Publishers, Amsterdam, pp 587–660
54. Hirschberg S, Burgherr P, Spiekerman G, Dones R (2004) Severe accidents in the energy sector: comparative perspective. J Hazard Mater 111(1–3):57–65
55. Hollnagel E, Woods DW, Leveson N (2006) Resilience engineering: concepts and precepts. Ashgate, Burlington
56. Hope BK (2012) Exposure gone "wild": a call for rational exposure scenarios. Hum Ecol Risk Assess 18:485–487

57. Hotelling H (1933) Analysis of a complex of statistical variables into principal components. J Educ Psychol 24:417–441 and 498–520
58. Huang IB, Keisler J, Linkov I (2011) Multi-criteria decision analysis in environmental sciences: ten years of applications and trends. Sci Total Environ 409:3578–3594
59. Inhaber H (2004) Risk analysis applied to energy systems. In: Cleveland CJ (ed) Encyclopedia of energy, vol 5. Elsevier, Amsterdam, pp 469–482
60. International Risk Governance Council (2005) White paper on risk governance. Towards an integrative approach. International Risk Governance Council (IRGC), Geneva
61. Jernelöv A (2010) The threats from oil spills: now, then, and in the future. AMBIO 39:353–366
62. Johansson B (2013) A broadened typology on energy and security. Energy 53:199–205
63. Johnson S, Low-Choy S, Mengersen K (2011) Integrating Bayesian networks and geographic information systems: good practice examples. Integr Environ Assess Manag 8(3):473–479
64. Jolliffe IT (2002) Principal component analysis, 2nd edn. Springer series in statistics. Springer, New York
65. Kaiser MJ, Yu Y (2010) The impact of Hurricanes Gustav and Ike on offshore oil and gas production in the Gulf of Mexico. Appl Energy 87:284–297
66. Kalantarnia M, Khan F, Hawboldt K (2010) Modelling of BP Texas City refinery accident using dynamic risk assessment approach. Process Safety Environ Prot 88(3):191–199
67. Knapp S, Franses PH (2009) Does ratification matter and do major conventions improve safety and decrease pollution in shipping? Marine Policy 33:826–846
68. Kontovas CA, Psaraftis HN, Ventikos NP (2010) An empirical analysis of IOPCF oil spill cost data. Mar Pollut Bull 60:1455–1466
69. Koornneef J, Ramírez A, Turkenburg W, Faaij A (2012) The environmental impact and risk assessment of CO2 capture, transport and storage: an evaluation of the knowledge base. Progress Energy Combust Sci 38, 62–86
70. Krausmann E, Renni E, Campedel M, Cozzani V (2011) Industrial accidents triggered by earthquakes, floods and lightning: lessons learned from a database analysis. Nat Hazards 59:285–300
71. Kröger W, Zio E (2011) Vulnerable systems. Springer, London
72. Kruyt BDP, van Vuuren HJM, Vries d, Groenenberg H (2009) Indicators for energy security. Energy Policy 37(6):2166–2181
73. Lefèvre N (2010) Measuring the energy security implications of fossil fuel resource concentration. Energy Policy 38:1635–1644
74. Linkov I, Satterstrom FK, Kiker G, Batchelor C, Bridges T, Ferguson E (2006) From comparative risk assessment to multi-criteria decision analysis and adaptive management: recent developments and applications. Environ Int 32:1072–1093

75. Lirong W, Zhongan J, Weimin C, Xiuwei Z, Dawei L, Yujing Y (2011) Major accident analysis and prevention of coal mines in China from the year of 1949 to 2009. Mining Sci Tech (China) 21:693–699
76. Liu KF-R, Ko C-Y, Fan C, Chen C-W (2012) Combining risk assessment, life cycle assessment, and multi-criteria decision analysis to estimate environmental aspects in environmental management system. Int J Life Cycle Assess 17:845–862
77. Löschel A, Moslener U, Rübbelke DTG (2010) Indicators of energy security in industrialised countries. Energy Policy 38(4):1665–1671
78. Mabel MC, Raj RE, Fernandez E (2010) Adequacy evaluation of wind power generation systems. Energy 35:5217–5222
79. Maiti J, Khanzode VV, Ray PK (2009) Severity analysis of Indian coal mine accidents: a retrospective study for 100 years. Safety Sci 47:1033–1042
80. Morgan MG, Henrion M (1990) Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis. Cambridge University Press, New York
81. Paté-Cornell E (2012) On "Black Swans" and "Perfect Storms": risk analysis and management when statistics are not enough. Risk Anal 32(11):1823–1833
82. Pearson K (1901) On lines and planes of closest fit to systems of points in space. Philos Mag 2:559–572
83. Perrons RK (2013) Assessing the damage caused by Deepwater Horizon: not just another Exxon Valdez. Mar Pollut Bull 71(1-2):20–22
84. Psarros G, Skjong R, Vanem E (2011) Risk acceptance criterion for tanker oil spill risk reduction measures. Mar Pollut Bull 62:116–127
85. Rasmussen NC (1981) The application of probabilistic risk assessment techniques to energy technologies. Annu Rev Energy 6:123–138
86. Ren J, Jenkinson I, Wang J, Xu DL, Yang JB (2009) An offshore risk analysis method using fuzzy Bayesian network. J Offshore Mechanics and Arctic Eng 131:1–12
87. Renn O, Klinke A, van Asselt M (2011) Coping with complexity, uncertainty and ambiguity in risk governance: a synthesis. Ambio 40(2):231–246
88. Roth S, Hirschberg S, Bauer C, Burgherr P, Dones R, Heck T, Schenler W (2009) Sustainability of electricity supply technology portfolio. Ann Nuclear Energy 36:409–416
89. Saleh JH, Cummings AM (2011) Safety in the mining industry and the unfinished legacy of mining accidents: safety levers and defense-in-depth for addressing mining hazards. Safety Sci 49:764–777
90. Schenler W, Hirschberg S, Burgherr P, Makowski M (2009) Final report on sustainability assessment of advanced electricity supply options. NEEDS deliverable nř D10.2 - research Stream 2b. In: NEEDS project "new energy externalities developments for sustainability", Brussels, Belgium
91. Scholtens B, Boersen A (2011) Stocks and energy shocks: the impact of energy accidents on stock market value. Energy 36:1698–1702

92. Scouras J, Parnell GS, Ayyub BM, Liebe RM (2009) Risk analysis frameworks for counterterrorism. In: Voeller JG (ed) Wiley handbook of science and technology for homeland security. Wiley, Hoboken
93. Skogdalen JE, Vinnem JE (2012) Quantitative risk analysis of oil and gas drilling, using Deepwater Horizon as case study. Reliab Eng Syst Safety 100:58–66
94. Sornette D, Ouillon G (2012) Dragon-kings: mechanisms, statistical methods and empirical evidence. European Phys J Special Topics 205:1–26
95. Sovacool BK (2008) The costs of failure: a preliminary assessment of major energy accidents, 1907–2007. Energy Policy 36:1802–1820
96. Sovacool BK (2013) An international assessment of energy security performance. Ecol Econ 88:148–158
97. Stirling A (1999) Risk at a turning point? J Environ Med 1:119–126
98. Sutton I (2012) Major events. In: Sutton I (ed) Offshore safety management. Elsevier, Amsterdam, pp 44–81
99. Szklo A, Schaeffer R (2007) Fuel specification, energy consumption and CO2 emission in oil refineries. Energy 32:1075–1092
100. Ten Hoeve JE, Jacobson MZ (2012) Worldwide health effects of the Fukushima Daiichi nuclear accident. Energy Environ Sci 5:8743–8757
101. Thioulouse J, Chessel D, Dolédec S, Olivier JM (1997) ADE-4: a multivariate analysis and graphical display software. Stat Comput 7:75–83
102. Tu J (2007) Coal mining safety: China's Achilles' heel. China Secur 3(2):36–53
103. Umbach F (2010) Global energy security and the implications for the EU. Energy Policy 38:1229–1240
104. Vinnem JE (2010) Risk indicators for major hazards on offshore installations. Safety Sci 48:770–787
105. Vinnem JE (2011) Evaluation of offshore emergency preparedness in view of rare accidents. Safety Sci 49:178–191
106. Vitázková J, Cazzoli E (2011) Estimate of consequences from the Fukushima disaster. In: Nordic PSA Conference, Stockholm
107. Vivoda V (2010) Evaluating energy security in the Asia-Pacific region: a novel methodological approach. Energy Policy 38:5258–5263
108. von Hippel D, Suzuki T, Williams JH, Savage T, Hayes P (2011) Energy security and sustain-ability in Northeast Asia. Energy Policy 39:6719–6730
109. Winzer C (2012) Conceptualizing energy security. Energy Policy 46:36–48
110. Yustaa JM, Correa GJ, Lacal-Arántegui R (2011) Methodologies and applications for critical infrastructure protection: state-of-the-art. Energy Policy 39:6100–6119
111. Zoback MD, Gorelick SM (2012) Earthquake triggering and large-scale geologic storage of carbon dioxide. PNAS 109(26):10164–10168

Index

R.M. Kovacevic et al. (eds.), *Handbook of Risk Management in Energy Production and Trading*, International Series in Operations Research & Management Science 199,
DOI 10.1007/978-1-4614-9035-7, © Springer Science+Business Media New York 2013