

Chapter 7

Predictive Analytics in Future Power Systems: A Panorama and State-Of-The-Art of Deep Learning Applications



Sakshi Mishra, Andrew Glaws, and Praveen Palanisamy

Abstract The challenges surrounding the optimal operation of power systems are growing in various dimensions, due in part to increasingly distributed energy resources and a progression towards large-scale transportation electrification. Currently, the increasing uncertainties associated with both renewable energy generation and demand are largely being managed by increasing operational reserves—potentially at the cost of suboptimal economic conditions—in order to maintain the reliability of the system. This chapter looks at the big picture role of forecasting in power systems from generation to consumption and provides a comprehensive review of traditional approaches for forecasting generation and load in various contexts. This chapter then takes a deep dive into the state-of-the-art machine learning and deep learning approaches for power systems forecasting. Furthermore, a case study of multi-time-horizon solar irradiance forecasting using deep learning is discussed in detail. Smart grids form the backbone of the future interdependent networks. For addressing the challenges associated with the operations of smart grid, development and wide adoption of machine learning and deep learning algorithms capable of producing better forecasting accuracies is urgently needed. Along with exploring the implementation and benefits of these approaches, this chapter also considers the strengths and limitations of deep learning algorithms for power

S. Mishra (✉)

Integrated Applications Center, National Renewable Energy Laboratory, Golden, CO, USA
e-mail: sakshi.mishra@nrel.gov

A. Glaws

Computational Science Center, National Renewable Energy Laboratory, Golden, CO, USA
e-mail: andrew.glaws@nrel.gov

P. Palanisamy

Perception Planning and Decision Systems, General Motors, Warren, MI, USA
e-mail: praveen.palanisamy@gm.com

© Springer Nature Switzerland AG 2020

M. H. Amini (ed.), *Optimization, Learning, and Control for Interdependent Complex Networks*, Advances in Intelligent Systems and Computing 1123,
https://doi.org/10.1007/978-3-030-34094-0_7

147

systems forecasting applications. This chapter, thus, provides a panoramic view of state-of-the-art of predictive analytics in power systems in the context of future smart grid operations.

Keywords Smart grid · Deep learning · Predictive analytic · Machine learning · Time series · Energy forecast · Power systems

7.1 Introduction

Overview The challenges surrounding the optimal operation of power systems are growing in various dimensions, due in part to increasingly distributed energy resources and a progression towards large-scale transportation electrification. Currently, the increasing uncertainties associated with both renewable energy generation and demand are largely being managed by increasing operational reserves—potentially at the cost of suboptimal economic conditions—in order to maintain the reliability of the system. This chapter looks at the big picture role of forecasting in power systems from generation to consumption and provides a comprehensive review of traditional approaches for forecasting generation and load in various contexts. This chapter then takes a deep dive into the state-of-the-art machine learning and deep learning approaches for power systems forecasting. Furthermore, a case study of multi-time-horizon solar irradiance forecasting using deep learning is discussed in detail. Smart grids form the backbone of the future interdependent networks. For addressing the challenges associated with the operations of smart grid, development and wide adoption of machine learning and deep learning algorithms capable of producing better forecasting accuracies is urgently needed. Along with exploring the implementation and benefits of these approaches, this chapter also considers the strengths and limitations of deep learning algorithms for power systems forecasting applications. This chapter, thus, provides a panoramic view of state-of-the-art of predictive analytics in power systems in the context of future smart grid operations.

Forecasting has long played an essential role in power systems planning and operations. With the introduction of deregulated markets, forecasting has emerged as a critical component of electricity markets as well. Reliable forecasting models allow electrical utilities and independent systems operations (ISOs) to make optimal capacity building and dispatch decisions by understanding their economic implications while still maintaining a reliable energy supply. Forecasting models are also used by the market participants to place strategic bids. The significance of forecasting has dramatically increased because of the rapidly changing landscape of traditional power systems. Some of the main drivers of this change are (1) increasing penetration of intermittent renewable energy resources on utility scale as well as distributed energy resources (DERs), (2) deployment of various smart grid technologies such as advanced metering infrastructure, (3) deregulation of electricity markets, (4) demand response programs turning static loads into dynamic

loads, (5) forthcoming electrification of the transportation fleet, (6) greenhouse gas reduction targets, and (7) declining costs of energy storage technologies, among others.

The forecasted data in power systems include meteorological variables such as solar irradiance, wind speed, and wind direction; energy production from renewable energy sources such as photovoltaic plants, wind farms, and hydroelectric dams; load or demand; price of electricity or locational marginal prices; price of fossil fuels such as coal, oil, and natural gas; electric vehicle (EV) charging loads, and so on. These quantities are forecasted for different timescales as well as different spatial resolutions. Long-term forecasts are useful for power systems infrastructure building decisions while short-term forecasts are utilized to inform optimal decision-making by system operators dispatching energy on the grid and market participants trading energy in the markets.

7.1.1 Motivation

Traditionally, in the regulated electricity sector, mostly vertically integrated utilities had a monopoly. The reliability of supply was primarily the utilities' responsibility and was maintained using short-term load forecasts. The fossil fuel-based generation sources were dispatchable so that variability associated with the demand was the primary source of uncertainty in the system. Electricity users were passive consumers; that is, there was neither a bidirectional flow of energy from the distribution grid end nor any provision of demand response. Planning and investment in new capacity were based on long-term demand forecasts and utilities were responsible for building the transmission capacity to serve their customers. Traditional forecasting methodologies served well in this regulated business scenario.

Competitive electricity markets have been introduced since the last decade of the twentieth century as a part of the deregulation of the electricity sector [1]. Consequently, energy is now traded in competitive markets, making electricity price and demand forecasts fundamental inputs to the day-to-day decision-making process of the various energy-selling entities, including the utilities, independent power producers, large industrial customers with significant amounts of distribution generation production, and so on.

Moreover, building new transmission capacity is not a straightforward decision made by a single utility anymore. FERC Order No. 1000 [2] established new rules regarding the transmission planning and cost allocation requirement for public utility transmission providers, which have made capacity expansion a competitive process as well. As a result, accurate long-term load forecasting for different geographic areas has become even more important for maintaining the reliability of the system and economically expanding the network to accommodate future demand growth as well as distributed generation penetration on the grid.

The shift toward a digital and electrified economy is causing increased research and planning for networks of electrified transportation and a smart grid, operating

interdependently. Forecasting will play an essential role in the transition to this new system as well. These future interdependent networks require reliable long-term EV growth forecasts for the planning of EV charging infrastructure as well as distribution network enhancements to accommodate the high penetrations of dynamic EV charging loads. In the operations domain, short-term forecasts of EV charging or discharging are required to get accurate load forecasts. Additional complexity is added when daily EV charging profiles are optimized using intelligent controls. The operational schedule of EV charging responds to the market price (even without the initiation of demand response events from the utility), making it even more challenging for traditional forecasting approaches to predict the dynamically changing demand.

The macrogrid in the USA (as well as many other industrialized countries) is a century old. Various components of generation, transmission, and distribution systems are reaching the end of their useful life and need to be refurbished. Though there are significant capital costs necessary to renovate the thousands of miles of distribution infrastructure, the reliability threats are even more dire. Several recent wildfires in California can be attributed to the aging power systems infrastructure of Pacific Gas and Electric [3]. The remaining useful life of the assets can be assessed to strategically plan the renovation of aging power systems infrastructure by leveraging advanced machine learning and deep learning-based predictive analytics. Accurate remaining useful life predictions for distribution grid components can inform economic investment such that the components with the highest risk of failure are replaced first.

Growing uncertainty in energy consumption, increasing penetration of intermittent renewable energy generation sources (at both utility scale and for small DERs), the burgeoning share of microgrid deployment, smart grid technologies enabling the internet of things (IoT), aging grid infrastructure, and the forthcoming revolution of electrified transportation are rapidly changing the landscape of power systems. Advanced and innovative predictive analytics approaches are urgently needed to enable more accurate forecasts to improve decision-making and provide the foundation for a smart, resilient, and sustainable grid of the future.

7.1.2 Classification of Power Systems Forecasting Models

Power systems forecasts may be done for various timescales based on the application of the predicted data. These forecasts can also be classified based on their application domain and their role in the power systems generation, transmission, distribution, and consumption areas. When a quantity is forecasted for different time horizons, the input variables used for producing the forecast also change. For example, when forecasting load for a long-term horizon in a given geographical area, inputs such as macroeconomic uncertainty, population growth, climate change patterns (for predicting the extreme loads), and distributed generation penetration projections are considered. For short-term load forecasting like day-

ahead forecasting, input variables such as day of week, time of day, and past load consumption are relied upon instead. The effectiveness of machine learning and deep learning algorithms varies based on the timescale and type of input variables. The classification of forecasting models is discussed in detail in the following sections.

7.1.2.1 Classification Based on the Domain of Application in Power Systems

There are three functionally different parts of power systems studies and management, which make it possible to provide reliable and economical electricity to consumers in the present and in the future. These three parts—planning, operations, and market—are described in the following sections. Different types of predictive models are used in these three parts for obtaining forecasts for different quantities, such as load, resource, production, and so on, as shown in Fig. 7.1.

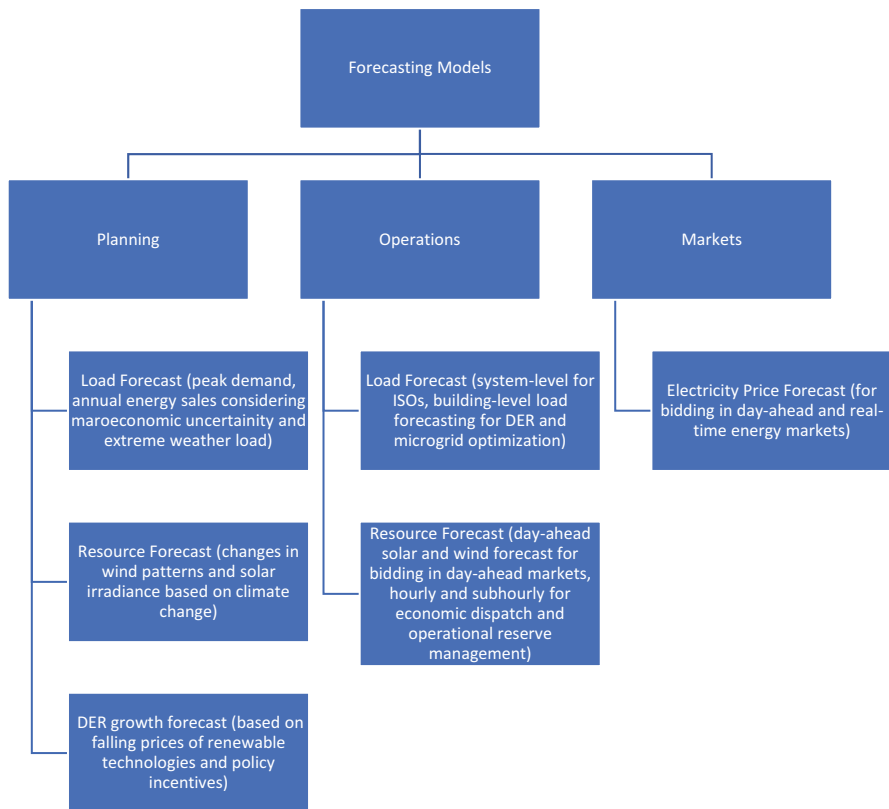


Fig. 7.1 Classification of forecasting models based on domain of applicability

7.1.2.1.1 Planning

The process of power systems planning is ever evolving and has the largest strategic impact on the future of power systems. Future power systems are planned using a set of forecasts, including *load forecasting* (which in turn depends on macroeconomic uncertainty, extreme weather, and climate changes for a given geographic location), *distributed generation technologies growth predictions* (which depend on the rate of decrease in the cost of these technologies and energy policy that provides incentives), and *resource forecasting* (which includes short-term production forecasts as well as long-term changes in solar irradiance and wind patterns for a given geographic area).

Smart and clean energy technologies form the foundation of the future smart grid. The key to enabling the adoption of clean energy technologies lies in how well power systems enhancements are planned to accommodate new technologies, enabling their smooth integration with the existing power systems. The goal of planning is to build and modify the generation, transmission, and distribution infrastructure that are needed to meet predicted future needs. Therefore, power systems planning has traditionally been divided into centralized generation planning, transmission planning, and distribution planning. The outcome of planning studies is to address what to build (more generation or transmission/distribution), how much to build, and where to build.

Traditionally, generation planning begins with load forecasting. Reliability evaluation is then conducted to determine if and when additional generation is needed. The remaining useful life of existing base load plants, which are largely powered by fossil fuels, is also accounted for in the next step. This is followed by capacity expansion studies based on economic considerations [4]. Nowadays, however, generation planning is not a solitary process. DER penetration forecasts, including behind-the-meter distributed generation, need to be accounted for in the process. Also, economical siting of utility-scale renewable generation plants depends on availability of solar and wind resources, which may or may not coincide with the demand pockets and existing transmission infrastructure. High penetration of utility-scale renewable energy resources, given their intermittent and variable nature, adds increased complexity to generation planning studies that depend on renewable resource forecasts [5, 6].

Transmission planning is aimed at optimizing the use of a generation portfolio by supplying loads from the most economical sources of power and improving the reliability of overall systems by operating generation stations flexibly [7]. Generation and transmission planning are closely related because the powerflows through the transmission system are a direct result of generation dispatch [8, 9]. Distribution system planning, on the other hand, is optimized for the lowest cost operation that meets the desired reliability of the electricity service. However, the introduction and increased adoption of DERs has changed the process of distribution system planning drastically. This is because components of distribution and transmission systems are not designed to handle the bidirectional flow of power from the DERs, so additional measures must be taken to refurbish the distribution grid with this capability [10, 11].

For generation and transmission planning, load forecasting is done for a long time horizon—often between 2 and 10 years. This is because system capacity expansion projects require a long lead time. Peak annual demand/load (in kilowatts) and total annual energy sales (in kilowatt-hours) are calculated for long-term load forecasts [12]. Peak load is highly correlated with weather. Therefore, peak load forecast is normalized based on extreme weather predictions. Projected EV and DER growth in the future has led to researching and employing methodologies that explicitly consider DERs as well as EV load along with its charging patterns [13, 14]. Load forecasting also needs to be specific to geographical locations, along with maintaining reasonable accuracies of the predicted magnitude.

7.1.2.1.2 Operations

Power systems operations are associated with making decisions regarding the use of existing equipment and infrastructure to generate, transmit, and deliver energy. It is primarily aimed at doing so safely, reliably, and efficiently. The operations domain deals with three different time horizons: (1) operations planning (a few weeks to months), (2) near real time (a few hours to days), and (3) real time (typically 5–10 min) [15].

Operations planning ensures that sufficient resources are available to meet demand for the next few months. It takes load forecasts (and associated errors), utility-scale renewable generation forecasts, and generation and transmission outages into account. Operations planning also defines the reserve capacity requirements to mitigate the risk imbalances because of forecast errors and unplanned outages of generation or transmission components [16]. The aim of near real-time operations is to select the most economic generation portfolio for the next few days using a process called unit commitment. Real-time operations are aimed at ensuring system reliability and supply sufficiency by revising the near real-time schedule on an as-needed basis.

Load forecasting is the first step of all three time horizons of power systems operations, making it a critical component. For the operations planning and near real-time applications, hourly load forecasts are used. For real-time applications, however, subhourly (minute-level) resolution is typically required. Once the magnitude and geographic location of demand are obtained using load forecasts, least-cost generation is scheduled to meet that demand. The production forecasts of utility-scale renewable generation plants are also considered while scheduling the generation. In the regions with high DER penetration, their production is also considered; behind-the-meter DERs are typically considered negative load.

7.1.2.1.3 Markets

The landscape of the power sector has substantially changed after the introduction of competitive markets coupled with the deregulation of the industry. This has led

to the trading of electricity under market rules using spot and derivative contracts. But the price dynamics of this unique commodity is different from any other commodity because of its unique properties, requirements, and dependencies. For example, energy typically experiences a constant balance between production and consumption because large quantities are not economically storable. Additionally, power demand can depend on weather factors, such as temperature, precipitation, and wind speeds, and on the magnitude of activity in different sectors (i.e., holidays vs. workdays, weekdays vs. weekends, on-peak vs. off-peak hours).

Electricity prices in the wholesale market, therefore, exhibit seasonality at various timescales (daily, weekly, annually) as well as abrupt and brief price spikes. According to [17], “[t]he costs of over-/under-contracting and then selling/buying power in the balancing (or real-time) market are typically so high that they can lead to huge financial losses or even bankruptcy. Extreme price volatility, which can be up to two orders of magnitude higher than that of any other commodity or financial asset, has forced market participants to hedge not only against volume risk but also against price movements.”

Short-term electricity price forecasting is done for the day-ahead market, where the bids are submitted for the delivery of electricity during each load period, which can be hourly or subhourly. Medium-term time horizons are used for risk management and derivative pricing. These forecasts can either be point-forecasts or probability distributions of the prices. Long-term electricity price forecasts are done for planning and economic feasibility analysis of future power plants, establishing long-term power purchase agreements, forward capacity markets, seasonal capacity markets, financial transmission rights auctions, and so on. The time horizon can vary from months to years for such applications.

Renewable generation forecasts in the short term are also required for owners to bid in the market. ISOs need the production forecasts of intermittent energy sources to schedule the generation with sufficient reserves to minimize the risk of underproduction. To avoid financial losses associated with underbidding or overbidding, renewable generation plant owners need reasonably accurate forecasts of solar and wind resources [18].

For each time horizon, the choice of input variables plays a significant role in the effectiveness of the model for both traditional forecasting approaches as well as deep learning methods. For short-term forecasts, the daily and hourly variability must be considered. On the other hand, medium-term forecasting favors annual variations more than weekly ones. For long-term price forecasts, seasonality itself becomes irrelevant. Instead, long-term trends such as load-growth in a certain geographic area, large penetration of cheap renewable energy resources in close proximity, and EV load demand play the major role.

7.1.2.2 Classification Based on Timescale

In the previous section, various power systems forecasting models were discussed in the context of their applicability to planning, operation, and market domains.

Table 7.1 Types of forecasting models based on timescale

Forecasting type	Time horizon	Applications	Methods
Nowcasting	5 min–6 h	Load frequency control, battery-use optimization, real-time market participation, economic dispatch	Satellite-based physical models (use cloud motion vector-based method); sky imagery-based physical models; statistical/machine learning models based on historical data (e.g., persistence, ARMA, SVRs, deep learning)
Short-term forecasting	6 h–1 week	Unit commitment, switching source, rescheduling means of production, day-ahead market participation	Hybrid NWP/statistical/machine learning models
Medium-term forecasting	1 week–2 year	Scheduling maintenance, capacity markets bidding, and pricing	Statistical models based on predicted growth/change (e.g., end-use, econometric models)
Long-term forecasting	2 year–5 year	Long-term purchase agreements, forward capacity market, management of multiyear reservoirs, nuclear fuel management	Statistical models based on predicted growth/change (e.g., end-use, econometric models)
Very long-term forecasting	25+ year	Capacity expansion, infrastructure retirement, policymaking	Statistical models based on predicted growth/change (e.g., end-use, econometric models); general circulation models (climate models)

Another way of classifying the forecasting models in power systems is based on the timescale for which the quantities are being forecasted. These timescales can mainly be classified into five types, as given in Table 7.1.

7.1.3 Organization of the Chapter

The introduction section first lays out the motivation behind exploring newer approaches such as deep learning for power systems predictive analytics. The power systems forecasting problems are then classified in broad categories based on

timescale of forecasting as well as their application in power systems planning, operations, and market domains. The second section, *forecasting power systems using classical approaches*, takes a deeper look at the widely used statistical times-series forecasting methods as well as traditional machine learning-based approaches. The third section then introduces state-of-the-art deep learning algorithms and explores their recent applications in the power systems forecasting literature. A solar irradiance forecasting case study is discussed in detail in the fourth section. The fifth section identifies future work areas in this domain and concludes the chapter.

7.2 Forecasting in Power Systems Using Classical Approaches

The power systems forecasting problems discussed in the previous section most closely align with the mathematical framework of the time series forecasting problem. This section introduces this general mathematical framework and provides a broad overview of several statistical and machine learning approaches to time series forecasting. Note that deep learning methods are left to Sect. 7.3 to be explored in more detail.

7.2.1 Time Series Data

A general time series dataset can be written as

$$\{x_1, x_2, x_3, \dots\}, \quad (7.1)$$

where each x_t for $t = 1, 2, 3, \dots$ represents the realization of some random variable. A common additive modeling approach to characterizing Eq. (7.1) is to partition the time series into a trend, seasonality, and stochastic term,

$$x_t = T_t + S_t + Z_t. \quad (7.2)$$

The trend term T_t represents the long-term, nonperiodic changes in the data, the seasonality term S_t describes any periodic behavior of the time series, and the stochastic term Z_t is stationary process (defined later) that models the random noise in the data.

Note that Eq. (7.1) frames the time series data in terms of scalar-valued quantities. This is done to simplify the discussion in this section in order to provide a clear and broad overview of traditional approaches to time series forecasting. The extension of this perspective to the multivariate case is relatively straightforward. One feature of multivariate time series data that is important to power systems mod-

eling is the concept of exogenous variables. In time series forecasting, exogenous variables are causally independent of other factors in the system. In the case of solar irradiance forecasting, examples of exogenous variables may include factors like wind speed and cloud cover. Including exogenous variables in the forecasting process may improve performance.

Recall that each x_t in Eq. (7.1) is a realization of some random variable. The complete time series is then fully characterized by the joint distribution of these random variables. However, such a perspective is typically impractical or impossible for real-world applications. A more reasonable approach is to characterize the time series in terms of secondary properties, such as the mean and covariance functions of the series,

$$\mu_t = \mathbb{E}[x_t] \quad \text{and} \quad \sigma_{t,s} = \mathbb{Cov}[x_t, x_s] = \mathbb{E}[(x_t - \mu_t)(x_s - \mu_s)]. \quad (7.3)$$

The dependence of the value of x_t on previous terms is characterized by the autocovariance function $\gamma_t(h) = \sigma_{t,t+h}$ and the autocorrelation function $\rho_t(h) = \gamma_t(h)/\gamma_t(0)$, where h is the lag parameter.

A key property of time series data is the idea of stationarity. A given time series is said to be strictly stationary if any two subseries,

$$\{x_t, x_{t+1}, x_{t+2}, \dots, x_{t+n}\} \quad \text{and} \quad \{x_s, x_{s+1}, x_{s+2}, \dots, x_{s+n}\} \quad \text{for } t, s, n \in \mathbb{N}, \quad (7.4)$$

have the same joint distribution. Notice that if each x_t in a given time series is independent and identically distributed (iid), then the time series is strictly stationary. Such a sequence drawn from a distribution with mean 0 and variance σ^2 is typically referred to as white noise.

As discussed earlier, characterizing the full joint distribution of a time series is not realistic for most real-world applications, making the identification of a time series as strictly stationary infeasible. Alternatively, a time series is said to be weakly (or wide-sense) stationary if any two subseries have the same mean and covariance functions, μ_t and $\sigma_{t,s}$, respectively. Equivalently, a weakly stationary time series has mean and covariance functions that are independent of t . That is, $\mu_t = \mu$ and $\sigma_{t,s} = \sigma$. Notice that this also implies that the autocovariance and autocorrelation functions only depend on the lag parameter, $\gamma_t(h) = \gamma(h)$ and $\rho_t(h) = \rho(h)$. Because this definition is of more practical use, it is common to use the term stationary to refer to weakly stationary and specifically refer to a time series as strictly stationary when the stricter definition is meant.

The next two sections explore various statistical and machine learning approaches to time series forecasting. In general, the goal of forecasting is to predict values of future datapoints $\{\hat{x}_{n+1}, \hat{x}_{n+2}, \dots\}$ given a finite set of observed data $\{x_1, x_2, x_3, \dots, x_n\}$. Generally, time series forecasting is classified according to the horizon out to which the forecast is made, as illustrated in Table 7.1.

The differences between short-, medium-, and long-term forecasts are highly dependent on the problem under consideration. However, short- and medium-

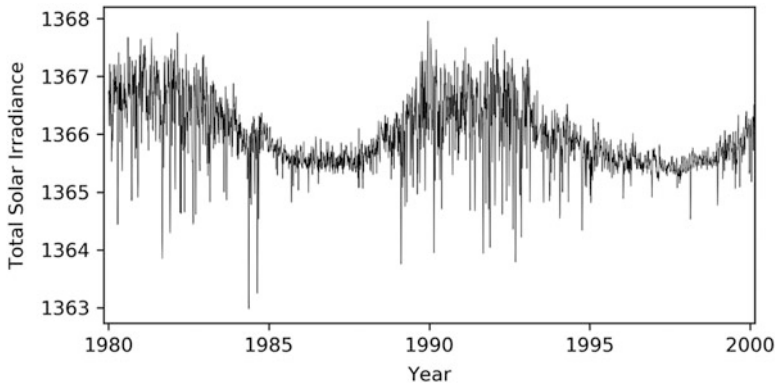


Fig. 7.2 Total solar irradiance measured daily from 1980 to 2000. These data come from the NOAA's National Centers for Environment Information database [19]

term forecasts typically are more dependent on autocorrelation factors and shorter seasonality behaviors. These prediction horizons tend to be more amenable to the types of data-driven methods covered here. Long-term forecasting seeks to model trends in the data and often depends on the additional models of the relevant systems to help predict changes in these trends.

Power systems forecasting is a particularly difficult problem. Figure 7.2 shows an example time series data of total solar irradiance over a 20-year range [19]. It is immediately obvious that this dataset is nonstationary (as is the case of many time series data arising from power systems). The data show seasonal cycles of increased and decreased solar irradiance that have a period of approximately 11 years. In addition to fluctuations in the mean of the data, the seasonality also changes the variance of the data. The time series varies more significantly during periods of high solar irradiance and less significantly during periods of low solar irradiance. Lastly, the example data in Fig. 7.2 highlight the differences in short-, medium-, and long-term forecasting. Short-term forecasts are focused on accurately capturing the high-frequency fluctuations in the data. Medium- and long-term predictions cannot hope to perfectly predict these behaviors and instead focus on the large-scale trends and seasonal characteristics in the time series.

7.2.2 Statistical Forecasting Approaches

7.2.2.1 Naïve Model Approach

The naïve model approach to time series forecasting simply predicts that the next value in the sequence is the same as the current value,

$$\hat{x}_{t+1} = x_t. \quad (7.5)$$

This approach produces the optimal prediction for random walk data and is therefore also known as the random walk model for time series forecasting. The main purpose of this model is to serve as a simple baseline to compare with more sophisticated models. This naïve model approach is also called the persistence model [20].

7.2.2.2 Exponential Smoothing

Exponential smoothing is a relatively simple approach to modeling time series that predicts new values in time series using a weighted moving average that more heavily favors recent datapoints [21]. Given time series data $\{x_1, x_2, x_3, \dots, x_n\}$, the simple exponential smoothing model computes the smoothed approximation of \hat{x}_{n+1} as

$$\hat{x}_{n+1} = \alpha x_n + (1 - \alpha) \hat{x}_n, \quad (7.6)$$

where $\alpha \in (0, 1)$ is the smoothing factor. Notice that this method computes a weighted average of the current true value and the current predicted (or smoothed) value. The current smoothed value was computed similarly. Thus, previous terms contribute to the current predict value with exponentially decreasing importance. The rate of this decay is controlled by the smoothing factor α . Extensions to simple exponential smoothing incorporate trends and seasonality [22, 23].

Simple exponential smoothing is among the earliest forecasting techniques applied to load forecasting [24]. In particular, this work explores the application of exponential smoothing to short-term forecasting at hourly intervals. More recently, several studies have explored the application of double seasonality exponential smoothing to short-term load forecasting and found this approach to be robust despite its relative simplicity [25, 26].

7.2.2.3 Autoregressive Moving Average (ARMA) Models

The autoregressive moving average (ARMA) model and its variations are powerful forecasting tools that are among the most popular statistical methods for power systems analysis. The ARMA model has long been used for power-related problems, such as solar irradiance and load forecasting [27, 28]. More recently, an ARMA variant called ARIMA (covered in the next section) has been applied to short-term solar forecasting [29, 30], next-day electricity pricing [31], and hourly load predictions [32].

As the name suggests, the ARMA model makes two key assumptions on the time series. The first is that the time series data can be modeled by an autoregressive process. An autoregressive process of order p , denoted by $AR(p)$, assumes a linear dependence of the current timestep on the previous p timesteps,

$$\hat{x}_t = \sum_{i=1}^p \omega_i x_{t-i} + z_t, \quad (7.7)$$

where ω_i are constants and z_t is a white noise term. The second assumption of the ARMA model is that of a moving average model. The moving average model of order q , denoted by $MA(q)$, represents the sequence as a linear relationship of some other white noise sequence,

$$\hat{x}_t = \sum_{i=1}^q \theta_i z_{t-i} + z_t, \quad (7.8)$$

where θ_i are constants and each z_t is an iid white noise term. The ARMA model of orders p and q , denoted by $ARMA(p, q)$, combines Eqs. (7.7) and (7.8) to form

$$\hat{x}_t = \sum_{i=1}^p \omega_i x_{t-i} + \sum_{i=1}^q \theta_i z_{t-i} + z_t. \quad (7.9)$$

The ARMA model is typically solved using the Box-Jenkins method [33]. This is an iterative process of specifying the model, fitting the parameters, and verifying the process. Specifying the model involves the order of the $ARMA(p, q)$ model (i.e., selecting the appropriate values of p and q). Heuristically, this can be accomplished by examining the autocorrelation function $\rho_t(h)$ and the partial autocorrelation function. Recall that the autocorrelation function explains the relationship between two terms with lag h . However, because this relationship can have a recursive structure, it may be difficult to distinguish between a time series that is dependent on the previous n points and one that is highly dependent only on the previous one. The partial autocorrelation addresses this concern by filtering out the influence of the intermediate terms $\{x_{t-1}, x_{t-2}, \dots, x_{t-h+1}\}$. This is computed by solving the linear system

$$\mathbf{\Sigma} \boldsymbol{\alpha} = \boldsymbol{\gamma}, \quad (7.10)$$

where $(\mathbf{\Sigma})_{i,j} = \gamma_t(i-j)$ and $(\boldsymbol{\gamma})_i = \gamma_t(i)$ for $i, j = 1, 2, \dots, h$. The partial autocorrelation with lag h is $\alpha_t(h) = (\boldsymbol{\alpha})_h$. Reasonable guesses of p and q for ARMA can be made from examining plots of the autocorrelation and partial autocorrelation functions. If the autocorrelation plot slowly decays to zero and the partial autocorrelation plot abruptly decays to zero after a lag of h , then the model is likely $ARMA(h, 0)$, or equivalently $AR(h)$. Alternatively, if the partial autocorrelation plot slowly decays to zero and the autocorrelation plot abruptly decays to zero after a lag of h , then the model is likely $ARMA(0, h)$ or $MA(h)$. If both values slowly decay to zero, then the model is likely $ARMA(p, q)$ where the orders are taken to be a lag after which the plots have sufficiently decayed. Selecting the appropriate value of p and q can be difficult and take some trial and error.

Once the order of the ARMA model has been determined, the parameters ω_i and θ_i must be fit. This is accomplished using any preferred numerical optimization method to solve for the maximum likelihood estimate of these parameters. Once ω_i and θ_i have been computed, the model is examined for errors and overfitting. If necessary, the process is repeated with a new model selection.

7.2.2.4 Autoregressive Moving Integrated Average (ARIMA) Models

The success and popularity of the ARMA model have led to multiple variations and extensions of the method. The autoregressive integrated moving average (ARIMA) model was introduced to address the stationarity assumption on the time series data. ARIMA has been used recently for predicting the EV charging demand for stochastic power systems operation [34]. It incorporates differencing of the time series data to attempt to remove any nonstationary behavior. The number of differencing steps d is treated as another modeling parameter so that the model is written *ARIMA*(p, d, q).

Notice that the discussion of the Box–Jenkins method discussed in the previous section appears to assume that both the autocorrelation and the partial autocorrelation functions will eventually decay to zero (whether slowly or rapidly). If this is not the case, then differencing may be applied to the data to remove the nonstationarity. Differencing is a common approach to producing a stationary time series. One differencing iteration produces a new time series with

$$y_t = x_{t+1} - x_t. \quad (7.11)$$

The Box–Jenkins method determines d by differencing on the time series until the autocorrelation and partial autocorrelation plots decay appropriately.

7.2.3 Machine Learning Forecasting Approaches

Supervised machine learning seeks to construct a predictive model $f_{\Theta}(\mathbf{x})$, based on a given training set of data $\{\mathbf{x}_i, y_i\}_{i=1}^N$, where \mathbf{x}_i and y_i represent the feature vector and the target value [35]. For time series forecasting, the feature vectors are typically constructed by a moving window over the given data, $\mathbf{x}_i = [x_i, x_{i+1}, x_{i+2}, \dots, x_{i+n}]^T$, and the target value is the first datapoint after this window, $y_i = x_{i+n+1}$. The subscript Θ in the model denotes the collection of parameters that are tuned to best fit the data. Machine learning methods fit the model parameters from the data through iterative updates to reduce some loss function, such as the squared-error loss $L = \sum_{i=1}^N (y_i - f_{\Theta}(\mathbf{x}_i))^2$ or the absolute-error loss $L = \sum_{i=1}^N |y_i - f_{\Theta}(\mathbf{x}_i)|$.

This section introduces two popular machine learning methods for time series forecasting: the support vector regression (SVR) and the Gaussian process regression (GPR). There are many more approaches that may be appropriate depending on the specific problem at hand, such as k-nearest neighbor regression or regression trees. A more comprehensive overview of these methods can be found in [20, 36, 37]. Deep learning (or neural network) methods tend to fall within the realm of machine learning as well. However, their discussion is reserved for Sect. 7.3 so that they may be explored more in depth.

7.2.3.1 Support Vector Regression

Support vector regression (SVR) is a form of the popular machine learning approach known as support vector machine (SVM) [38]. The linear SVR attempts to fit the model

$$y = \boldsymbol{\theta}^\top \mathbf{x} + \theta_0 \quad (7.12)$$

to the data while minimizing $\|\boldsymbol{\theta}\|$. A linear model may be insufficient to describe the complex relationships underlying real-world datasets. Nonlinear or kernel SVR reformulates the model as

$$y = \sum_{i=1}^N \theta_i k(\mathbf{x}_i, \mathbf{x}) + \theta_0, \quad (7.13)$$

where $k(\cdot, \cdot)$ is a kernel function such as the radial basis function, or squared-exponential kernel $k(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma(\mathbf{x}_i - \mathbf{x}_j)^2)$, where γ is a hyperparameter that can be tuned using a grid search with cross-validation. The use of the kernel function implicitly defines a nonlinear mapping of the feature vector to some higher-dimensional space where a linear model is applied. This nonlinear mapping provides greater flexibility than simply applying the linear model directly to the features as in Eq. (7.12). Such a mapping is guaranteed to exist, provided that the kernel satisfies the so-called Mercer condition [39].

As mentioned previously, training any machine learning model requires the formulation of some loss function that informs the optimal set of model parameters. For SVR, it is common to use the ϵ -insensitive loss function. This loss ignores any points within $\pm\epsilon$ of the model prediction and is equal to the absolute error in the model for datapoints outside this range. Using the ϵ -insensitive loss, the SVR learning problem can be stated as

$$\min_{\Theta} \|\Theta\| + c \sum_{i=1}^N \xi_i \quad \text{subject to} \quad \left| y_i - \sum_{i=1}^N \theta_i k(\mathbf{x}_i, \mathbf{x}) + \theta_0 \right| \leq \epsilon + \xi_i \quad (7.14)$$

where ξ_i are slack variables that penalize deviation outside the ϵ -insensitive region of the loss function.

SVR is a popular machine learning approach to load forecasting [40–42]. SVRs have also been combined with other approaches to enhance performance on the short-term load forecasting problem. For example, an SVR can be combined with a locally weighted regression method that more heavily favors nearby points when making predictions [43]. Another approach combines SVRs with the empirical mode decomposition that separates out the high- and low-frequency components of a time series [44]. Both of these hybrid approaches were found to outperform the classical SVR method.

7.2.3.2 Gaussian Process Regression

Gaussian process regression (GPR) approaches time series forecasting from a Bayesian perspective by assuming that the underlying model for the data is drawn from prior distribution of functions [45]. For GPR, this prior is assumed to be a mixture of multivariate Gaussian random variables, or a Gaussian process,

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')), \quad (7.15)$$

where $m(\mathbf{x})$ and $k(\mathbf{x}, \mathbf{x}')$ are the mean and variance function, respectively. Often, the problem is formulated with mean zero and the kernel function equal to the squared-exponential from SVR.

Conditioning on the given dataset generates the posterior distribution of f , which is also a Gaussian process with mean and variance

$$\begin{aligned} \mathbb{E} \left[f(\mathbf{x}) \mid \{\mathbf{x}_i, y_i\}_{i=1}^N \right] &= \mathbf{k}(\mathbf{x})^\top \mathbf{K}^{-1} \mathbf{y}, \\ \mathbb{V} \text{ar} \left[f(\mathbf{x}) \mid \{\mathbf{x}_i, y_i\}_{i=1}^N \right] &= k(\mathbf{x}, \mathbf{x}) - \mathbf{k}(\mathbf{x})^\top \mathbf{K}^{-1} \mathbf{k}(\mathbf{x}), \end{aligned} \quad (7.16)$$

where $(\mathbf{y})_i = y_i$, $(\mathbf{K})_{i,j} = k(x_i, x_j)$ and $(\mathbf{k}(\mathbf{x}))_i = k(\mathbf{x}_i, \mathbf{x})$ for any \mathbf{x} . Using the decaying exponential kernel, it can be shown that the model interpolates the data without any variance. By assuming the data are corrupted by Gaussian noise with variance σ^2 , the posterior distribution then has mean and variance

$$\begin{aligned} \mathbb{E} \left[f(\mathbf{x}) \mid \{\mathbf{x}_i, y_i\}_{i=1}^N \right] &= \mathbf{k}(\mathbf{x})^\top \left(\mathbf{K} + \sigma^2 \mathbf{I} \right)^{-1} \mathbf{y}, \\ \mathbb{V} \text{ar} \left[f(\mathbf{x}) \mid \{\mathbf{x}_i, y_i\}_{i=1}^N \right] &= k(\mathbf{x}, \mathbf{x}) - \mathbf{k}(\mathbf{x})^\top \left(\mathbf{K} + \sigma^2 \mathbf{I} \right)^{-1} \mathbf{k}(\mathbf{x}). \end{aligned} \quad (7.17)$$

Gaussian processes have been applied to the load forecasting problem with promising success [46]. Additionally, GPR has been used for renewable energy forecasting relating to solar radiation [47] and wind power [48]. One study used GPRs with time-based composite covariance to handle seasonality in solar radiation data [49].

7.2.4 Shortcomings of Classical Approaches

Statistical and machine learning approaches to time series forecasting are powerful tools for understanding and modeling power systems forecasts. These methods have been performing reasonably well for short- and medium-term forecasting with traditionally acceptable level of accuracies. However, these methods can require significant data preprocessing that is not explored deeply here. For example, most of the statistical approaches assume stationary time series data with variable-independence and normality assumptions. Extensions to these methods that effectively deal with nonstationary data require manual tuning of various meta-parameters that essentially transform the data to be stationary.

Additionally, with the increasing dynamism in the future power systems, there is a need to obtain forecasts with higher accuracy than what is being achieved with traditional statistical and machine learning methods. The operations of power systems are getting more dynamic in nature with bidirectional flow of power through distributed energy resources, prosumer participation with demand-response, and other smart grid technologies. Increasing renewable energy penetration and decreasing synchronous generation resources are reducing the overall inertia of the grid [50]. This requires a finer temporal resolution of the forecasts in order to maintain reliable real-time operations of the grid. Furthermore, price forecasting for electricity markets can benefit greatly from a small percentage gain in the prediction accuracy, and better renewable energy forecasts are required by ISOs to lower the amount of the costly operational reserves [51].

The next section examines the history and current state-of-the-art in deep learning methods for power systems forecasting. With their ability to represent complex nonlinear behaviors in nonstationary, high-frequency, and high-dimensional time series data, these methods have been shown to be more robust to some of the abovementioned pitfalls of traditional approaches, but at the expense of some new hurdles.

7.3 Forecasting in Power Systems Using Deep Learning

7.3.1 Deep Learning

Artificial neural networks (ANN) are universal function approximators [52]; that is, it is possible to represent complex nonlinear behavior in a high-dimensional,

high-frequency, and nonstationary time series using ANNs. A deep neural network is an ANN with multiple hidden layers and nodes cascaded between input and output layers. Deep neural networks are sophisticated neural networks that have been successfully applied to analyze data in many disciplines in the past several years such as computer vision, image recognition, automatic speech recognition, bioinformatics, finance, and nature language processing [53].

In general, supervised machine learning algorithms are particularly task specific. However, deep learning networks are capable of learning intricate structures in large datasets, allowing them to generalize better to scenarios not present in the training data. Because of their capability to learn nonlinear relationships between input features, these networks can identify and ignore features that do not impact the target variable by minimizing the appropriate weights. Consequently, deep learning algorithms typically do not require the type of extensive data preprocessing and feature engineering that is required of other traditional machine learning methods. Additionally, deep learning algorithms are also capable of managing high-dimensional datasets better than traditional machine learning algorithms.

Recurrent neural networks (RNN), long short-term memory networks (LSTM), convolutional neural networks (CNN), autoencoders, restricted Boltzmann machines, deep belief networks, and deep Boltzmann machines are all common types of deep learning algorithms. The following sections introduce deep learning algorithms that are often applied to power systems forecasting problems and describe their mathematic framework briefly.

7.3.1.1 Recurrent Neural Network

Unlike traditional feedforward neural networks in which information flows from each layer to the next, RNNs allow the output from a layer to flow back into itself. This allows RNNs to process sequential data without assuming the independence among the time series samples or the datapoints [54]. Feedforward networks lose any knowledge of the system state after processing each time series sample, thereby failing to account for the relationship between exogenous variables along the temporal dimension. The recurrent edges in an RNN introduce temporal coupling into the model. The internal memory, formed by the feedback connections of the neurons in the hidden-layer nodes, updates the states of each neuron in the network with the previous input. The addition of this temporal coupling, which unfolds over time, allows RNNs to learn and exhibit complex system dynamics, making them efficient at time series forecasting problems.

The input to an RNN is a sequence of real-valued datapoints $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \dots\}$, where \mathbf{x}_t represents the value of time series variables timestep t . Given a finite input subsequence of length n , the target output for the RNN is the next value \mathbf{y}_{n+1} . Note that the target \mathbf{y} values may contain the same variables as the input \mathbf{x} values but at future timesteps, or they may be different if the input includes exogenous variables. The network output (i.e., the predictions from the network) is denoted

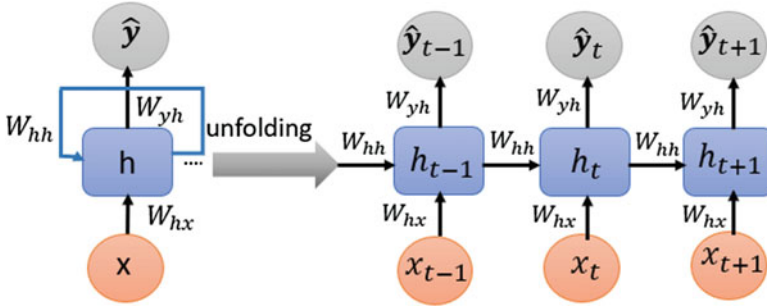


Fig. 7.3 Unfolding of an RNN over the temporal dimension

by \hat{y}_t . Figure 7.3 shows how the network unfolds the data along the temporal dimension. Mathematically, this unfolding is written as

$$\mathbf{h}_t = f_h (\mathbf{W}_{hx}\mathbf{x}_t + \mathbf{W}_{hh}\mathbf{h}_{t-1} + \mathbf{b}_h),$$

$$\hat{y}_{t+1} = f_o (\mathbf{W}_{yh}\mathbf{h}_t + \mathbf{b}_y), \quad (7.18)$$

where the current input datapoint \mathbf{x}_t is fed into the network along with the output of the hidden layer from the previous timestep \mathbf{h}_{t-1} , and the output from the hidden layer is used to generate the prediction \hat{y}_{t+1} . The remaining terms in Eq. (7.18) include the activation functions f_o and f_h , the weight matrices \mathbf{W}_{hx} , \mathbf{W}_{hh} , and \mathbf{W}_{yh} , and the biases for each layer \mathbf{b}_h and \mathbf{b}_y .

7.3.1.2 Long Short-Term Memory Network

In theory, RNNs should be capable of handling long-term temporal relationships because of their ability to retain information from previous timesteps. In practice, vanishing gradients make it difficult for them to learn long-term dependencies. Long short-term memory networks (LSTM) are a variation on the traditional RNN that are more effective at learning long-term trends in data, making them efficient at time series forecasting problems.

The key difference between RNNs and LSTMs is that the latter replaces hidden nodes with a more complex memory cell that handles the recurrent transfer of information (see Fig. 7.4). Four layers of neural connections, which exchange information in a particular special way, form the foundation of these memory cells. LSTMs are capable of learning long-term dependencies because the memory cells retain the existing information and append the unit with the new information; in RNNs, the content of the hidden node is replaced with the new value calculated from the current input.

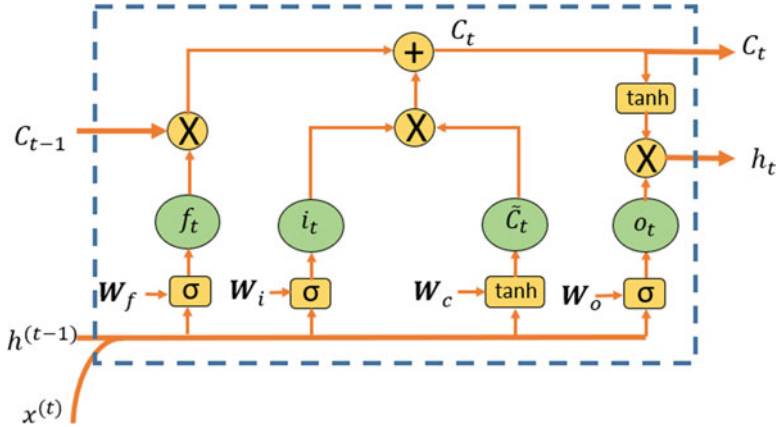


Fig. 7.4 LSTM memory cell diagram (from [55])

The mathematical formulation governing the flow of information in a LSTM cell is

$$\mathbf{f}_t = \sigma \left(\mathbf{W}_f \left[\mathbf{h}_{t-1}^\top \quad \mathbf{x}_t^\top \right]^\top + \mathbf{b}_f \right),$$

$$\mathbf{i}_t = \sigma \left(\mathbf{W}_i \left[\mathbf{h}_{t-1}^\top \quad \mathbf{x}_t^\top \right]^\top + \mathbf{b}_i \right),$$

$$\tilde{\mathbf{C}}_t = \tanh \left(\mathbf{W}_c \left[\mathbf{h}_{t-1}^\top \quad \mathbf{x}_t^\top \right]^\top + \mathbf{b}_c \right),$$

$$\mathbf{C}_t = \mathbf{f}_t \circ \mathbf{C}_{t-1} + \mathbf{i}_t \circ \tilde{\mathbf{C}}_t,$$

$$\mathbf{o}_t = \sigma \left(\mathbf{W}_o \left[\mathbf{h}_{t-1}^\top \quad \mathbf{x}_t^\top \right]^\top + \mathbf{b}_o \right),$$

$$\mathbf{h}_t = \mathbf{o}_t \circ \tanh(\mathbf{C}_t). \quad (7.19)$$

Note that the nodes in the cell operate on the concatenated vector $\left[\mathbf{h}_{t-1}^\top \quad \mathbf{x}_t^\top \right]^\top$ where \mathbf{x}_t is the current input vector and \mathbf{h}_{t-1} denotes the output from the cell at the previous timestep. The value \mathbf{C}_t is the current state of the cell and is defined by a combination of the information from the forget gate \mathbf{f}_t and the input gate \mathbf{i}_t (where \circ denotes the element-wise Hadamard product). The output gate \mathbf{o}_t is acted on by the cell state to produce the output of the cell \mathbf{h}_t . The various

\mathbf{W} 's and \mathbf{b} 's represent the weights and biases in the cell, while σ and \tanh are the sigmoid and hyperbolic tangent activation functions, respectively.

Once the model is chosen, there are two main iterative phases in the learning algorithm: (1) forward propagation and (2) weight update. For an RNN or LSTM, the architecture first unfolds the time series input along the temporal dimension, making the network similar to a traditional feedforward neural network. In the forward propagation phase, the input vector propagates through the hidden layers (using randomly initialized values for the weight matrices and biases) to compute the output vector. The mismatch between the interim prediction output and the actual target is calculated as a loss function (e.g., the mean squared-error loss). The weights are updated using gradient descent with the gradients with respect to the loss function computed using the backpropagation through time algorithm.

7.3.1.3 Other Relevant Models

WaveNet deep learning models were recently introduced that apply deep learning techniques from audio signal processing and computer vision models to time series (sequential) data [56]. Convolutional neural networks (CNN) are a type of deep feedforward ANN that have been used to analyze visual imagery on a large scale. A deep convolutional WaveNet architecture, which is variation of CNN, has been successfully used for conditional time series forecasting [57].

7.3.2 Deep Learning Applications

Deep learning has been applied to a variety of power systems prediction problems recently, including solar forecasting, building load forecasting, system load forecasting, wind forecasting, and electricity price forecasting. RNN and LSTMs are the most popular architectures published in the literature for power systems forecasting problems. The following sections discuss the recent literature of power systems predictive analytics using deep learning.

7.3.2.1 Load Forecasting

Load forecasting may be done at either a systems level or building level and for different time horizons. Deep neural networks have been used for building energy load forecasting using an LSTM and an LSTM-based sequence to sequence modeling approach [58]. Short-term residential load forecasting is done using an LSTM in [59]. Shi et al. [60] propose a pooling-based RNN architecture, which outperforms traditional RNNs, along with other traditional machine learning algorithms in residential load forecasting. Another variation of RNN, called the

gated recurrent unit network, is used in [61] for daily peak load forecasting. CNNs with k-means clustering have also been applied for short-term load forecasting for smart grids [62].

7.3.2.2 Generation Forecasting

In the field of renewable energy, deep learning has been applied for wind and solar forecasting problems. A short-term wind forecasting problem is addressed using RNNs with a so-called infinite feature selection method in [63] and using CNNs in [64]. A hybrid deep learning approach is proposed for day-ahead wind power forecasting in [65]. Wind and solar irradiance forecasting are done using CNNs with input data obtained from numerical weather prediction in [66].

Solar forecasting methodologies vary widely based on the type of inputs being used for the process. For example, a standard time series forecasting problem may only make use of previous solar irradiance measurement (endogenous variables). Alternatively, one might use ground-based meteorological parameters (exogenous variables) or sky imagery/video for predicting solar irradiance. Siddiqui et al. propose a deep learning-based approach for solar irradiance forecasting using sky videos [67]. LSTMs are used for solar power forecasting by Gensler et al. in [68] and RNNs are used in [69] for solar irradiance forecasting. Section 7.4 in this chapter examines a case study in multi-time-horizon solar irradiance forecasting using RNNs and LSTMs [55, 70].

7.3.2.3 Electricity Price Forecasting and Electric Vehicle Charging

Electricity price forecasting in competitive energy markets is a challenging prediction problem because of the rare characteristics of electricity. Electricity cannot be treated like other commodities because trading requires a balance between supply and demand at every point in time. The failure to maintain this balance results in blackouts and brownouts that are hugely detrimental to the society as a whole. The research around deep learning-based approaches to electricity price forecasting is growing. There have been a few articles exploring this topic [71–75]. Deep learning has also been applied for demand-side management for smart charging of EVs [76].

7.3.3 Deep Learning Strengths and Shortcomings

Deep learning has shown promising results in the field of predictive analysis, because of its ability to model complex, nonlinear relationships between various exogenous input variables and the associated output. It is capable of uncovering trends in the historical dataset, providing highly accurate forecasts. For power systems forecasting problems, deep learning algorithms are increasingly outpacing

traditional approaches for nowcasting and short-term forecasting. In order to produce such accurate predictions for these time horizons, deep learning algorithms require a relatively significant amount of training data. The next two sections summarize the strengths and weaknesses of deep learning approaches in the context of power systems forecasting problems.

7.3.3.1 Strengths

Deep learning algorithms with recurrent connections (e.g., RNN and LSTM) are capable of capturing short- and long-term trends in time series data. When trained using exogenous variables, these algorithms are effective at finding and modeling the complex temporal relationships between various input variables. Deep learning also has the rather unique capability of performing in situ feature engineering; that is, extensive manual feature engineering is not required for deep learning algorithms like traditional machine learning algorithms. The data availability in power systems has exploded in recent years, creating a natural environment for the emergence of deep learning algorithms. For this reason, it is reasonable to assume that deep learning has yet to reach its full potential in revolutionizing power systems predictive analytics field.

7.3.3.2 Shortcomings

Deep learning models have traditionally been difficult to train because of their expensive computational costs. This limitation has been overcome in recent years with technical advances in GPUs, network architectures, and development of performance optimization techniques. While ANNs act as universal function approximators, they are also often a black-box approach to modeling. They lack interpretability and are prone to overfitting because of the high capacity to learn (especially deep neural networks). Also, deep learning algorithms require significant amounts of data for training. For cases where data are limited, deep learning algorithms may not be the optimal method to use. Lastly, for long-term forecasting horizons (5+ years), statistical methods still provide reasonably good predictions, given the limited data availability.

7.4 Case Study: Multi-Timescale Solar Irradiance Forecasting Using Deep Learning

This section reviews an example of using deep learning for real-time forecasting of solar irradiance [55, 70], where a unified architecture is proposed for predicting multi-time-horizon solar irradiance. This work uses both RNNs and LSTMs to make



Fig. 7.5 The seven SURFRAD research stations distributed across the continental USA

predictions of global horizontal irradiance (GHI), also referred to as the total solar irradiance. Recall from Sect. 7.3 that these deep learning architectures use data from previous timesteps to inform the current one. This allows the models to learn the underlying dynamics of system in order to enhance their predictive capabilities.

7.4.1 Data

The data for this study come from the seven Surface Radiation Budget Network (SURFRAD) measurement stations, scattered across the continental USA (see Fig. 7.5) that measure various meteorological parameters, including solar radiation. The distribution of these research stations across various climate zones demonstrates the robustness of the constructed networks in predicting GHI.¹ Minute-by-minute meteorological data for 2009–2011 from this database is used in this study. The data are averaged over each hour to obtain mean hourly GHI values for forecasting.

¹Because the supervised training approach relies on the data from the given geographic location to learn the relationships between the various meteorological inputs and the GHI, it is difficult to transport/reuse this model for a climatically different geographical location. Thus, the model will need to be trained with a location-specific dataset for using it in various geographical locations. Therefore, the algorithm itself is robust for various locations, but the model needs to be retrained for different locations.

Data from 2010 and 2011 at each location are used for training while corresponding data from 2009 are used for measuring performance.

7.4.1.1 Global Horizontal Irradiance

Global horizontal irradiance (GHI) refers to the total solar power per unit area that is incident on some surface (e.g., a photovoltaic solar panel) and is typically measured in W/m^2 . This value has two main components: (1) direct normal irradiance (DNI) and (2) diffuse horizontal irradiance (DHI). The GHI at a particular time t can be expressed as

$$GHI_t = DNI_t \times \cos(\theta) + DHI_t, \quad (7.20)$$

where θ denotes the solar zenith angle, which is the angle between the zenith (overhead) and the sun. This value is important to understanding the availability of solar energy on the grid.

The constructed networks (discussed in Sect. 7.4.2) directly predict a value known as the clear-sky index K_t . This value is a ratio of the true GHI to the expected GHI in a cloud-free scenario,

$$K_t = GHI_t / GHI_t^{\text{clear}}. \quad (7.21)$$

The clear-sky index is a dimensionless value that describes the total solar irradiance relative to a theoretical upper limit, which occurs in cloud-free situations. This acts as a type of normalization for the model that can increase robustness to location or seasonality. The clear-sky GHI (GHI_t^{clear}) in Eq. (7.21) is calculated using the Bird clear-sky model [77] based on latitude, longitude, elevation, and atmospheric parameters, such as column water vapor, ozone optical thickness, and aerosol optical depth. Based on this calculation and the predicted clear-sky index from the deep learning model, one can easily obtain the predicted GHI.

7.4.1.2 Exogenous Input Variables

The input to the deep learning model is a vector of 20 exogenous variables for each timestep:

- downwelling global solar (W/m^2),
- upwelling global solar (W/m^2),
- direct-normal solar (W/m^2),
- downwelling diffuse solar (W/m^2),
- downwelling thermal infrared (W/m^2),
- downwelling infrared case temperature (K),
- downwelling infrared dome temperature (K),

- upwelling thermal infrared (W/m^2),
- upwelling infrared case temperature (K),
- upwelling infrared dome temperature (K),
- global UVB (mW/m^2),
- photosynthetically active radiation (W/m^2),
- net solar (W/m^2),
- net infrared (W/m^2),
- net radiation (W/m^2),
- 10-mean air temperature (C),
- relative humidity (%),
- wind speed (m/s),
- wind direction ($^\circ$),
- station pressure (mb).

Not all of the variables listed here are necessarily important to the solar irradiance forecast, but they have been used in this case study. As an extension to this work, further experiments can be conducted to understand the relevance of the individual input variables and accordingly reduce the dimensionality of the dataset.

7.4.1.3 Data Preprocessing and Postprocessing

The algorithmic approach in this case study begins by preprocessing the data. This includes removing extreme outliers (values which are ± 4 standard deviation away from the mean) as well as nighttime values, filling in missing data with the mean value of surrounding points, and normalizing the input data vectors. The clear-sky GHI is computed using the Bird model (see Sect. 7.4.1.1) and used to transform target GHI values to the clear-sky index K_t . Postprocessing the data includes recovering the predicted GHI from the clear-sky index and computing the performance of the network using the mean squared error.

7.4.2 Model Architecture and Training

This case study examines two scenarios: (1) a fixed-time horizon that is similar to other statistical and machine learning forecasting approaches (such as those discussed in Sects. 7.2.2 and 7.2.3) and (2) a multi-time-horizon that is better suited for the flexibility of a deep learning model. In the fixed-time case, separate models are trained for each desired time horizon (1, 2, 3, and 4 h) while a single model is used to predict all of the time horizons in the multi-time case. In both scenarios, separate models are trained for each of the seven SURFRAD locations.

For the fixed-time-horizon problem, this study only considers traditional RNN models and compares the performance of this deep learning method to standard machine learning approaches. The network is constructed using rectified linear units (ReLU) activation functions for all hidden layers and a linear activation on the

output layer. The output is scalar-valued because the goal is to predict GHI for a single time horizon.

The multi-time-horizon networks predict GHI at 1-, 2-, 3-, and 4-h time horizons simultaneously, producing a four-dimensional output vector. This work also proposes an extension to the unified architecture for predicting multi-time-scale solar irradiance, which covers 5-min, 15-min, and other such intrahour time horizons. This work compares LSTMs and RNNs; however, no comparison is made to traditional machine learning methods because these approaches are unable to perform multi-time-horizon predictions. The RNNs have similar activation architectures to the fixed-time-horizon case. The LSTM networks use sigmoid and hyperbolic tangent activations within the memory cells.

For training, the deep learning models have access to the target GHI values so that the mean squared loss can be computed. The models are trained using stochastic gradient descent where the gradients with respect to this loss are computed using backpropagation through time. The training minibatch sizes are $n = 100$, and the networks are trained for 1000 epochs.

7.4.3 Results

7.4.3.1 Single Time Horizon Model

Table 7.2 shows the comparison between the RNN performance and the performance of other machine learning forecasting approaches. The values in the ML column are those presented in [78] where the authors perform the same fixed-time-horizon study using several traditional machine learning algorithms (SVRs, random forests, and gradient boosting) as well as a traditional feedforward neural network. The listed performance is the optimal performance across all testing algorithms for each horizon/location combination. In each case, the RNN approach significantly outperforms the others.

7.4.3.2 Multi-Time-Horizon Model

Table 7.3 contains the results of the RNN/LSTM comparison study for the multi-time-horizon problem.² Recall that for this study, a single RNN or LSTM network is trained for each location that forecasts GHI out to all four time horizons. Neither network architecture outperformed the other across all seven locations. However, within each location there is a significant increase in error from the 3-h to the 4-h

²Note that Table 7.2 compares the GHI W/m² values between RNN and traditional ML approaches while Table 7.3 is comparing the performance of RNN and LSTM algorithms based on clear-sky index (which is a dimensionless parameter).

Table 7.2 Root mean squared error of predicted GHI (W/m^2) for fixed-time-horizon RNNs and machine learning (ML) approaches

	Bondville, IL		Boulder, CO		Desert Rock, NV		Fort Peck, MT		Goodwin Creek, MS		Penn State, PA		Sioux Falls, SD	
	RNN	ML	RNN	ML	RNN	ML	RNN	ML	RNN	ML	RNN	ML	RNN	ML
1-h	16.8	62	17	74	41.7	52	21.2	56	24.8	71	8.64	67	27.2	52
2-h	20.73	98	20.7	108	57.23	72	29.7	81	25.2	103	10.5	97	32.1	81
3-h	18.78	116	21.2	123	60.54	83	25.5	94	26.9	125	11.8	114	30.6	96
4-h	17.98	121	22.9	125	49.71	82	29.4	93	22	120	10.7	117	35.3	103
Mean	18.57	99.25	20.45	107.5	52.29	72.25	26.45	81	24.73	104.8	10.41	98.75	31.3	83

Table 7.3 Root mean squared error of the predicted clear-sky index for multi-time-horizon RNN and LSTM models

	Bondville, IL		Boulder, CO		Desert Rock, NV		Fort Peck, MT		Goodwin Creek, MS		Penn State, PA		Sioux Falls, SD	
	RNN	LSTM	RNN	LSTM	RNN	LSTM	RNN	LSTM	RNN	LSTM	RNN	LSTM	RNN	LSTM
1-h	0.71	0.65	0.59	0.58	0.47	0.46	0.71	0.72	0.65	0.67	0.68	0.69	0.69	0.71
2-h	1.12	1.18	1.06	1.06	0.98	0.96	1.35	1.34	1.17	1.20	1.13	1.11	1.25	1.23
3-h	4.52	4.30	4.01	4.45	4.53	4.71	6.31	6.18	4.59	4.58	3.16	3.12	5.06	4.99
4-h	52.13	43.65	49.45	58.88	153.27	126.87	69.99	77.99	58.98	54.48	28.40	24.70	73.41	77.16
Mean	14.62	12.45	13.78	16.24	39.81	33.25	19.59	21.56	16.35	15.23	8.34	7.40	20.10	21.02

forecasting horizon. This could indicate that autocorrelation in the time series data decays after 3 h.

7.5 Summary and Future Work

Given the rapidly occurring technological changes in power systems and their forthcoming transformation into the smart grid, which operates as a part of a complex amalgamation of interdependent transportation, communication, and IoT networks, there is an urgent need for developing and deploying better algorithms for forecasting the various power systems quantities. Moreover, operational uncertainties continue to increase with the burgeoning share of utility-scale as well as DER-scale renewable energy generation on the grid, calling for better short-term forecasts. These play a significant role in the optimization of the operational efficiency of power systems, both economically and in terms of reliability.

Forecasting accuracies for electricity price prediction play a major role in maintaining the economic viability of energy producers' businesses in the market. Resource and load forecasting also have a major role to play in large-scale deployment of microgrids because these quantities are the main inputs to the optimization algorithms aimed at operating the microgrids intelligently (i.e., maximizing economic benefit while maintaining the reliability of the local supply).

Deep learning algorithms (e.g., RNNs and LSTMs) have been applied to power systems forecasting problems with promising results in the recent literature. They also offer the potential to continue improving as they are trained further on the continuous stream of newly generated data. As with any research area, the goal is to ultimately move these algorithms to the industry deployment phase. Because of their low forward inference time (on the order of milliseconds), these algorithms and architectures can provide forecasts in the near real-time horizon. The performance of the deployed systems can be further improved by implementing sophisticated hyperparameter tuning mechanisms.

The following two sections briefly note some areas where there is plenty of scope as well as a need for further development of deep learning applications for power systems.

7.5.1 *Deterministic Versus Probabilistic Forecasting*

The forecasted values from the deep learning models discussed in Sect. 7.3 are deterministic in nature. That is, given the same sequence of inputs, the networks will always produce the same output. Furthermore, there is no measure of confidence related to the predictions. Recall the Gaussian process regression (GPR) from Sect. 7.2.3.2. This probabilistic forecasting approach naturally produces a measure of confidence based on the variance in the GP. As the predicted values get further from any given data, the variance grows and the confidence in the predicted value

decreases. Such understanding of uncertainties in power systems forecasting is critical because of the highly variable nature of the data and the large cost of grid blackouts and brownouts. Some work has considered how deep learning can be recast as a probabilistic model [79, 80], but continued research into the topic is critical.

7.5.2 *Other Potential Applications*

Anomaly detection in smart grids is a timely and relevant topic as the distribution grid infrastructure in industrialized countries like the USA has aged and needs refurbishment and replacement to maintain the reliability of the supply. There has been relatively less progress in applying deep learning algorithms for anomaly detection in power systems prognostics and fault prediction problems [81–83]. The application areas include anomaly detection for predicting the remaining useful life of the components of power systems, predicting impending fault on power systems, and predicting building level faults based on the data from building sensors.

Acknowledgments The authors wish to thank Kate Anderson and Adam Warren (National Renewable Energy Laboratory) for the encouragement to pursue this research work. The authors would also like to thank Manajit Sengupta and Ryan King (National Renewable Energy Laboratory) for providing useful suggestions to refine the manuscript. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government.

References

1. M. Shahidpour, H. Yamin, Z. Li, *Market Operations in Electric Power Systems: Forecasting, Scheduling, and Risk Management* (Wiley-IEEE Press, Hoboken, 2002)
2. Federal Energy Regulatory Commission, Order no. 1000 - Transmission planning and cost allocation, 21 July 2011. (Online). <https://www.ferc.gov/whats-new/comm-meet/2011/072111/E-6.pdf>. Accessed 27 June 2019
3. I. Penn, P. Eavis, J. Glanz, How PG&E ignored fire risks in favor of profits, *New York Times*, 2019
4. H. Chen, Y. Zhang, H. Ngan, Power System Planning, in *Power System Optimization: Large-Scale Complex Systems Approaches* (Wiley, Singapore, 2016), pp. 76–130
5. A.S. Malik, C. Kuba, Power generation expansion planning including large scale wind integration: a case study of Oman. *J. Wind Energy* **2013**, 1 (2013)
6. M. Child, C. Kemfert, D. Bogdanov, C. Breyer, Flexible electricity generation, grid exchange and storage for the transition to a 100% renewable energy system in Europe. *Renew. Energy* **139**, 80–101 (2019)
7. I.F. Abdin, E. Zio, An integrated framework for operational flexibility assessment in multi-period power system planning with renewable energy production. *Appl. Energy* **222**, 898–914 (2018)
8. O.J. Guerra, D.A. Tejada, G.V. Reklaitis, An optimization framework for the integrated planning of generation and transmission expansion in interconnected power systems. *Appl. Energy* **170**, 1–21 (2016)

9. H. Kim, S. Lee, S. Han, W. Kim, K. Ok, S. Cho, *Integrated generation and transmission expansion planning using generalized Bender's decomposition method*, in IEEE International Conference on Computational Intelligence & Communication Technology, Ghaziabad, India, 2015
10. S. Hasanvand, M. Nayeripour, H. Fallahzadeh-Abarghouei, A new distribution power system planning approach for distributed generations with respect to reliability assessment. *J. Renewable Sustainable Energy* **8**, 045501 (2016)
11. B. Singh, C. Pal, V. Mukherjee, P. Tiwari, M.K. Yadav, Distributed generation planning from power system performances viewpoints: a taxonomical survey. *Renew. Sust. Energy. Rev.* **75**, 1472–1492 (2017)
12. M.R. AlRashidi, K.M. EL-Naggar, Long term electric load forecasting based on particle swarm optimization. *Appl. Energy* **87**(1), 320–326 (2010)
13. A. Poghosyan, D.V. Greetham, S. Haben, T. Lee, Long term individual load forecast under different electrical vehicles uptake scenarios. *Appl. Energy* **157**(1), 699–709 (2015)
14. M. Zhao, W. Liu, J. Su, L. Zhao, X. Dong, W. Liu, *Medium and long term load forecasting method for distribution network with high penetration DGs*, in 2014 China International Conference on Electricity Distribution (CICED), Shenzhen, China, 2014
15. E. Vaahedi, *Practical power system operation* (Wiley, Hoboken, 2014)
16. M. Matos, R. Bessa, A. Botterud, Z. Zhou, Forecasting and setting power system operating reserves, in *Renewable Energy Forecasting - From Models to Applications*, ed. by G. Kariniotakis (Woodhead Publishing Elsevier, Duxford, 2017), pp. 279–308
17. R. Weron, Electricity price forecasting: a review of the state-of-the-art with a look into the future. *Int. J. Forecast.* **30**, 1030–1081 (2014)
18. N. Mazzi, P. Pinson, Wind power in electricity markets and the value of forecasting, in *Renewable Energy Forecasting - From Models to Applications*, Woodhead Publishing Series in Energy (Woodhead Publishing Elsevier, Duxford, 2017), pp. 259–278
19. C. Frohlich, J. Lean, Total Solar Irradiance (TSI) composite database, National Oceanic and Atmospheric Administration (Online). <https://www.ngdc.noaa.gov/stp/solar/solarirrad.html>. Accessed July 2019
20. C. Voyant, G. Notton, S. Kalogirou, M.-L. Nivet, C. Paoli, F. Motte, A. Fouilloy, Machine learning methods for solar radiation forecasting: a review. *Renew. Energy* **105**, 569–582 (2017)
21. R.G. Brown, Exponential smoothing for predicting demand. *Oper. Res.* **5**(1), 145 (1957)
22. C.C. Holt, *Forecasting Seasonals and Trends by Exponentially Weighted Moving Averages* (Office of Naval Research, Arlington, 1957)
23. P.R. Winters, Forecasting sales by exponentially weighted moving averages. *Manag. Sci.* **6**(3), 324–342 (1960)
24. W.R. Christiaanse, Short-term load forecasting using general exponential smoothing. *IEEE Trans. Power Syst.* **PAS-90**(2), 900–911 (1971)
25. J.W. Taylor, L.M. de Menezes, P.E. McSharry, A comparison of univariate methods for forecasting electricity demand up to a day ahead. *Int. J. Forecast.* **22**, 1–16 (2006)
26. J.W. Taylor, P.E. McSharry, Short-term load forecasting methods: an evaluation based on European data. *IEEE Trans. Power Syst.* **22**(4), 2213–2219 (2007)
27. T.N. Goh, K.J. Tan, Stochastic modeling and forecasting of solar radiation data. *Sol. Energy* **19**, 755–757 (1977)
28. S. Vemuri, W.L. Huang, D.J. Nelson, On-line algorithms for forecasting hourly loads of an electric utility. *IEEE Trans. Power Syst.* **PAS-100**(9), 3775–3784 (1981)
29. G. Reikard, Predicting solar radiation at high resolutions: a comparison of time series forecasts. *Sol. Energy* **83**(3), 342–349 (2009)
30. A. Moreno-Munoz, J. J. G. de la Rosa, R. Posadillo, F. Bellido, *Very short term forecasting of solar radiation*, in 2008 33rd IEEE Photovoltaic Specialists Conference, 2008
31. J. Contreras, R. Espinola, F.J. Nogales, A.J. Conejo, ARIMA models to predict next-day electricity prices. *IEEE Trans. Power Syst.* **18**(3), 1014–1020 (2003)
32. N. Amjady, Short-term hourly load forecasting using time-series modeling with peak load estimation capability. *IEEE Trans. Power Syst.* **16**(3), 498–505 (2001)

33. G.E.P. Box, G.M. Jenkins, *Time Series Analysis: Forecasting and Control* (Prentice Hall, Englewood Cliffs, 1994)
34. H. Amini, A. Kargarian, O. Karabasoglu, ARIMA-based decoupled time series forecasting of electric vehicle charging demand for stochastic power system operation. *Electr. Power Syst. Res.* **140**, 379–390 (2016)
35. T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning* (Springer, New York, 2001)
36. N.K. Ahmed, A.F. Atiya, N.E. Gayar, H. El-Shishiny, An empirical comparison of machine learning models for time series forecasting. *Econ. Rev.* **29**(5–6), 594–621 (2010)
37. R.H. Inman, H.T.C. Pedro, C.F.M. Coimbra, Solar forecasting methods for renewable energy integration. *Prog. Energy Combust. Sci.* **39**(6), 535–576 (2013)
38. V. Vapnik, *The Nature of Statistical Learning Theory* (Springer, New York, 1995)
39. N. Cristianini, J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods* (Cambridge University Press, Cambridge, 2000)
40. E. Ceperic, V. Ceperic, A. Baric, A strategy for short-term load forecasting by support vector regression machines. *IEEE Trans. Power Syst.* **28**(4), 4356–4364 (2013)
41. J. Dhillon, S.A. Rahman, S.U. Ahmad, M.D.J. Hossain, *Peak electricity load forecasting using online support vector regression*, in IEEE Canadian Conference on Electrical and Computer Engineering, 2016
42. J. Jose, V. Margaret, K.U. Rao, *Impact of demand response contracts on short-term load forecasting in smart grid using SVR optimized by GA*, in Innovations in Power and Advanced Computing Technologies, 2017
43. E.E. Elattat, J. Goulermas, Q.H. Wu, Electric load forecasting based on locally weighted support vector regression. *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.* **40**(4), 438–447 (2010)
44. L. Ghelardoni, A. Ghio, D. Anguita, Energy load forecasting using empirical mode decomposition and support vector regression. *IEEE Trans. Smart Grid* **4**(1), 549–556 (2013)
45. C.E. Rasmussen, C.K.I. Williams, *Gaussian Processes for Machine Learning* (The MIT Press, Cambridge, 2006)
46. H. Mori, M. Ohmi, *Probabilistic short-term load forecasting with Gaussian processes*, in Proceedings of the 13th International Conference on Intelligent Systems Application to Power Systems, 2005
47. A. Rohani, M. Taki, M. Abdollahpour, A novel soft computing model (Gaussian process regression with K-fold cross validation) for daily and monthly solar radiation forecasting (part: I). *Renew. Energy* **115**, 411–422 (2018)
48. N. Chen, Z. Qian, I.T. Nabney, X. Meng, Wind power forecasts using Gaussian processes and numerical weather prediction. *IEEE Trans. Power Syst.* **29**(2), 656–665 (2014)
49. S. Salcedo-Sanz, C. Casanova-Mateo, J. Munoz-Mari, G. Camps-Valls, Prediction of daily global solar irradiation using temporal Gaussian processes. *IEEE Geosci. Remote Sens. Lett.* **11**(11), 1936–1940 (2014)
50. A. Ulbig, T. Borsche, G. Andersson, Impact of low rotational inertia on power system stability and operation. *IFAC Proc. Vol.* **47**(3), 7290–7297 (2014)
51. E. Bitar, P.P. Khargonekar, K. Poolla, Systems and control opportunities in the integration of renewable energy into the smart grid. *IFAC Proc. Vol.* **44**(1), 4927–4932 (2011)
52. K. Hornik, M. Stinchcombe, H. White, Multilayer feedforward networks are universal approximators. *Neural Netw.* **2**(5), 359–366 (1989)
53. Y. LeCun, Y. Bengio, G. Hilton, Deep learning. *Nature* **521**, 436–444 (2015)
54. A. Graves, *Supervised sequence labelling with recurrent neural networks* (Springer-Verlag, Berlin, 2012)
55. S. Mishra, P. Palanisamy, An integrated multi-time-scale modeling for solar irradiance forecasting using deep learning, arXiv:1905.02616, 2019
56. A.v.d. Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, K. Kavukcuoglu, *WaveNet: a generative model for raw audio*, in DeepMind, 2016

57. A. Borovykh, S. Bohte, C.W. Oosterlee, Conditional time series forecasting with convolutional neural networks, arXiv:1703.04691, 2018
58. D.L. Marino, K. Amarasinghe, M. Manic, *Building energy load forecasting using deep neural networks*, in IECON 2016 - 42nd Annual Conference of the IEEE Industrial Electronics Society, Florence, Italy, 2016
59. W. Kong, Z.Y. Dong, Y. Jia, D.J. Hill, Y. Xu, Y. Zhang, Short-term residential load forecasting based on LSTM recurrent neural network. *IEEE Trans. Smart Grid* **10**(1), 841–851 (2019)
60. H. Shi, M. Xu, R. Li, Deep learning for household load forecasting - a novel pooling deep RNN. *IEEE Trans. Smart Grid* **9**(5), 5271–5280 (2018)
61. Z. Yu, Z. Niu, W. Tang, Q. Wu, Deep learning for daily peak load forecasting—a novel gated recurrent neural network combining dynamic time warping. *IEEE Access* **7**, 17184–17194 (2019)
62. X. Dong, L. Qian, L. Huang, *Short-term load forecasting in smart grid: a combined CNN and K-means clustering approach*, in 2017 IEEE International Conference on Big Data and Smart Computing, Jeju, South Korea, 2017
63. H. Shao, X. Deng, Y. Jiang, A novel deep learning approach for short-term wind power forecasting based on infinite feature selection and recurrent neural network. *J. Renewable Sustainable Energy* **10**, 043303 (2018)
64. C.-J. Huang, P.-H. Kuo, A short-term wind speed forecasting model by using artificial neural networks with stochastic optimization for renewable energy systems. *Energies* **11**, 2777 (2018)
65. Y.-Y. Hong, C.L.P.P. Rioflorido, A hybrid deep learning-based neural network for 24-h ahead wind power forecasting. *Appl. Energy* **250**, 530–539 (2019)
66. D. Diaz-Vico, A. Torres-Barran, A. Omari, J.R. Dorronsoro, Deep neural networks for wind and solar energy prediction. *Neural Process. Lett.* **46**, 829 (2017)
67. T.A. Siddiqui, S. Bharadwaj, S. Kalyanaraman, *A deep learning approach to solar-irradiance forecasting in sky-videos*, in 2019 IEEE Winter Conference on Applications of Computer Visions (WACV), Waikoloa Village, HI, USA, 2019
68. A. Gensler, J. Henze, B. Sick, N. Raabe, *Deep learning for solar power forecasting - an approach using AutoEncoder and LSTM neural networks*, in IEEE International Conference on Systems, Man, and Cybernetics (SMC), Budapest, Hungary, 2016
69. A. Alzahrani, P. Shamsi, C. Dagli, M. Ferdowsi, Solar irradiance forecasting using deep neural networks. *Proc. Comput. Sci.* **114**, 304–313 (2017)
70. S. Mishra, P. Palanisamy, *Multi-time-horizon solar forecasting using recurrent neural network*, in 2018 IEEE Energy Conversion Congress and Exposition (ECCE), Portland, OR, 2018
71. W. Zhang, F. Cheema, D. Srinivasan, *Forecasting of electricity prices using deep learning networks*, in 2018 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC), Kota Kinabalu, Malaysia, 2018
72. R.A. Chinnathambi, S.J. Plathottam, T. Hossen, A.S. Nair, P. Ranganathan, *Deep neural networks (DNN) for day-ahead electricity price markets*, in 2018 IEEE Electrical Power and Energy Conference (EPEC), Toronto, ON, Canada, 2018
73. J. Lago, F.D. Ridder, B.D. Schutter, Forecasting spot electricity prices: deep learning approaches and empirical comparison of traditional algorithms. *Appl. Energy* **221**, 386–405 (2018)
74. A. Brusafferri, M. Matteucci, P. Portlani, A. Vitali, Bayesian deep learning based method for probabilistic forecast of day-ahead electricity prices. *Appl. Energy* **250**, 1158–1175 (2019)
75. S. Atef, A.B. Eltawil, *A comparative study using deep learning and support vector regression for electricity price forecasting in smart grids*, in 2019 IEEE 6th International Conference on Industrial Engineering and Applications (ICIEA), Tokyo, Japan, 2019
76. K.L. Lopez, C. Gagne, M.-A. Gardner, Demand-side management using deep learning for smart charging of electric vehicles. *IEEE Trans. Smart Grid* **10**(3), 2683–2691 (2019)
77. R.E. Bird, R.L. Hulstrom, *Simplified Clear Sky Model for Direct and Diffuse Insolation on Horizontal Surfaces* (Solar Energy Research Institute, Golden, CO, 1981)

78. A. Dobbs, T. Elgindy, B.-M. Hodge, A. Florita, J. Novacheck, *Short-Term Solar Forecasting Performance of Popular Machine Learning Algorithms* (National Renewable Energy Laboratory, Golden, CO, 2017)
79. G. Li, J. Shi, J. Zhou, Bayesian adaptive combination of short-term wind speed forecasts from neural network models. *Renew. Energy* **36**(1), 352–359 (2011)
80. N. Bassamzadeh, R. Ghanem, Multiscale stochastic prediction of electricity demand in smart grids using Bayesian networks. *Appl. Energy* **193**, 369–380 (2017)
81. Z. Fengming, L. Shufang, G. Zhimin, W. Bo, T. Shiming, P. Mingming, Anomaly detection in smart grid based on encoder-decoder framework with recurrent neural network. *J. China Univ. Posts Telecommun.* **24**(6), 67–73 (2017)
82. L. Banjanovic-Mehmedovic, A. Hajdarevic, M. Kantardzic, F. Mehmedovic, I. Džananovic, Neural network-based data-driven modelling of anomaly detection in thermal power plant. *J. Control Meas. Electron. Comput. Commun.* **58**(1), 69 (2017)
83. C. Zhang, D. Song, Y. Chen, X. Feng, C. Lumezanu, W. Cheng, J. Ni, B. Zong, H. Chen, N.V. Chawla, *A deep neural network for unsupervised anomaly detection and diagnosis in multivariate time series data*, in Proceedings of the AAAI Conference on Artificial Intelligences, 2019