

Chapter 10

How Robust is a Robust Policy? Comparing Alternative Robustness Metrics for Robust Decision-Making

Jan H. Kwakkel, Sibel Eker, and Erik Pruyt

Abstract Nowadays, decision-makers face deep uncertainties from a myriad of external factors such as climate change, population growth, new technologies, and economic developments. The challenge is to develop robust policies, which perform well across all possible resolutions of the uncertainties. One approach for achieving this is to design a policy to be adapted over time in response to how the future actually unfolds. A key determinant for the efficacy of such an adaptive policy is the specification of when and how to adapt it. This specification depends on how robustness is being operationalized. To date, there is little guidance for selecting an appropriate robustness metric. In this chapter we address this problem, using a case study of designing a policy for stimulating the transition of the European energy system towards more sustainable functioning using five different robustness metrics. We compare the policies as identified by each metric and discuss their relative merits. We highlight that the different robustness metrics emphasize different aspects of what makes a policy robust. More specifically, measures that separate dispersion and the mean, effectively doubling the number of objectives, provide very valuable information on the trade-offs between the mean performance of the policy and dispersion around this mean. We also discuss, based on our case, why analysts should use multiple robustness metrics.

J.H. Kwakkel (✉) • S. Eker • E. Pruyt
Faculty of Technology, Policy and Management, Delft University
of Technology, Delft, The Netherlands
e-mail: J.H.Kwakkel@tudelft.nl; s.eker@tudelft.nl; e.pruyt@tudelft.nl

10.1 Introduction

In many planning problems, planners face major challenges in coping with uncertain and changing physical conditions, and rapid unpredictable socio-economic development. How should society prepare itself for this confluence of uncertainty? Given the presence of irreducible uncertainties there is no straightforward answer to this question. Effective decisions must be made under unavoidable uncertainty [10, 23]. The acceptance of uncertainty as an inevitable part of long-term decision-making has given rise to the development of new tools and approaches (see Walker et al. [36] for a review).

Robust decision making is a paradigm example of the new approach to model-based decision support in the face of deep uncertainty [13, 22]. In this approach, a very large ensemble of plausible futures spanning the various key uncertain factors is created [20]. This ensemble serves as a test bed for candidate policies. Through scenario discovery, the key vulnerabilities and opportunities of a candidate policy are identified [6]. In light of this, an iterative process of (re)design of candidate policies takes place, aimed at improving the overall robustness of the policy.

Typically, the iterative redesign of candidate policies involves the inclusion of actions whose implementation is conditional on how the future unfolds [13]. The challenge here is to avoid implementing these actions either too early or too late. Very recently, robust multi-objective optimization has been suggested as a technique for supporting the search for finding the right conditions [14]. However, a variety of alternative operationalizations of robustness have been used within a multi-objective optimization framework (e.g., [13, 15, 17, 21, 22]). In these various operationalizations, robustness is understood either as reducing the uncertainty about the expected consequences of a given policy. So no matter how the future plays out, the policy performance falls in a narrow bandwidth. Or, alternatively, robustness is understood as minimizing the undesirable outcomes. So, no matter how the future unfolds, policy performance will be satisfactory. This raises the question how the choice of the robustness metric affects the final design of an adaptive policy or plan. It also raises the question whether some robustness metrics always outperform other robustness metrics. Therefore, insight into the consequences of different robustness metrics can help analysts in choosing a (set of) metric(s) that is appropriate for the case at hand, and improve awareness regarding the relative merit of alternative robustness metrics.

In this chapter, we apply five different robustness metrics to the same case, allowing us to compare the results and providing insight into the relative merits of each of these five metrics. We start from the European energy transition case studied by Hamarat et al. [14]. This case focuses on finding an adaptive plan, built on the European emission-trading scheme that maximizes the potential of achieving the emission reduction targets set out by the European Commission. A variety of actions can be taken to help in achieving the emission reduction targets. The key question is to identify the conditions under which a given action should be taken. This can be formulated as a robust multi-objective optimization problem. In this chapter, we explore the consequences of alternative robustness metrics using this case.

The chapter is structured accordingly. In Sect. 10.2 we introduce the alternative robustness metrics. Section 10.3 provides additional details on the multi-objective optimization formulation and the approach for solving this problem, and briefly introduces the simulation model of the European energy system, the key uncertain factors that need to be accounted for, and the case specific formulation of the robust multi-objective optimization problem. Section 10.4 contains the results for each of the five robustness metrics and their comparison. A discussion of the results and their implications is presented in Sect. 10.5.

10.2 Measuring Robustness

Robust optimization methods aim at finding, in the presence of uncertainty about inputs, optimal outcomes that are not overly sensitive to any specific realization of the uncertainties [1, 3–5, 19]. In robust optimization, the uncertainty that exists about the outcomes of interest is described through a set of scenarios [25]. Robustness is then defined over this set of scenarios. This definition differs from worst-case formulations such as minimax, which can produce very costly and conservative solutions [25].

The way in which robustness is defined over the scenario set can affect the solutions that are being found. A careful choice of the robustness metric(s) is thus paramount. Three families of metrics are available: regret, satisficing, and statistical. Regret based metrics are comparative in character. They compare the performance of a policy option in a given scenario with some performance measure in the same scenario. Regret-based metrics originate from Savage [30]. Savage defines regret as the difference between a given policy's performance in a specific possible future and the performance of the best performing policy option in that specific possible future. A robust policy is the one that minimizes the maximum regret across all alternative possible futures. Alternative regret metrics use some type of baseline performance for a given scenario instead of best performing option [17, 22, 26]. In this chapter, we will not further consider regret-based metrics, for they are very hard to embed efficiently within an optimization routine.

Satisficing metrics aim at maximizing the number of scenarios, which meet a minimum performance threshold. They thus rely on the upfront specification of this performance threshold. A well-known example of this is the domain criterion [31, 32], which focuses on the fraction of the space where a given performance threshold is met; the larger this space, the more robust the policy. Often, this is simplified to looking at the fraction of scenarios, rather than the volume of the space. Recently, the domain criterion made a resurgence under the label of info-gap decision theory [2]. Because of their reliance on a user specified performance threshold, satisficing criteria introduce a new source of uncertainty into the analysis. In part because of this, we ignore these criteria in this chapter.

The third family of robustness metrics are statistical or density based. That is, they look at the distributional character of the outcomes of interest. The basic intuition of these metrics is that a robust policy leaves little uncertainty about the

expected outcomes. That is, the more peaked the distribution of expected outcomes, or the more skewed towards the desired region of outcomes, the more robust the policy. In this chapter, we focus our analysis on this family of metrics and we highlight several exemplary metrics.

For comparative purposes, we include the conservative maximin (or minimax) criterion as our first metric. It is the most conservative and focuses only on the worst case and making this worst case as good as possible

$$f_i(x) = \begin{cases} \max(\mathbf{x}_i), & \text{minimization} \\ \min(\mathbf{x}_i), & \text{maximization} \end{cases} \quad (10.1)$$

where \mathbf{x}_i is a vector of the performance of outcome indicator i in each scenario. In case of minimization, the worst case is the maximum performance. In case of maximization, the worst case is the minimum performance. Note that we take the maximum or minimum over the performance across the set of scenarios, rather than use an optimization for finding the globally worst case. Since this metric only considers the worst case, it is expected to result in conservative and costly solutions. The second metric is based on the intuition that a robust solution will have a good average result with very limited dispersion around it. In mathematical form

$$f_i(x) = \begin{cases} (\mu_i + 1)(\sigma_i + 1), & \text{minimization} \\ (\mu_i + 1)/(\sigma_i + 1), & \text{maximization} \end{cases} \quad (10.2)$$

where μ_i is the mean over the set of scenarios for outcome indicator i and σ_i is the standard deviation. The +1 is included to handle situations where either μ_i or σ_i is close to zero. This metric is essentially a signal to noise ratio, or a form of risk discounting. There are three downsides to this first metric. First, it does not provide insight into the trade-off between improving the mean and reducing the standard deviation. Second, functions that combine the mean and variance are not always monotonically increasing [28]. Third, by using the standard deviation, good and bad deviations from the mean are treated equally [34]. In many cases, however, robustness is understood as minimizing the undesirable deviations from the mean. These problems can be solved in various ways. In this chapter we consider three alternative approaches. Common to these approaches is that they have separate objectives for the mean and some measure of the deviation from this, effectively doubling the number of objectives in the eventual optimization problem.

The third metric is a variant of the approach used by Takriti and Ahmed [34], where we measure the mean and the undesirable deviations away from some target value as separate objectives

$$f_i(x) = \begin{cases} -\mu_i, \sum_{k=1}^k (x_k - q_{50})^2 [x_k > q_{50}], & \text{minimization} \\ \mu_i, -\sum_{k=1}^k (x_k - q_{50})^2 [x_k < q_{50}], & \text{maximization} \end{cases} \quad (10.3)$$

where q_{50} is the median performance, k is a scenario, x_k is the score for the i -th outcome indicator in scenario k , and the sum is only taken over the cases that meet the specified condition. So, we are taking the sum of squared differences from the median in the undesirable direction. Effectively, this metric uses the sum of squared differences as proxy for the skewness of the distribution.

The fourth metric offers an alternative way of measuring the skewness of the distribution of outcomes, using a quantile-based definition following Voudouris et al. [35]. This quantile-based definition is motivated by the potential unreliability of moment-based definitions of skewness if the density estimate is fat-tailed [8], i.e. if there are many scenarios (data points) in the tail regions.

$$f_i(x) = \begin{cases} -\mu_i, & \frac{(q_{90} + q_{10})/2 - q_{50}}{(q_{90} - q_{10})/2}, & \text{minimization} \\ \mu_i, & \frac{(q_{90} + q_{10})/2 - q_{50}}{(q_{90} - q_{10})/2}, & \text{maximization} \end{cases} \quad (10.4)$$

where q_{10} , q_{50} , and q_{90} , are the 10th, 50th, and 90th quantile respectively of the distribution of outcomes for outcome indicator i . The more positive the values of this metric, the more skewed the density estimate is towards the right (higher values). The more negative this value is, the more skewed the distribution is towards the left (lower values). Therefore, for outcome indicators to be maximized, the policy option with higher skewness is preferred, whereas for outcome indicators to be minimized, lower skewness is preferred.

Instead of considering the skewness of the distribution, one can consider the peakedness. Kurtosis is a well-known measure, established originally to describe peakedness. Among several formulations of kurtosis [16], following Voudouris et al. [35], a simple quantile-based metric is adopted here

$$f_i(x) = \begin{cases} -\mu_i, & \frac{q_{90} - q_{10}}{q_{75} - q_{25}}, & \text{minimization} \\ \mu_i, & \frac{q_{90} - q_{10}}{q_{75} - q_{25}}, & \text{maximization} \end{cases} \quad (10.5)$$

where q_{10} , q_{25} , q_{75} , and q_{90} , are the 10th, 25th, 75th, and 90th quantile, respectively, of the distribution of outcomes for outcome indicator i . The higher this metric, i.e. the smaller the inter-quartile range compared to the interval between 90th and 10th quantiles, the more peaked the density estimate is around the mean.

10.3 Case

The European Union (EU) has targets for the reduction in carbon emissions and the share of renewable technologies in the total energy production by 2020 [7]. The main aim is to reach 20% reduction in carbon emission levels compared to 1990

levels and to increase the share of renewables to at least 20 % by 2020. In order to meet the 2020 goals, the EU adopted the European Emissions Trading Scheme (ETS) for limiting the carbon emissions [7]. ETS imposes a cap-and-trade principle that sets a cap on the allowed greenhouse gas emissions and an option to trade allowances for emissions. However, current emissions and shares of renewables show a fragile progress of reaching the 2020 targets. Moreover, the energy system includes various uncertainties related to e.g. technology lifetimes, economic growth, costs, learning curves, and investment preferences. Whether the policy will achieve its targets is at least partly contingent on how these various uncertainties play out.

10.3.1 Model

In this study, a System Dynamics [11, 27, 33] model is used for simulating plausible futures of the EU electricity system. The model represents the power sector in the EU and includes congestion on interconnection lines by distinguishing seven different regions in the EU. These are United Kingdom, Ireland, Italy, and northwest, northeast, middle, southwest, and southeast Europe. Nine power generation technologies are included. These are: wind, PV solar, solid biomass, coal, natural gas, nuclear energy, natural gas with Carbon Capture and storage (CCS), coal gasification with CCS, and large scale hydro power. The model includes endogenous mechanisms and processes related to the competition between technology investments, market supply-demand dynamics, cost mechanisms, and interconnection capacity dynamics.

Figure 10.1 shows the main sub-models that constitute this model at an aggregate level. These are installed capacity, electricity demand, electricity price, profitability and levelised costs of electricity. At an aggregated level, there are two main factors that drive new capacity investments: electricity demand and expected profitability. An increase of the electricity demand leads to an increase in the installed capacity, which will affect the electricity price. This will cause a rising demand, in turn resulting in more installed capacity. On the other hand, decreasing electricity prices will lead to lower profitability and less installed capacity, which will result in electricity price increases. Each sub-model has more detailed interactions within itself and with the other sub-models and exogenous variables and these causal relationships drive the main dynamics of the EU electricity system. More detail on the model can be found in Loonen [24], including a detailed description of all equations and variables.

We are interested in exploring and analyzing the influence of a set of deeply uncertain input variables on the key output variables. In order to explore this uncertainty space, not only parametric but also structural uncertainties are included. For exploring structural uncertainties, several alternative model formulations have been specified and a switch mechanism is used for switching between these alternative formulations. Parametric uncertainties are explored over pre-defined ranges. Table 10.1 provides an overview of the uncertainties, 46 in total.

In many scenarios, the ETS policy alone will not be sufficient to achieve the stated CO₂ emission reduction targets. It thus needs not be complemented with additional actions. The reasons for not achieving the reduction targets differ from one

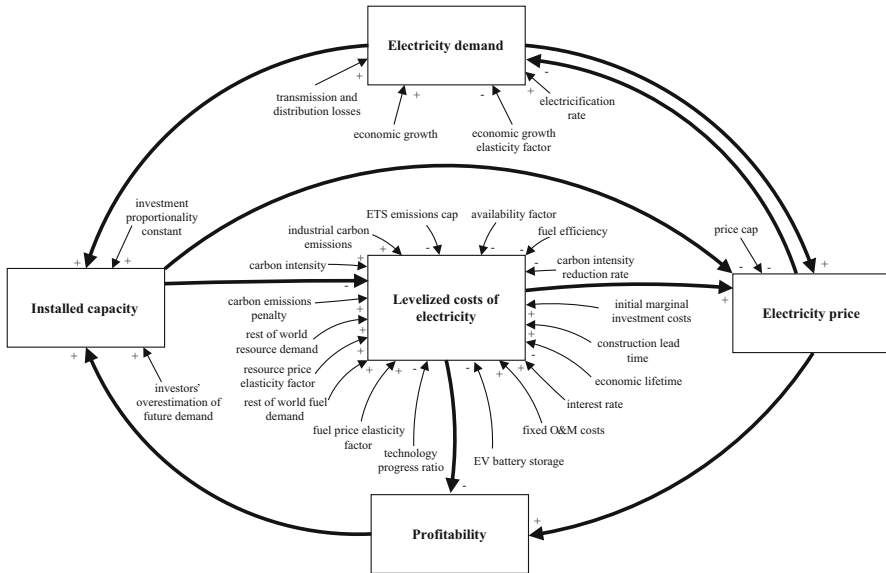


Fig. 10.1: The main causal loops in the EU energy model

scenario to the next. Therefore, the additional actions are scenario dependent. We consider three possible additional actions:

1. Obligatory phase out of older energy generation technology.
2. Subsidy for sustainable energy generation technology. The amount of the subsidy is conditioned on the differences in marginal costs between sustainable and non-sustainable technology.
3. Obligatory decommissioning of non renewable technologies to maintain the achieved fraction of renewable technologies.

Each of these actions is good only in some scenarios. For example, the third action is aimed at scenarios where the use of renewable energy collapses after a quick uptake. To address this, the challenge is to specify up front the conditions under which each of these actions should be implemented. We address this problem using multi-objective robust optimization.

Table 10.1: Specification of the uncertainties to be explored

Name	Description
Economic lifetime	For each technology, the average lifetime is not known precisely. Different ranges for the economic lifetimes are explored for each technology.
Learning curve	For each technology, the extent to which costs will decrease with increasing experience is uncertain. Different progress ratios are explored for each technology.
Economic growth	It is deeply uncertain how the economy will develop over time. Six possible developments of economic growth behaviors are considered.
Electrification rate	The rate of electrification of the economy is explored by means of six different electrification trends.
Physical limits	The effect of physical limits on the penetration rate of a technology is unknown. Two different behaviors are considered.
Preference weights	Investor perspectives on technology investments are treated as being deeply uncertain. Growth potential, technological familiarity, marginal investment costs and carbon abatement are possible decision criteria.
Battery storage	For wind and PV solar, the availability of (battery) storage is difficult to predict. A parametric range is explored for this uncertainty.
Time of nuclear ban	A forced ban for nuclear energy in many EU countries is expected between 2013 and 2050. The time of the nuclear ban is varied between 2013 and 2050.
Price demand elasticity	A parametric range is considered for price demand elasticity factors.

10.3.2 Formulating the Problem

The general optimization problem we are solving is

$$\text{minimize } F(L) = [f_{costs}, -f_{renewables}, -f_{reduction}]$$

where $L = [l_{df}, l_{ad}, l_{sf}, l_{sd}, l_{pr}, l_{dcf}, l_{fth}, l_{tr}]$

$$\begin{aligned} \text{subject to } & 0.5 \leq c_{df} \leq 1, \\ & 0.0 \leq c_{ad} \leq 0.75 \\ & 0.0 \leq c_{sf} \leq 0.5 \\ & 0.0 \leq c_{sd} \leq 20.0 \\ & 1.0 \leq c_{pr} \leq 2.0 \\ & 0.0 \leq c_{dcf} \leq 0.5 \\ & 0.0 \leq c_{fth} \leq 1.0 \\ & 10 \leq c_{tr} \leq 40 \end{aligned}$$

Table 10.2 offers an explanation of each policy lever l , and the meaning of the subscripts for both l_i and C_i . f_{costs} , $f_{renewables}$, and $f_{reduction}$ are the costs of the policy, the fraction of renewables at the end of the simulation, and the reduction

of emissions of green house gases respectively. The constraints c_i are taken from Hamarat et al. [14] and are based on common sense and case specific considerations.

Various approaches exist for solving multi-objective optimization problems. Over the last decade, substantial advances have been made through the use of genetic algorithms (GA). GA use a population of solutions, which are evolved over the course of the run of the algorithm. This population can be evolved in such a way that it maintains diversity, while continually moving towards the Pareto frontier. In this

Table 10.2: List of triggers and their descriptions

	Trigger	Brief description
Action 1	Desired fraction (df)	Desired fraction of renewable technologies.
	Additional decommissioning (ad)	Additional fraction of non-renewable technologies to be decommissioned.
Action 2	Subsidy factor (sf)	Additional fraction of subsidy for renewables.
	Subsidy duration (sd)	Duration for how long the subsidy for the renewables will be active.
	Proximity (pr)	Proximity of cost to the cost of the most expensive non-renewable technology.
Action 3	Decommissioning factor (dcf)	Fraction to be decommissioned for non-renewables when the gap between desired and forecasted fraction for renewables is above the Trigger.
	Forecast time horizon (fth)	Time horizon over which the forecast for the level of renewable fraction is done.
	Trigger (tr)	Proximity of the forecasted renewable fraction to the desired fraction.

way, multiple Pareto front solutions can be found in a single run of the algorithm [9]. Currently, a wide variety of alternative multi-objective evolutionary algorithms are available for solving multi-objective optimization problems [12, 18]. In this study, we use Borg, a state of the art GA where the evolutionary operators co-evolve with the search [12]. In various comparisons, Borg has been demonstrated to be among the best available genetic algorithms for solving multi objective optimization problems [12, 29].

10.4 Results

Figure 10.2 shows the ϵ -progress of the genetic algorithm for each of the five alternative robustness metrics (10.1)–(10.5). ϵ -Progress measures how often the genetic algorithm has been able to find a substantially better solution. Over the course of the optimization, this will occur less frequently. As a result, ϵ -progress stabilizes, indicating that the algorithm has converged. The number of function evaluations required to achieve a stable set of solutions differs from one robustness metric to

the next. For example, the second robustness metric stabilized very quickly, while the third and fourth robustness metrics require a substantially larger number of function evaluations. A first reason for this is that the ε values are different for non-comparative metrics. A second explanation is the fact that the first two robustness metrics use a single metric, while metrics three, four and five use two metrics. This effectively doubles the solution space.

Figure 10.3 shows the values for the decision variables for all of the solutions, grouped by metric. In this figure, each line represents a solution, and the intersection point of this line with a vertical axis shows the value of the corresponding decision variable in this solution. We observe several things. First, virtually all metrics use the edges of the solutionspace for several of the decision variables. Second, many of

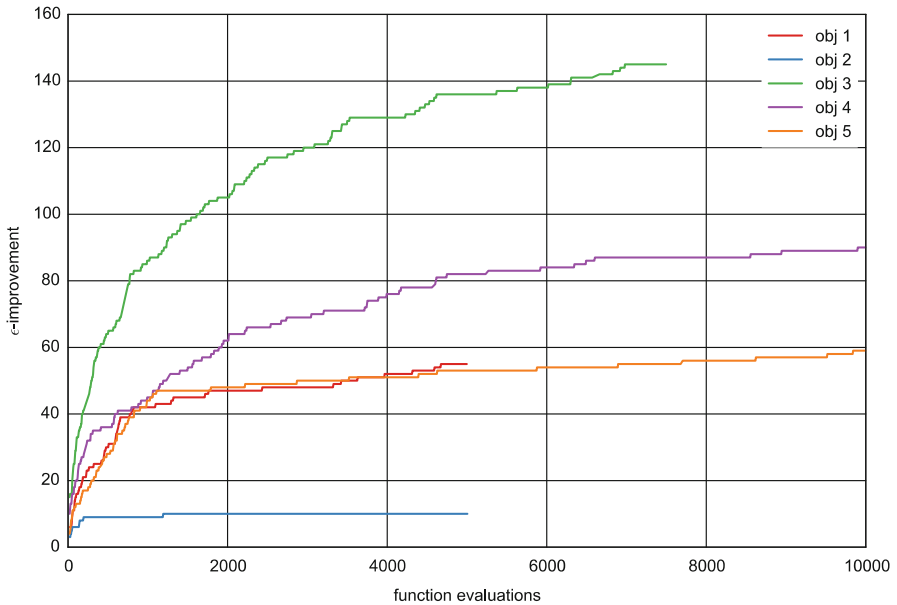


Fig. 10.2: ε -Progress for the five objective functions (robustness metrics)

the metrics go to the upper extreme of the decommissioning factor. This means that obligatory additional decommissioning of fossil-based energy generation is effective across metrics. We also see that virtually all solutions combine this with a low value for the trigger. This means that even if there are small differences between the desired and forecasted fraction of renewables, the various robustness metrics favor aggressive decommissioning. On the other levers, the pattern is less clear. Still, we observe that for example metric 1 uses the upper extreme of the desired fraction of renewables in combination with aggressive additional commissioning of sustainable energy generation technologies. Metric 1 aims at minimizing the worst case, so to minimize the worst case, very aggressive promotion of sustainable energy should be pursued. Similar observations can be made for the other metrics.

Solving the multi-objective optimization problem for each of the metrics produces a set of solutions that are non-dominated. Figure 10.3 shows the values for the decision variables for each solution per robustness metric. There are clear differences in this, but how does this translate to differences in outcomes? To analyze this, we look at the cumulative distribution of the terminal values for the outcomes of interest. This is shown in Fig. 10.4. We see that the shape of the distribution is quite similar for all solutions found by each of the five metrics, but some metrics produce better results than others. For example, if we look at the fraction of renewables (Fig. 10.4a), we see that robustness metric 5 produces a set of solutions that all

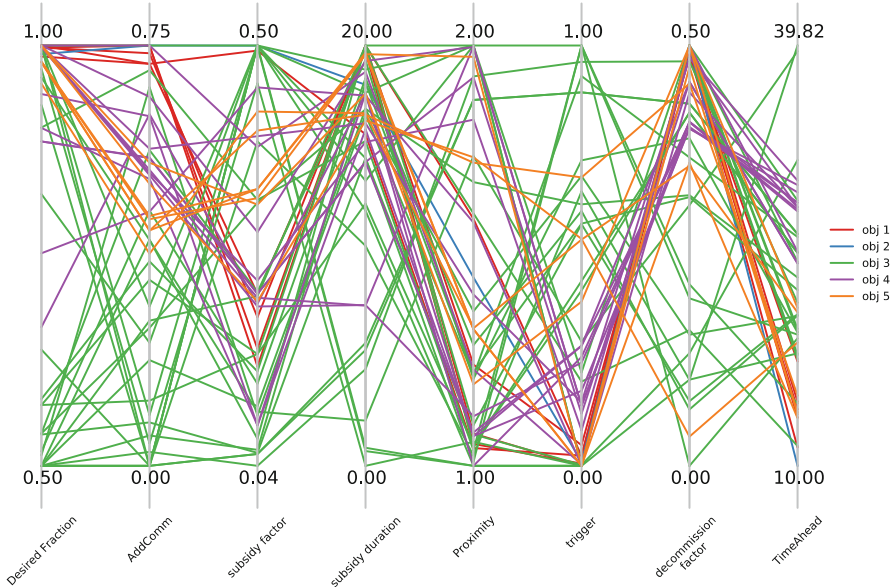


Fig. 10.3: Values for policy levers for all solutions, colored by objective function (robustness metric) used

result in a high fraction of cases with substantial adoption of renewables. In contrast, metrics 3 and 4 produce a set of solutions with a wider spread of results.

In order to get sharper insight into how different the results are for all of the solutions, we calculate the distance between all cumulative distributions using the Kolmogorov-Smirnov statistic. The result for each of the three outcome indicators is shown in Fig. 10.5a–c. Each of these figures shows a matrix, and each cell is colored according to the difference between the cumulative distributions of the two solutions in the row and column of this cell. Here, we can see that the one solution found by using metric number 2 (marked as obj 2–1) is equal to or better than all other solutions. The second best set of solutions is produced by metric 5 where almost all solutions are very close to the solution found by metric 2. This observation is true for all three indicators. So both metric 2 and 5 produce solutions that are

both relatively cheap, and effective in reducing CO₂ emissions and increasing the adoption of sustainable energy generation technologies.

So far, we have focused on the individual outcomes of interest. A major advantage of the multi-objective optimization approach is that we can get insight into the trade-offs between the various outcomes of interest. To this end, we created a set of pair-wise scatter plots for each of the five robustness metrics, with Gaussian kernel density estimates on the diagonal. The resulting set of figures is shown in Fig. 10.6. We observe that irrespective of the robustness metric, the solutions with a high adoption of sustainable energy generation also tend to be cheaper across virtually all scenarios. This is explained by the fact that pursuing high adoption requires aggressive

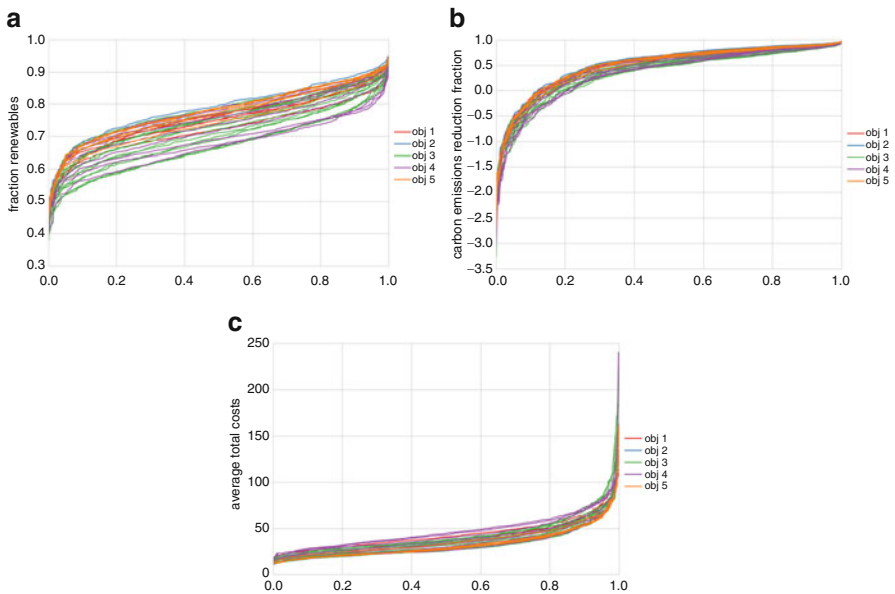


Fig. 10.4: Cumulative distributions of the terminal values of the three performance indicators across all solutions. **(a)** The fraction of renewables. **(b)** The fraction of emission reduction. **(c)** The costs of the policy

strategies in the short term which result in learning effects, in turn lowering the price and increasing the efficiency of sustainable energy generation technologies.

Figure 10.6 also highlights some of the key differences between the various metrics. For example, metric 1, which focuses on the worst case only, performs poorly across the range of scenarios even though its worst case is the least worst case possible. Metrics 3, 4, and 5 which include both the average performance as well as a metric related to the deviation from this average produce a larger set of solutions. For example, metric 4 offers some examples of solutions that appear to have a slightly lower average performance, but are more reliable with respect to their expected performance, or at least have a less wide range of downside outcomes. This is most clearly observable for the fraction of renewables.

10.5 Discussion

Comparing the results identified by the five different robustness metrics, we observe several things. First, metric 1 focuses on minimizing the worst case, but at the expense of the performance in all other scenarios. A worst case approach should thus be used only in very specific circumstances. Second, metrics 2 and 5, both of which focus on the peakedness of the distribution appear to produce the best sets of solutions. For this case, there appears to be little benefit to using a metric that focuses on the downside risk of outcomes instead of the overall distribution of outcomes. Still, both metric 3 and 4, which do this, can produce results on par with metrics 2 and 5. We see in Fig. 10.5 that both solution 4–11 and 3–16 are on par with the best solutions produced by metrics 2 and 5.

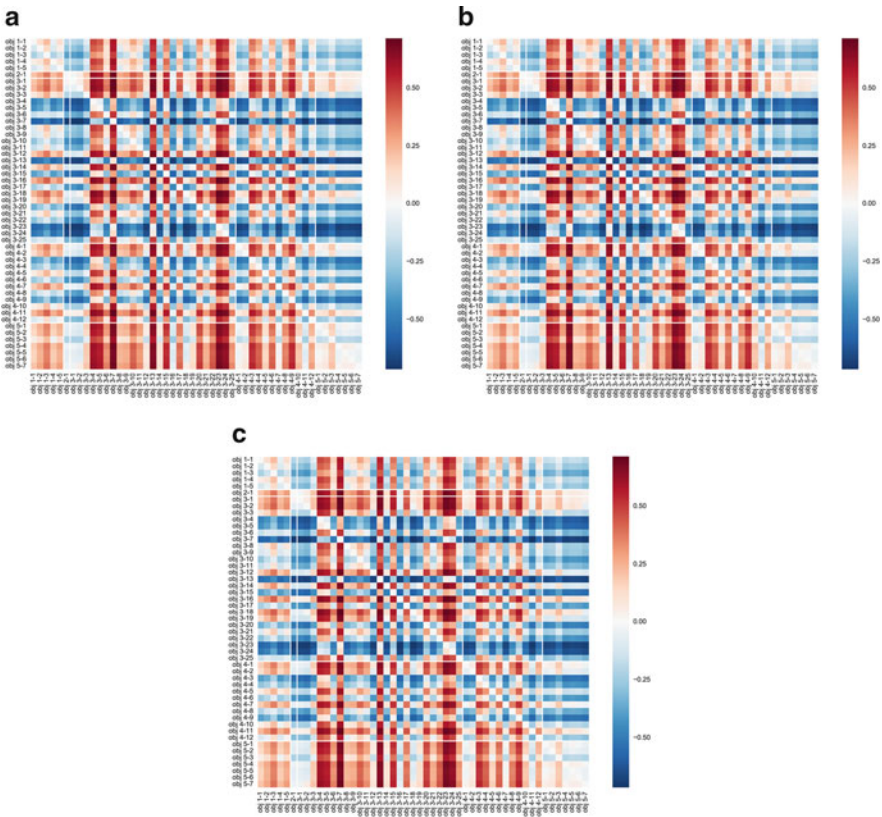


Fig. 10.5: Distances between cumulative distributions of the terminal values for the three outcome indicators as measured by the Kolmogorov-Smirnov statistic. (a) The fraction of renewables. (b) The fraction of emission reduction. (c) The costs of the policy

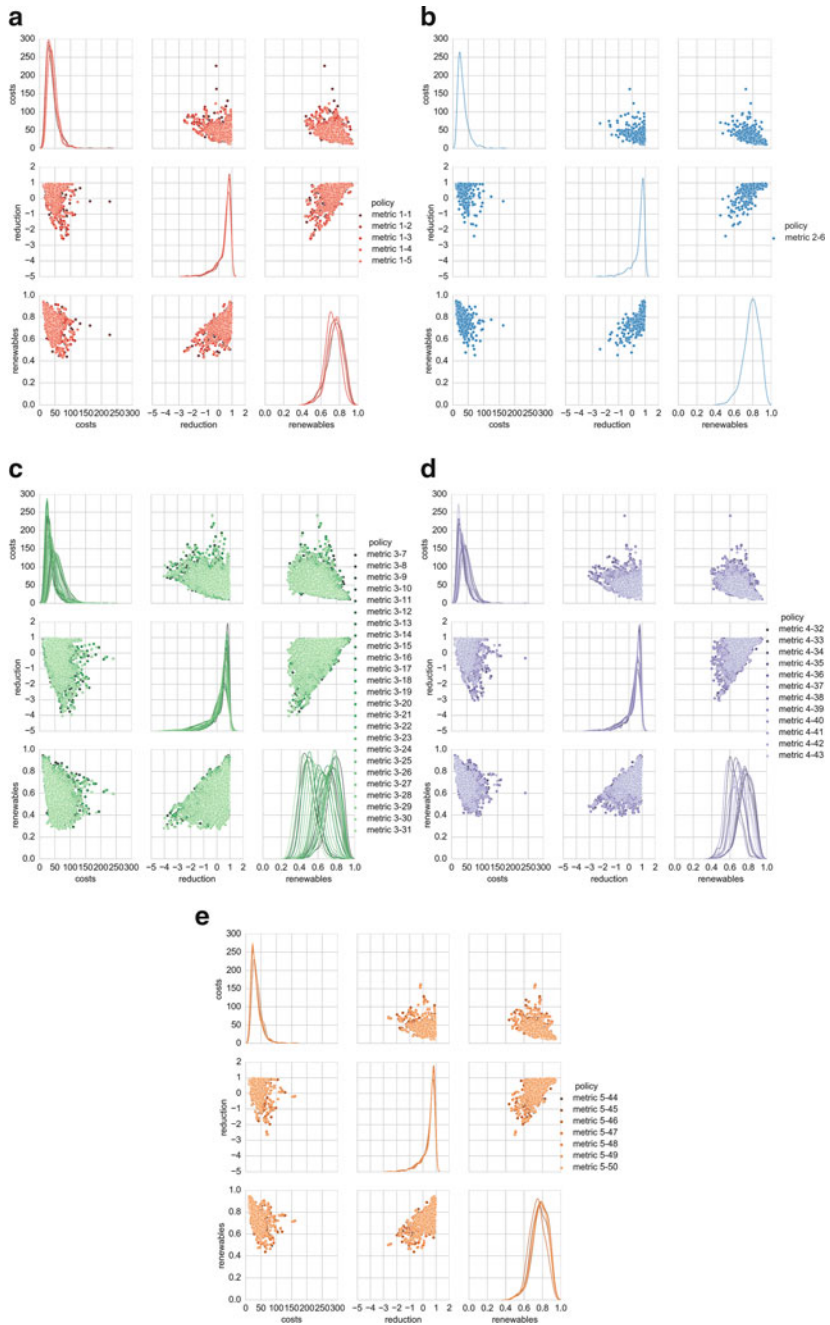


Fig. 10.6: Pairwise scatter plots for each solution across all scenarios for all five robustness metrics. (a) Metric 1, (b) metric 2, (c) metric 3, (d) metric 4, (e) metric 5

The results suggest that there is no single best robustness metric. Metric 1 is useful if the worst case is the sole concern. Metric 2 and 5 are useful if one is interested in reducing the uncertainty about expected outcomes but one is indifferent with respect to upside and downside deviations. Metrics 3 and 4 are useful if one is primarily concerned about negative deviations from the average performance. Metrics 3, 4, and 5 produce insight into the trade-offs between average performance and the deviations from this average. This can be very useful decision support information, but it doubles the solution space. In the present chapter, we have used the same robustness metric for all outcome indicators. Evidently a mixed approach can be pursued. For example, users might prefer certainty about costs, while being concerned about negative deviations for adoption of sustainable energy generation. In this case either metric 2 or 5 can be used for costs, while either metric 2 or metric 3 would be used for the fraction of renewables. In short, case specific considerations and system characteristics should be considered in choosing statistical robustness metrics.

The results are based on a single run of the optimization algorithm for each metric. It is good practice to assess the adequacy of solutions found through genetic algorithms by performing several replications. Genetic algorithms exploit stochasticity for effective searching, but this also introduces some randomness in the algorithm. Performing several replications with different random seeds can enhance the confidence that the identified solutions are indeed good approximations of the true Pareto front.

Robust optimization requires evaluating the performance of a given solution over a set of scenarios. This creates substantial runtime concerns. In the case reported here, we evaluated the robustness over a set of 500 scenarios. Each of these scenarios requires the running of the simulation model, which takes a few seconds. Finding effective ways of reducing the size of the set of scenarios needed for calculating the robustness metric can help in substantially reducing the calculation time. If we are able to reduce the size of the set from a 500 to 250, the runtime would be halved.

In this chapter, we focused on five examples of statistical robustness metrics. We have not considered satisficing or regret based metrics. It will be quite interesting to extend the presented analysis to also include examples of both families of robustness metrics. This would offer a more comprehensive insight into the merits of examples of all three families of robustness metrics.

The implication of the results presented in this chapter is twofold. First, there is no clearly superior single robustness metric. Case specific consideration and system characteristics affect the merits of the various robustness measures. This implies that an analyst has to choose carefully which robustness measure is being used and assess its appropriateness. Second, because of the different insights generated by the different robustness metrics, it is advised to consider multiple robustness measures simultaneously and explore their joint implications for decision-making. For example, in this case, all metrics favor aggressive additional decommissioning of existing fossil fuel based energy generation. This might be a strong argument for including this action as part of the overall strategy. In short, when deciding on robustness metrics, use multiple and choose with care.

References

1. Bai, D., Carpenter, T., Mulvey, J.: Making a case for robust optimization models. *Manag. Sci.* **43**(7), 895–907 (1997)
2. Ben-Haim, Y.: *Information-Gap Decision Theory: Decisions Under Severe Uncertainty*. Academic, London (2001)
3. Ben-Tal, A., Nemirovski, A.: Robust convex optimization. *Math. Oper. Res.* **23**(4), 769–805 (1998)
4. Ben-Tal, A., Nemirovski, A.: Robust solutions of linear programming problems contaminated with uncertain data. *Math. Program.* **88**(3), 411–424 (2000)
5. Bertsimas, D., Sim, M.: The price of robustness. *Oper. Res.* **52**(1), 35–53 (2004)
6. Bryant, B.P., Lempert, R.J.: Thinking inside the box: a participatory, computer-assisted approach to scenario discovery. *Technol. Forecast. Soc. Chang.* **77**(1), 34–49 (2010)
7. Commission, E.: *Europe 2020 a strategy for smart, sustainable and inclusive growth*. COM **2020**(33) (2010)
8. Davidson, R.: Statistical inference in the presence of heavy tails. *Econ. J.* **15**(1), C31–C53 (2012)
9. Deb, K., Pratap, A., Agarwal, S., Meyarivan, T.: A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evol. Comput.* **6**(2), 182–197 (2002)
10. Dessai, S., Hulme, M., Lempert, R., Pielke, R.: Do we need better predictions to adapt to a changing climate? *EOS* **90**(13), 111–112 (2009)
11. Forrester, J.W.: *Industrial Dynamics*. MIT, Cambridge (1961)
12. Hadka, D., Reed, P.: Borg: An auto-adaptive many-objective evolutionary computing framework. *Evol. Comput.* **21**(2), 231–259 (2013)
13. Hamarat, C., Kwakkel, J.H., Pruyt, E.: Adaptive robust design under deep uncertainty. *Technol. Forecast. Soc. Chang.* **80**(3), 408–418 (2013)
14. Hamarat, C., Kwakkel, J.H., Pruyt, E., Loonen, E.T.: An exploratory approach for adaptive policymaking by using multi-objective robust optimization. *Simul. Model. Pract. Theory* **46**, 25–39 (2014)
15. Herman, J.D., Reed, P.M., Zeff, H.B., Characklis, G.W.: How should robustness be defined for water systems planning under change? *J. Water Resour. Plan. Manag.* **141**(10), 04015012 (2015)
16. Joanes, D.N., Gill, C.A.: Comparing measures of sample skewness and kurtosis. *J. R. Stat. Soc. D (The Statistician)* **47**(1), 183–189 (1998)
17. Kasprzyk, J.R., Nataraj, S., Reed, P.M., Lempert, R.J.: Many objective robust decision making for complex environmental systems undergoing change. *Environ. Model. Softw.* **42**, 55–71 (2013)
18. Kollat, J.B., Reed, P.M.: Comparing state-of-the-art evolutionary multi-objective algorithms for long-term groundwater monitoring design. *Adv. Water Resour.* **29**(6), 792–807 (2006)
19. Kouvelis, P., Yu, G.: *Robust Discrete Optimization and Its Applications*. Kluwer Academic Publishers, Dordrecht (1997)
20. Kwakkel, J.H., Pruyt, E.: Exploratory modeling and analysis, an approach for model-based foresight under deep uncertainty. *Technol. Forecast. Soc. Chang.* **80**(3), 419–431 (2013)
21. Kwakkel, J.H., Haasnoot, M., Walker, W.E.: Developing dynamic adaptive policy pathways: a computer-assisted approach for developing adaptive strategies for a deeply uncertain world. *Clim. Chang.* **132**, 373–386 (2015)
22. Lempert, R.J., Collins, M.T.: Managing the risk of uncertain threshold responses: comparison of robust, optimum, and precautionary approaches. *Risk Anal.* **27**(4), 1009–1026 (2007)
23. Lempert, R.J., Popper, S., Bankes, S.: *Shaping the Next One Hundred Years: New Methods for Quantitative, Long-Term Policy Analysis*. RAND Corporation, Santa Monica (2003)
24. Loonen, E.: Exploring carbon pathways in the EU power sector, using exploratory system dynamics modelling and analysis to assess energy policy regimes under deep uncertainty. Master's thesis, Delft University of Technology, Delft (2012)

25. Mulvey, J.M., Vanderbei, R.J., Zenios, S.A.: Robust optimization of large-scale systems. *Oper. Res.* **43**(2), 264–281 (1995)
26. Popper, S.W., Griffin, J., Berrebi, C., Light, T., Min, E.Y.: *Natural Gas and Israel's Energy Future: A Strategic Analysis Under Conditions of Deep Uncertainty*. RAND, Santa Monica (2009)
27. Pruyt, E.: *Small System Dynamics Models for Big Issues: Triple Jump Towards Real-World Complexity*. TU Delft Library, Delft (2013)
28. Ray, P.A., Watkins, Jr. D.W., Vogel, R.M., Kirshen, P.H.: Performance-based evaluation of an improved robust optimization formulation. *J. Water Resour. Plan. Manag.* **140**(6), 04014006 (2013)
29. Reed, P.M., Hadka, D., Herman, J.D., Kasprzyk, J.R., Kollat, J.B.: Evolutionary multiobjective optimization in water resources: the past, present, and future. *Adv. Water Resour.* **51**, 438–456 (2013)
30. Savage, L.: *The Foundations of Statistics*. Wiley, New York (1954)
31. Schneller, G.O., Sphicas, G.P.: Decision making under uncertainty: Starr's domain criterion. *Theor. Decis.* **15**(4), 321–336 (1983)
32. Starr, M.K., Starr, M.K.: *Product Design and Decision Theory*. Prentice-Hall, Englewood Cliffs (1963)
33. Sterman, J.D.: *Business Dynamics—Systems Thinking and Modeling for a Complex World*. McGraw-Hill, Irwin (2000)
34. Takriti, S., Ahmed, S.: On robust optimization of two-stage systems. *Math. Program.* **99**(1), 109–126 (2004)
35. Voudouris, V., Matsumoto, K., Sedgwick, J., Rigby, R., Stasinopoulos, D., Jefferson, M.: Exploring the production of natural gas through the lenses of the ACEGES model. *Energy Policy* **64**, 124–133 (2014)
36. Walker, W.E., Haasnoot, M., Kwakkel, J.H.: Adapt or perish: a review of planning approaches for adaptation under deep uncertainty. *Sustainability* **5**(3), 955–979 (2013)