ESCUELA DE
NEGOCIOS

UNIVERSIDAD ADOLFO IBÁÑEZ

WORKING PAPER Nº2

A Machine Learning and Distributionally Robust Optimization Framework for Strategic Energy Planning under Uncertainty

AUTHOR: BABONNEAU, FREDERIC

School Of Business Universidad
Adolfo Ibáñez

A Machine Learning and Distributionally Robust Optimization Framework for Strategic Energy Planning under Uncertainty

E. Guevara[1], F. Babonneau[2], T. Homem-de-Mello[2], and S. Moret[3]

[1]PhD Program in Industrial Engineering and Operations Research, Universidad Adolfo Ibáñez

[2]Escuela de Negocios, Universidad Adolfo Ibáñez

[3]Business School, Imperial College London

Technical report

Abstract

This paper investigates how the choice of stochastic approaches and distribution assumptions impacts strategic investment decisions in energy planning problems. We formulate a two-stage stochastic programming model assuming different distributions for the input parameters and show that there is significant discrepancy among the associated stochastic solutions and other robust solutions published in the literature. To remedy this sensitivity issue, we propose a combined machine learning and distributionally robust optimization (DRO) approach which produces more robust and stable strategic investment decisions with respect to uncertainty assumptions. DRO is applied to deal with ambiguous probability distributions and Machine Learning is used to restrict the DRO model to a subset of important uncertain parameters ensuring computational tractability. Finally, we perform an out-of-sample simulation process to evaluate solutions performances. The Swiss energy system is used as a case study all along the paper to validate the approach.

Keywords: Strategic energy planning, Electricity generation, Uncertainty, Distributionally robust optimization, Machine learning.

1 Introduction

Long-term energy planning for large-scale energy systems identifies strategic capacity investment decisions in energy conversion technologies to guarantee our future energy supply. The planning horizon is generally long enough, i.e., 20-50 years, to offer a possibility for the energy system to have a complete technology mix turnover. Optimization models, in particular, aim at finding an optimal strategy that minimizes the total investment and operations cost on the whole planning horizon. Among the most commonly used optimization energy models are MARKAL/TIMES [Krzemie, 2013], OSeMOSYS [Howells et al., 2011], ETEM [Babonneau et al., 2017], MESSAGE [Sullivan et al., 2013], SMART [Powell et al., 2012], oemof [Hilpert et al., 2018], Calliope [Pfenninger and Pickering, 2018] and EnergyScope [Limpens et al., 2019]. Usually, these large-scale models are multi-sector (e.g., electricity, heating, mobility) and consider multiple investment periods and few typical days for each period. An inherent characteristic of these models, as shown in Moret et al. [2017], is the lack of reliable data (due to errors in long-term forecasts) and, more generally, the presence of many uncertain input parameters. Such features lead to difficulties in analyzing the solutions and expose the identified strategies to a high risk of sub-optimality when the future deviates from the forecast expectations.

Both Stochastic Programming (SP) and, more recently, Robust Optimization (RO) have been widely used to deal with uncertainty in optimization energy models. In short, SP finds the decision that optimizes the expected value (or a more general risk function) of the objective, where the expectation is computed with respect to the probability distribution of the random variables representing the uncertainty in the problem. Because such probability distributions are often defined over a very large or even infinite number of possible realizations, sampling and/or decomposition approaches are typically applied in order to solve such problems numerically. Comprehensive discussions of theoretical and algorithmic aspects of SP can be found in Birge and Louveaux [2011], Shapiro et al. [2014]. A well-known limitation of SP, however, is the difficulty in defining the probability distribution functions (PDF) and the high sensitivity of the computed solutions to the assumed PDFs.

The RO method can be regarded as a min-max approach to consider uncertainty in optimization models. Unlike SP, it does not require the definition of specific PDFs. Instead, RO defines first an uncertainty set of possible realizations in an explicit way as, e.g., ranges of variation, based on partial known information on the uncertain parameters. Then, it looks for solutions that remain feasible for all realizations of the uncertain parameters within the uncertainty set. A drawback of such formulation is that it typically generates very conservative solutions, thereby increasing the investment cost of the solutions. Some approaches to circumvent that problem have been proposed—for instance, the definition of an *uncertainty budget* so that not all variables are allowed to take on their worst-case values simultaneously [Bertsimas and Sim, 2004]. A comprehensive discussion of RO can be found in Ben-Tal et al. [2009].

As a direct consequence of the aforementioned limitations, in the literature the use of both SP and RO in long-term energy planning models has been restricted to few uncertain parameters. In Babonneau et al. [2012], the authors address the issue of uncertain energy supplies in a robust formulation of the TIMES model. Powell et al. [2012] apply SP to cope with different sources of uncertainty, such as the energy of wind, energy demands and resource prices. In a more recent contribution, Moret et al. [2020a] propose a robust optimization framework that allows for the consideration of all uncertain parameters in the long-term energy planning EnergyScope model [Limpens et al., 2019]. However, the dynamics of recourse actions is not modelled, essentially to keep their formulation tractable, which according to the authors may lead to conservative solutions.

The present paper proposes alternative approaches to address these modelling and computational issues. More precisely, its contribution is twofold. First, we implement a SP formulation of the EnergyScope model considering, as in Moret et al. [2020a], all sources of uncertainty and assuming different PDFs to highlight their potential impact on the strategic decisions of investment in such long-term models. We compare these solutions with the robust ones published in Moret et al. [2020a].

Second, we propose a novel combined Machine Learning and Distributionally Robust Optimization (DRO) approach which allows us to obtain a numerically tractable, recourse-based robust formulation of the EnergyScope model that is far less sensitive to the choice of PDFs.

DRO has been introduced in the literature to compute robust solutions for stochastic problems assuming ambiguous probability distributions, i.e, when the true PDF of the uncertain parameters is unknown. DRO is based on the design of a set of distributions —called an *ambiguity set*—and it aims at providing the model with protection against the worst distribution within that set; see, for instance, Wiesemann et al. [2014]. The ambiguity set is calibrated assuming a distance measure (e.g., Wasserstein, [Gibbs and Su, 2002]) that differs according to the different DRO approaches.

DRO has been recently applied to energy problems, mainly to unit commitment (UC). In Xiong et al. [2017], the authors consider a UC model with uncertain wind power generation which is captured by an ambiguity set describing a family of wind energy distributions. They show that DRO generally outperforms the conventional RO method yielding lower expected costs. In Duan et al. [2018], where uncertainty on the forecasting of renewable generation and load is considered, similar results are obtained; the DRO operating costs appear to be lower than the ones associated to the standard RO solution and higher than the cost of the SP solution. However, DRO solutions vary less with respect to the underlying distributions, thus producing more robust decisions. Recently, DRO has been applied to a generation expansion planning (GEP) model [Han and Hug, 2019] where the goal is to minimize investment and operating cost, with uncertain demand, wind and PV generation forecasts. The work of Han and Hug [2019] focuses on investment of decentralized energy resources (DERs) at the distribution level and does not consider strategic centralized investments. DRO has also been applied to deal with uncertainty in problems of economic dispatch [Chen et al., 2016], day ahead scheduling of energy and reserve [Xiong and Singh, 2017], optimal power flow [Guo et al., 2019] and transmission expansion planning [Pozo et al., 2018, Velloso et al., 2018].

Here, we build an ambiguity set in which we assume that the true PDFs are close (using a Wasserstein distance) to a given reference distribution. Then, we generate a robust two-stage long-term energy planning model with a large number of uncertain parameters which unfortunately is not computationally tractable. Thus, to avoid computational issues, we use Machine Learning (ML) tools to rank and select the most important uncertain parameters to be included in the definition of the ambiguity set. In short, ML is a branch of artificial intelligence devoted to developing intelligent systems that learn from data. In the context of supervised learning, the modeler gives the machine/algorithm information about a set of characteristics and responses called labels, in order to learn how to make predictions or classifications. On the other hand, not all information that can be given to the machine/algorithm will provide better learning, which leads to the issue of choosing the most relevant variables to the model, a process called *variable selection*. Doing so reduces considerably the number of variables used, which produces several benefits: ease to visualize and understand the data, elimination of irrelevant or redundant variables, reduction of storage requirements, and reduction of computational times, to name a few.

By combining the ML-based selection and the DRO approach in this novel way, we are able to select the important variables of the problem in a more systematic fashion than what is accomplished with classical sensitivity analysis techniques. Such an approach yields a tractable robust version of the EnergyScope model that uses probability distributions for the uncertainty but is not very sensitive to variations in those PDFs. Finally, the DRO solutions are compared to previously computed RO and SP solutions. To the best of our knowledge, it is the first implementation of a DRO strategic energy planning model that considers an entire national energy system.

The rest of this paper is organized as follows. In Section 2, we present a compact formulation of the EnergyScope model and we describe the different sources of uncertainty in such model. In Section 3, we introduce a two-stage SP formulation assuming different underlying PDFs to assess the impact of such assumptions on strategic investment decisions. In section 4, we describe the combined ML and DRO framework we implement to produce a tractable robust dynamic planning energy model. The experimental results and main findings using the ML-DRO framework are discussed in Section 5. We show in particular that our approach produces robust and stable strategic solutions with respect to assumptions on the reference probability distributions. Finally, concluding remarks are presented in Section 6.

2 Strategic energy models and uncertainty

In this section, we first describe the strategic energy model introduced in Moret et al. [2020a], Moret [2017] that we use in the present paper. For the sake of simpler notations throughout the paper, we present a compact mathematical formulation and report the complete model in Appendix A for interested readers. Then we discuss model uncertainties as identified and characterized in Moret et al. [2017].

2.1 Compact mathematical formulation

A mixed-integer linear programming (MILP) formulation for strategic planning of energy systems was first introduced by Moret et al. [2016] and used in Moret [2017], Moret et al. [2020a][1]. It is a multi-sector multi-energy model calibrated on the national energy system of Switzerland. It considers the long-term planning of the energy system until 2035, with a single period "snapshot" formulation (optimization over one target year) which takes into account the seasonality of the year by months. The investment strategy is decided under the "here and now" paradigm, considering the demands and operations constraints in last year of planning. It incorporates information on the demand for end-use (electricity, heating and transportation), the efficiency and cost of technologies, the cost of resources (imported and local) and their availability as well as storage units characteristics. The demand for heating is divided into industrial, centralized and decentralized; the demand for transport is divided into the passengers and freight sectors.

The compact MILP formulation of the energy planning model is given as follows:

$$\begin{aligned} \text{minimize} \quad & c^T\mathbf{x} + e^T\mathbf{y} && \text{(1a)} \\ \text{subject to} \quad & A\mathbf{x} \leq b, && \text{(1b)} \\ & T\mathbf{x} + W\mathbf{y} \geq d, && \text{(1c)} \\ & \mathbf{x} \in X, && \text{(1d)} \\ & \mathbf{y} \in Y, && \text{(1e)} \end{aligned}$$

[1]The code is publicly available at https://github.com/energyscope/EnergyScope/tree/v1.0

where $\mathbf{x}$ represents the strategic investment decisions, and the set $X \subseteq \mathbb{R}_{+}^{n_1-q_1} \times \mathbb{Z}_{+}^{q_1}$ imposes constraints related to the nature of the variables (continuous and integer). The variables $\mathbf{y}$ represent the operation decisions, where the set $Y \subseteq \mathbb{R}_{+}^{n_2}$.

The objective of the problem is to minimize the total discounted cost of investment and operation over the planning horizon. The first term of the objective function defines annualized investment and maintenance costs for each technology and the second term defines the annualized operations cost. Constraints (1b) represent in a simplified way several system constraints that do not depend on the operation variables, such as: the existing capacity, the potential for each technology and additional system specifications on for example electricity and decentralized heating networks. Constraints (1c) are related to system operations, defining the annual and monthly capacity availability for technologies, imported and local resources bounds, supply-demand balance and the constraints on operation of storage units. It can be said that system operations depend on both investment ($\mathbf{x}$) and operation ($\mathbf{y}$) decisions, in the sense that investment decisions alter the available capacity configurations and thus the operations of the system.

Although the model has a multi-sector description (i.e., electricity, heating and transportation), we focus our analysis in the rest of the paper on the electricity sector to assess the impact of uncertainty on strategic investment decisions in power generation.

2.2 Model uncertainty

As discussed in Moret et al. [2017], uncertain parameters in (1) appear everywhere in the model, both in the objective and the constraints. The authors classified these parameters according to their similarities, with a total of 240 important uncertain parameters. In the objective function, there are 160 uncertain parameters broken down into: discount rate (1 parameter), resources costs (8 parameters), investment costs of technologies (52 parameters), maintenance costs of technologies (48 parameters) and lifetimes of technologies (51 parameters). In the constraints, there are 80 uncertain parameters broken down into: technology efficiencies (65 parameters) and end-use energy demands (15 parameters). In Table 1 we summarize the main uncertain parameters, with their range of variation relative to their nominal values (corresponding to the median) estimated in Moret et al. [2017] and their localization in the compact model formulation (1).

Parameters	Min %	Max %	Element in Problem (1)
	Investment Cost		
PV	-39.6%	39.6%	c
Wind	-21.6%	22.9%	c
Nuclear	-21.6%	119.3%	c
Hydro Dam	-21.6%	73.8%	c
Hydro River	-21.6%	21.6%	c
Geothermal	-39.7%	62.1%	c
Thermal power plant	-21.6%	25.0%	c
District Heating Network	-39.3%	39.3%	c
Decentralized NG Boilers	-21.6%	21.6%	c
	Resources Cost		
Local	-2.9%	2.9%	e
Import	-47.3%	89.9%	e
	End-Uses Demand		
Transportation	-3.4%	3.4%	d
Services	-7.4%	4.1%	d
Industry	-10.5%	5.9%	d
Households	-6.9%	4.3%	d
	Technologies efficiency		
Boilers	-5.7%	5.7%	W
Gasoline car	-20.6%	20.6%	W
PV	-20.8%	20.8%	W
Fuel Cell Car	-28.7%	28.7%	W
	Others		
Discount rate	-46.2%	46.2%	c
Maintenance Cost	-48.2%	35.7%	c
Technology lifetime	-26.5%	26.5%	c
Monthly capacity factor	-11.1%	11.1%	T

Table 1: Ranges of variations relative to the nominal values for the main uncertain parameters, taken from Moret et al. [2017]. The parameters are identified in Problem (1) through the elements c, e, d, T and W.

3 Assessing the impact of distribution assumptions in stochastic solutions

In this section, we evaluate the potential effect of uncertainty assumptions onto strategic decisions in the context of stochastic modeling. To do so, we consider a two-stage SP formulation of the energy model described above and generate scenarios assuming different probability distributions. Then, we analyze the resulting stochastic solutions and compare them with robust solutions reported in the literature.

3.1 The classical two-stage stochastic approach

Problem (1) under uncertainty can be formulated as a two-stage SP model. One chooses the first-stage investment decision variables $\mathbf{x}$ before the realization of uncertain parameters minimizing the associated investment cost plus the expected second-stage cost that depends on the recourse operation variables $\mathbf{y}$. The second-stage variables $\mathbf{y}$ adapt optimally to the revealed uncertainty. A standard formulation of the two-stage stochastic model is as follows:

$$\begin{aligned} \underset{\mathbf{x} \in X}{\text{minimize}} \quad & c^T \mathbf{x} + \mathbb{E}[Q(\mathbf{x}, \xi)] \\ \text{subject to} \quad & A\mathbf{x} \geq b, \end{aligned} \tag{2}$$

where $Q(\mathbf{x}, \xi)$ is the recourse function

$$Q(\mathbf{x}, \xi) := \min_y \{e^T \mathbf{y} : W\mathbf{y} \geq d - T\mathbf{x},\ \mathbf{y} \in Y\}$$

and $\xi := (e, T, W, d)$ indicates that the uncertainty can be present in any of the coefficients of the second-stage problem. In this formulation,we minimize the total expected value assuming nominal values for the first-stage uncertainties (e.g., investment costs) and a probability distribution function for the second-stage uncertain parameters ξ. The recourse function $Q(\mathbf{x}, \xi)$ depends on the first-stage decision $\mathbf{x}$ and the parameters ξ.

Problem (2) involves the expectation of $Q(\mathbf{x}, \xi)$ with respect to ξ. In general, such an expectation corresponds to a multi-dimensional integral and as such is virtually impossible to compute. Even when ξ has only a finite number of possible outcomes (also called *scenarios*), the number of scenarios may grow quickly with the number of uncertain parameters, so that the recourse function becomes intractable. For example, for m independent uncertain parameters, with three possible values each one, it gives a total of 3^m scenarios. To overcome this difficulty, the Sample Average Approximation (SAA) approach is used. Let $(\xi_i)_{i=1}^N$ be a set of N samples generated from the distribution of ξ. Then, the expected value of Q in (2) is approximated by the average of the realizations:

$$\mathbb{E}[Q(\mathbf{x}, \xi)] \approx \frac{1}{N} \sum_{i=1}^{N} Q(\mathbf{x}, \xi_i).$$

Note that the number N of samples yields a trade-off between accuracy and computational tractability needed to solve the problem. Discussions on related issues in the SAA approach can be found in Shapiro et al. [2014] and Homem-de-Mello and Bayraksan [2014]. We explain in Appendix B.1 the approach we implement to generate a reduced set of 1500 samples that yields an acceptable optimality gap and thus an acceptable approximation of the expected value.

3.2 Uncertainty assumptions: An empirical assessment

To illustrate the potential impact of uncertainty assumptions on strategic investment decisions, we perform a numerical experiment considering different PDFs for the uncertain parameters in (2). We compute two stochastic solutions and compare them with solutions reported in Moret et al. [2020a] which rely on the robust optimization paradigm which is discussed in the Introduction. The two stochastic solutions are defined as follows:

- *Stochastic-U*: The first stochastic solution is obtained by solving Problem (2) and assuming, as in

Moret et al. [2020a], uniform distributions for all uncertain second-stage parameters with variation ranges as reported in Table 1.

- *Stochastic-L*: The second stochastic solution is obtained by solving Problem (2) and assuming uniform distributions for uncertain second-stage parameters with symmetrical variation ranges and truncated lognormal distributions for uncertain second-stage parameters with asymmetrical variation ranges. Note that we choose the truncated lognormal distribution to satisfy the median property of the nominal value and ranges.

These stochastic solutions are compared with:

- The *Deterministic* one which does not consider uncertainty and assumes the nominal value for all parameters.
- The *Worst-case* solutions which assumes worst-case values for uncertain first- and second-stage parameters, as in Soyster [1973].
- The *Robust* solution computed in Moret et al. [2020a] and based on the robust optimization (RO) techniques [Bertsimas and Sim, 2004]. It adopts a min-max approach protecting against any realization of uncertain first- and second-stage parameters within the controlled uncertainty set. We refer the reader to Moret et al. [2020a] for more details.

3.2.1 Impact of distribution on installed capacity

Figure 1 shows the investment decisions for the electricity sector proposed in five solutions, i.e., *Deterministic*, *Robust*, *Worst-case*, *Stochastic-U* and *Stochastic-L*. For each solution, the left bar represents the installed capacities F and the right bar shows the available capacities Fcp for production, i.e., taking into account the yearly available factor cp of the technologies.

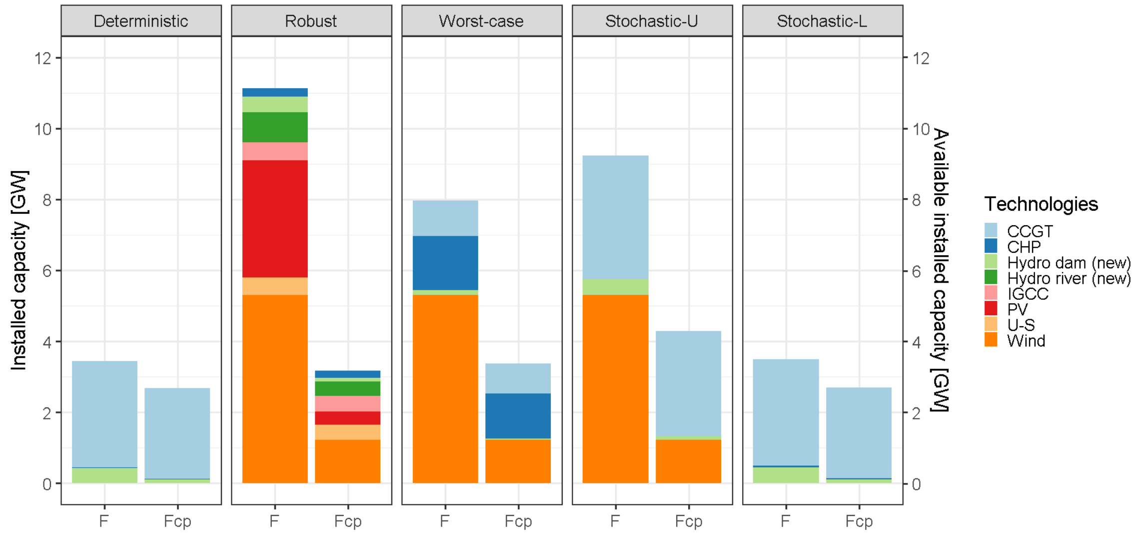

Figure 1: Electricity capacity mix (Full capacity F and available capacity Fcp) for the five investment strategies: *Deterministic*, *Robust*, *Worst-case*, *Stochastic-U* and *Stochastic-L*. (Acronyms: Photovoltaics (PV), Combined Cycle Gas Turbine (CCGT), Cogeneration of Heat and Power (CHP), Integrated Coal Gasification Combined Cycle (IGCC), Ultra-Supercritical Coal (U-S))

We observe the significant difference among the computed solutions depending on the chosen stochastic approach and/or the underlying probabilistic assumptions. On the one hand, the *Stochastic-U* solution invests only in renewable (Wind and Hydro dams) and fossil energy sources, while the *Stochastic-L* solution consists mostly of investments in natural gas (NG) similarly to the strategy of the *Deterministic* solution. This can be explained by the lognormal assumption which puts higher probability on low costs for gas imports, thereby making gas more competitive. On the other hand, *Robust* and *Worst-case* solutions are the only ones to invest significantly in PV and CHP capacities, respectively. These differences in strategic investments make the design of an efficient and robust energy policy highly hazardous for any decision maker.

3.2.2 "Out-of-Sample" simulation process

To assess and compare the economic performance of the five solutions in Figure 1, we perform an "Out-of-Sample" simulation process. We generate two sets of $n_{sample} = 10,000$ scenarios of first- and second-stage uncertain parameters assuming the probability settings used in the optimization process, i.e., in the first set, we assume uniform distributions for stochastic parameters whereas in the second set we use truncated lognormal distributions (for parameters with asymmetric ranges). For completeness, we perform an additional out-of-sample analysis using Triangular distributions centered on nominal values in order to assess the performance of the solutions on a different distribution setting. Then we solve the optimization problem for each parameter scenario with fixed investment decision variables, $\mathbf{x} = F$. In other words, installed capacity of the technologies is fixed (first-stage decisions) and the operation variables are determined by the second-stage optimization. Note that the electricity demand can always be satisfied by relying when needed on electricity imports. For the heating sector, we introduce a slack variable with a high penalty cost to ensure feasibility of the second-stage problem in each simulation run. As the paper focuses on the electricity sector, these infeasibility-related costs are not included in the reported computed cost results. infeasibilities are given in Table 3.

In Table 2, we report some cost statistics of the various strategies from simulations: the mean, the half-width of a 95% confidence interval for the mean and the standard deviation. Note that the maintenance cost component is included in the investment cost since it depends on the installed capacity.

		mean ± half-width/std		
Distributions		Investment cost	Operations cost	Total
Uniform	*Deterministic*	1406.4 ± 2.34/119.7	7137.4 ± 30.00/1530.5	8543.8 ± 30.12/1536.7
	Robust	3506.8 ± 5.80/296.3	5451.7 ± 16.48/841.2	8958.5 ± 17.34/885.0
	Worst-case	2254.7 ± 3.60/183.8	6389.5 ± 26.08/1330.8	8644.2 ± 26.57/1355.4
	Stochastic-U	2847.7 ± 4.85/247.8	5598.5 ± 20.94/1068.3	8446.2 ± 21.56/1100.0
	Stochastic-L	1446.7 ± 2.38/121.5	7159.4 ± 29.09/1484.1	8606.1 ± 29.21/1490.4
Lognormal	*Deterministic*	1405.2 ± 2.32/118.4	6048.3 ± 17.00/866.8	7453.5 ± 17.18/876.7
	Robust	3473.5 ± 5.73/292.3	4668.6 ± 8.87/452.0	8142.1 ± 10.56/539.0
	Worst-case	2241.5 ± 3.55/181.3	5388.2 ± 14.57/743.5	7629.7 ± 15.17/774.3
	Stochastic-U	2842.8 ± 4.88/249.0	4789.9 ± 11.82/603.1	7632.7 ± 12.81/654.0
	Stochastic-L	1445.6 ± 2.35/120.3	6092.8 ± 16.35/834.4	7538.4 ± 16.55/844.7
Triangular	*Deterministic*	1402.9 ± 1.63/83.2	6740.2 ± 21.86/1115.6	8143.1 ± 21.94/1119.3
	Robust	3479.3 ± 4.05/206.9	5184.3 ± 11.42/583.0	8663.6 ± 12.00/612.2
	Worst-case	2252.9 ± 2.47/126.4	6005.0 ± 18.74/956.2	8257.9 ± 19.11/975.3
	Stochastic-U	2836.4 ± 3.41/174.0	5326.9 ± 15.15/773.3	8163.3 ± 15.55/793.5
	Stochastic-L	1443.7 ± 1.65/65.6	6783.9 ± 21.10/1076.9	8227.6 ± 21.18/1080.6

Table 2: Simulation results for the various models for different out-of-sample distributions.

First, we observe in Table 2 that the estimates for the mean costs are are very precise in all cases since they have half-widths always smaller than 0.3%. The simulations with the lognormal distribution produce lower standard deviations of the output since that distribution corresponds to input parameters with lower variance than the other two distributions. As expected, the *Robust* solution yields a high average investment cost, but with the lowest standard deviation in operations costs as it protects the energy system against extreme second-stage operations costs. The *Worst-case* strategy should lead in theory to the most expensive investment solution to limit also operations costs but, given the worst-case investment costs, the system privileges energy sources with small uncertainty on investment costs (e.g., CHP and Wind). As expected, *Deterministic* and *Stochastic-L* solutions, with similar investments, have close performances with low average investment costs and high average yearly operating costs. However, the *Deterministic* solution leads more frequently to infeasibility in the second stage as discussed shortly.

We conclude from this simulation study that the best model in terms of average total cost depends on the choice of the out-of-sample distribution. For example, the cost performance of the *Stochastic-U* and *Stochastic-L* solutions depends on the assumed distribution in the simulations: *Stochastic-L* performs better with the lognormal distribution while *Stochastic-U* gives lower cost estimates assuming the uniform and triangular distributions. This is a clear illustration that the assumption on the distribution is very impacting and one can generate solutions that are suboptimal in practice and possibly undesirable. This motivates the use of Distributionally Robust Optimization (DRO) techniques to produce solutions that will remain good whatever the true probability for uncertain parameters is.

		Infeasibility		Elec. Imports
Distributions		% of simulations	Demand shortage	in GWh/h
Uniform	*Deterministic*	53.7%	0.84%	5.39
	Robust	0.03%	0.22%	10.20
	Worst-case	0%	0%	1.12
	Stochastic-U	0.19%	0.12%	5.73
	Stochastic-L	0.19%	0.14%	5.61
Lognormal	*Deterministic*	54.0%	0.86%	1.70
	Robust	0.03%	0.05%	10.28
	Worst-case	0%	0%	0.28
	Stochastic-U	0.19%	0.10%	1.80
	Stochastic-L	0.18%	0.12%	1.78
Triangular	*Deterministic*	55.2%	0.58%	3.07
	Robust	0%	0%	10.42
	Worst-case	0%	0%	0.56
	Stochastic-U	0%	0%	3.29
	Stochastic-L	0%	0%	3.18

Table 3: Simulation results in terms of infeasibility (% of simulations with unmet heating demand and percentage of conditional unmet demand) and imported electricity (in GW) for Uniform, Lognormal and Triangular distributions.

For the sake of completeness, we report in Table 3 additional simulation results, i.e, percentage of simulations with unsatisfied heating demand, the associated percentage of conditional unmet heating demand and the electricity imports that are needed to meet electricity demand. We can see that although the *Deterministic* solution appeared to produce solutions with low total cost, it fails to deal with demand variability within the heating sector. All other models yield acceptable feasibility performances. By construction, the two min-max solutions (*Robust* and *Worst-case*) are the ones with lowest infeasibility.

4 A distributionally robust optimization energy model

In this section, we introduce a distributionally robust optimization (DRO) approach to deal with the potentially high impact of distribution assumptions on stochastic solutions as observed in the previous section. As discussed in the Introduction, DRO is based on the design of a set of probability distributions (called *ambiguity set*) so that the model protects against the worst-case distribution within that set. In addition, we perform a Machine Learning (ML) analysis to identify a reduced subset of the uncertain parameters containing the most significant ones to be considered in the DRO approach. This ML-based selection process prevents the DRO model from having a numerically intractable large-scale formulation.

4.1 Distributionally robust optimization for two-stage models

The DRO formulation of the two stage model (2) can be written as follows:

$$\underset{\mathbf{x} \in X}{\text{minimize}} \quad c^T\mathbf{x} + \max_{\mathbb{P}\in\mathcal{D}} \mathbb{E}_{\mathbb{P}}[Q(\mathbf{x}, \xi)], \tag{3}$$

The objective function of DRO optimizes the worst-case expectation of the recourse function $Q(\mathbf{x}, \xi)$ over the ambiguity set $\mathcal{D}$ that includes all possible distributions $\mathbb{P}$ of the random vector variable ξ that have a certainty property, as discussed below. The set X is the feasibility region of the decision variable $\mathbf{x}$.

An important element in DRO is the design of the ambiguity set $\mathcal{D}$. There are multiple ways to define the ambiguity set, which must be appropriate for the application at hand [Gao and Kleywegt, 2016]. Moment-based ambiguity sets are utilized to model known structural properties such as symmetry [Roald et al., 2015], unimodality [Li et al., 2016], multimodality, independence patterns, among others, or moment constraints such as mean [Goh and Sim, 2010], variance, covariances, higher order moments, mean-absolute deviation, etc. Another ambiguity set is metric-based, which is constructed by using a function to measure the distance between two distributions in the probability space. Typically, this ambiguity set corresponds to a ball that is centered on a reference distribution and measures the distance between this reference distribution to the worst distribution within the ambiguity set. There are several ways to measure such distance; for instance, ϕ-divergence[2] [Ben-Tal et al., 2013], Wasserstein distance [Mohajerin Esfahani and Kuhn, 2018] and total variation distance [Rahimian et al., 2019]. A comprehensive review of DRO models and methods can be found in Rahimian and Mehrotra [2019]. It is also worthwhile mentioning that, via a dual representation, problem (3) can be written as a risk-averse version of (2) whereby the expectation is replaced by a coherent risk function; in that context, the size of the ambiguity set is directly related to the level of risk aversion—the larger the ambiguity set, the more risk-averse the model is. We refer to Shapiro et al. [2014] and references therein for details.

Solving model (3) exactly is in general very challenging. Using the Wasserstein distance, different reformulations of the model (3) have been proposed in the literature to obtain computationally tractable problems. Mohajerin Esfahani and Kuhn [2018] show that the two-stage DRO is reduced to a linear program if 1-norm or ∞-norm is used in the definition of the Wasserstein distance and the objective function belongs to a class of loss functions. Xu and Burer [2018] reformulate the maximum expected optimal value of uncertain mixed binary linear programming problem as a copositive program under standard assumptions, using a ambiguity set based on Wasserstein distance. They also show the effectiveness of their approach compared to the moment-based ambiguity set through numerical results. Hanasusanto and Kuhn [2018] consider a two-stage distributionally optimization with uncertainty in

[2]The ϕ-divergence is not actually a distance since it is not symmetric; however, it has the property that it is equal to zero if and only if the two distributions coincide.

the cost vector e and in the technology matrix T. They show that with a 2-norm Wassertein distance centered on a discrete reference distribution, any two-stage DRO problem with full resource is equivalent to a copositive program of polynomial size. Bansal et al. [2018] study a two-stage DRO problem using Wasserstein distance, where each probability distribution $\mathbb{P} \in \mathcal{D}$ has finite support. They propose decomposition algorithms (TSDR-LPs and TSDR-MBPs) that use a distribution separation procedure to solve, respectively, two-stage DRO linear programming and two-stage DRO mixed binary programming, under necessary conditions ensuring finite convergence.

In this paper, we use the algorithm TSDR-LPs described by Bansal et al. [2018] along with Benders decomposition to solve the strategic energy planning problem where the ambiguity set is defined by the Wasserstein distance, since it has following desirable characteristics: 1) Its formulation in LP allows for the use of existing solver and for the decomposition of the original problem; 2) The uncertainty in the second stage can be considered in any element of the model, that is, in the vectors e and d, also in the matrices W and T. Since we are considering finite support, this method is the most suitable for our DRO model. In the next section we give more details about our approach. We present the Wasserstein distance in the discrete setting and discuss some challenges of this metric.

4.1.1 Wasserstein-based ambiguity set

Let $\mathcal{M}_m(\Omega)$ be the set of all probability distributions $\mathbb{P}$ with support on $\Omega \subseteq \mathbb{R}^m$, (where m is the number of uncertain parameters) and which satisfy $\mathbb{E}_{\mathbb{P}}[\|\xi\|^p] < \infty$, with $p \geq 1$. The Wasserstein distance of order p between two distributions $\mathbb{P}_1$ and $\mathbb{P}_2 \in \mathcal{M}_m(\Omega)$ is defined as

$$W_p(\mathbb{P}_1, \mathbb{P}_2) := \left(\inf_{\Pi \in \Gamma_m(\mathbb{P}_1, \mathbb{P}_2)} \mathbb{E}_{\Pi}[\|\xi - \xi'\|^p] \right)^{1/p}, \tag{4}$$

where $\xi \sim \mathbb{P}_1$, $\xi' \sim \mathbb{P}_2$, and $\Gamma_m(\mathbb{P}_1, \mathbb{P}_2)$ represent the set of all distributions with support on $\Omega \times \Omega$ with marginals $\mathbb{P}_1$ and $\mathbb{P}_2$. The Wasserstein distance transports the probability mass from one distribution to another at a minimum cost. Indeed, the Wasserstein distance between two discrete distributions with a finite number of positive masses corresponds to a transportation planning problem, which can be formulated as a linear program.

The distance $W_p(\cdot, \cdot)$ is well-defined regardless of whether the distributions are continuous or discrete. We thus define the Wasserstein ambiguity set $\mathcal{D}_\epsilon$ as a ball of radius $\epsilon \geq 0$ with respect to the Wasserstein distance of order 1, centered at a prescribed reference distribution $\mathbb{P}_0$ as:

$$\mathcal{D}_\epsilon := \{\mathbb{P} \in \mathcal{M}_m(\Omega) : W_1(\mathbb{P}, \mathbb{P}_0) \leq \epsilon\}. \tag{5}$$

That is, the ambiguity set $\mathcal{D}_\epsilon$ contains all probability distributions whose Wasserstein distances to the reference distribution $\mathbb{P}_0$ are no more than ϵ. The radius ϵ explicitly controls the conservativeness of the resulting strategic decision; large ϵ will produce decisions that depend less on the assumed reference distribution, but in turn are more conservative. Note that the case of $\epsilon = 0$ corresponds to using the (non-DRO) expected value problem (2), whereas a value of $\epsilon = \infty$ (in practice, a large value of ϵ) corresponds to solving a robust version of (2) that minimizes the cost of the worst-case scenario instead of the expected cost. We can see then that the DRO formulation provides a continuum between those two extremes.

The ambiguity set (5) is one of the ambiguity sets proposed by Bansal et al. [2018] to solve two-stage DRO problems with finite-support distributions. Let L denote the number of points in the support Ω. The algorithm presented in that paper is a cutting-plane method that generates cuts for each $\mathbb{P} \in \mathcal{D}_\epsilon$. At each iteration, the following distribution separation problem is solved, for fixed $\mathbf{x} \in X$:

$$\max \left\{ \sum_{l=1}^{L} \mathbb{P}(\xi^l) Q(\mathbf{x}, \xi^l) : \mathbb{P} \in \mathcal{D}_\epsilon \right\}, \tag{6}$$

where the decision variables are the weights $\mathbb{P}(\xi^l)$ that the distribution $\mathbb{P}$ assigns to each outcome ξ^l for $l = 1, ..., L$. As discussed in Section 3, the number L of possible outcomes can grow exponentially with the number m of uncertain parameters of the model. Hence, it is impractical to have random vectors ξ even of moderate dimension, especially considering that the separation problem (6) is solved multiple times. To circumvent this problem, we propose to use machine learning techniques to select the most important parameters, as we will explain in the next section.

4.2 A Machine Learning approach for variable selection

To identify the most important parameters of the optimization model and thus reduce the computational time of the DRO algorithm, we rely on variable selection tools from machine learning. For this purpose, we use the Extreme Gradient Boosting (XGBoost) method, which is a predictive model based on a regression tree model [Friedman, 2001]. XGBoost is focused on computational speed and model performance, and can be used for supervised learning tasks such as Regression, Classification, and Ranking. In a nutshell, the XGBoost algorithm builds trees sequentially, where each new tree is created according to the margin of error left by the predictive variables of the previous tree, until the algorithm stabilizes and the performance of all trees combined reaches a maximum threshold of adjustment [Chen and Guestrin, 2016].

4.2.1 The XGBoost model

The idea of using XGBoost in our optimization model is to predict the installed capacity of different technologies of the electricity sector, which are summarized in eight target variables: Wind, Photovoltaics (PV), Combined Cycle Gas Turbine (CCGT), Combined Heat and Power (CHP), Integrated Coal Gasification Combined Cycle (IGCC), Ultra-Supercritical Coal (U-S), Hydro dam (new) and Hydro river (new).

The first step consists in generating a large sample of random parameter scenarios and in solving a large number of deterministic problems (1) (one per scenario) independently. This produces a dataset whose columns are the random values of the uncertain parameters and the values of the target output variables. Once the observations are obtained, the dataset is divided into two groups. The first one is the training sample, containing 70% of the data, on which the XGBoost algorithm is trained to obtain the impact of the predictors on target variables; then, the validation/prediction process is performed on the remaining data (30%), with the purpose of comparing real values with predicted ones and so to evaluate the precision of the ML models — one model per target variable.

To evaluate the quality of the XGBoost models, three indices of performance were used, including root-mean-squared error (RMSE), determination coefficient (R^2) and mean absolute error (MAE). Although these are standard measures of error, we include their expressions below for completeness:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \widetilde{y}_i)^2} \tag{7}$$

$$R^2 = 1 - \frac{\sum_i^n (y_i - \widetilde{y}_i)^2}{\sum_i^n (y_i - \bar{y}_i)^2} \tag{8}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \widetilde{y}_i| \tag{9}$$

where n is the number of instances, $\widetilde{y}_i$ is the predicted value of y_i, and $\bar{y}_i$ is the mean value of y_i.

4.2.2 Selection results

We performed the ML analysis with predictor variables corresponding to the second-stage uncertain parameters of (2) and 8 target variables as described above. To do so, we used the *xgboost* package by Chen et al. [2019]. In Table 4, we report the performance measures for each of the XGBoost models using the test samples (30% dataset).

	Target variables							
Indices	CHP	IGCC	U-S	Hydro dam (new)	Hydro river (new)	PV	Wind	CCGT
RMSE	0.308	0.587	0.356	0.038	0.087	0.639	0.473	0.433
R^2	0.779	0.702	0.944	0.945	0.926	0.838	0.969	0.879
MAE	0.205	0.410	0.184	0.019	0.034	0.260	0.225	0.306

Table 4: Performances of the XGBoost models on the testing dataset.

We observe in Table 4 that the R^2 values are close to 1 for most models, indicating good fits. In addition, the RMSE and MAE indices evaluate the errors between the observed and predicted values. Both have values close to zero, which means that the predictions are very close to those observed.

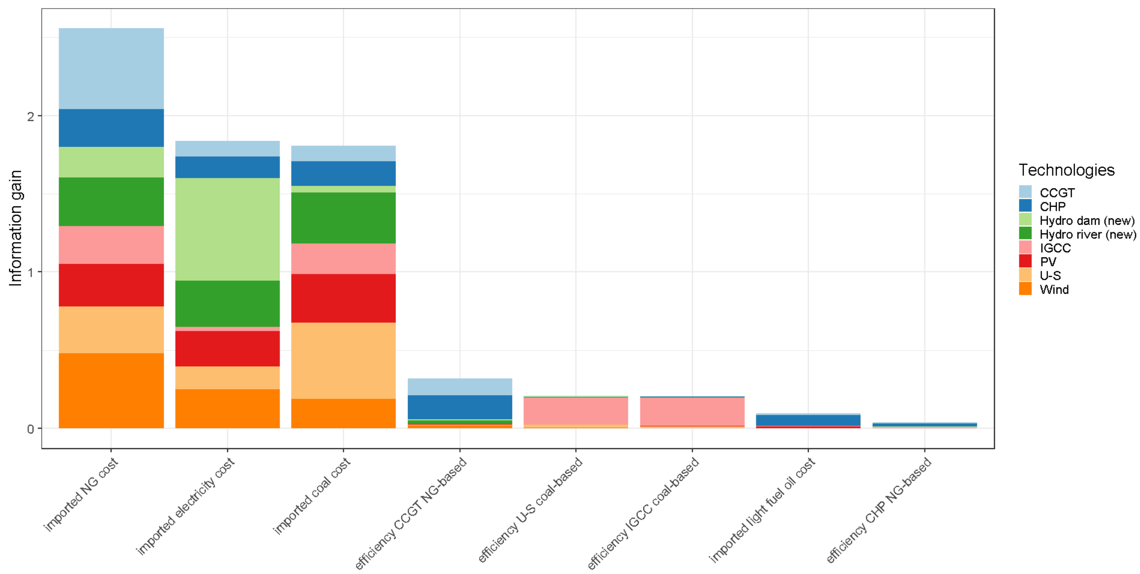

Figure 2: Information gains to the improvement of the XGBoost models stacked by parameters.

Figure 2 displays the results of the XGBoost analysis for the most important parameters in term of *information gains*, which are indices between 0 and 1 that indicate how well each uncertain parameter can be used to predict the target variable. Each bar in the figure displays the information gains on investments on each of the eight technologies corresponding to a given uncertain parameter—each color is a different technology. As we can see, the three most influencing parameters for the investment decisions are the importation costs of natural gas, electricity and coal, followed by three other parameters of smaller importance, i.e., for the efficiency of CCGT, U-S and IGCC.

In order to avoid over-fitting, ensuring these results are dataset independent, we carried out the same process 50 times, with different training and test sets, for all target variables. For each experiment, we obtain a ranking of the parameters in order of importance. The statistics of the rankings are summarized on Figure 3 for the eight most important parameters.

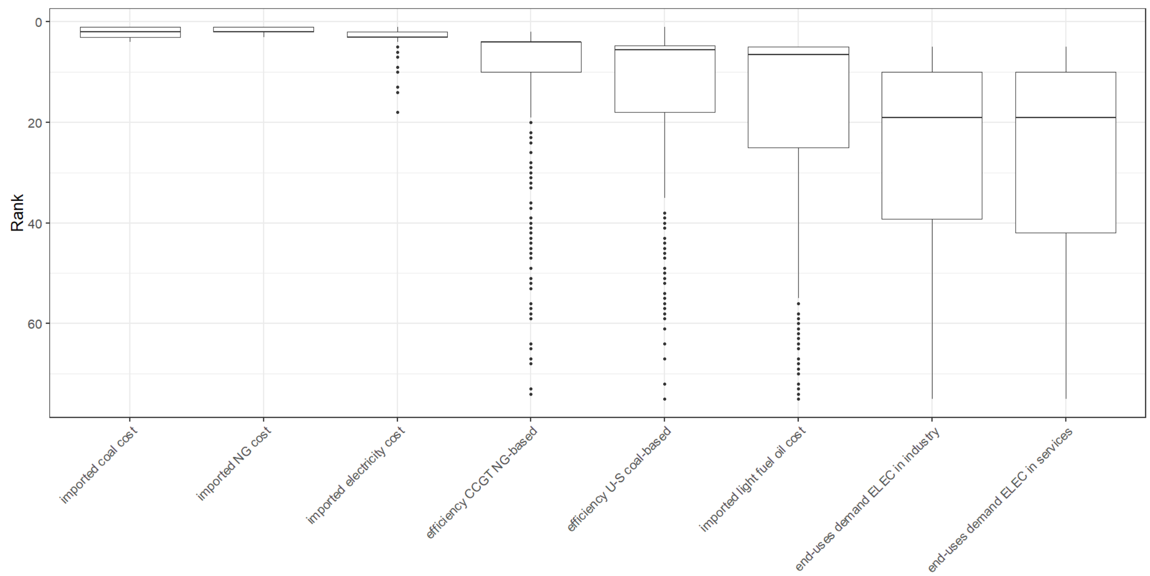

Figure 3: Ranking statistics for the most important parameters over 50 runs.

The results confirm that the most influencing parameters are the three import costs (i.e., gas, electricity and coal) followed by two efficiency parameters (i.e., CCGT and U-S). The efficiency of the IGCC technology is not anymore considered as an important uncertain parameter. We thus retain for our DRO model these five cost and efficiency parameters which appear both in Figures and 2 and 3.

5 Numerical experiments

In this section, we solve the two-stage DRO model considering the five most important uncertain parameters identified through the ML-based analysis and assuming different reference PDFs. Then we perform out-of-sample simulations to assess the performances of generated DRO solutions and compare them with stochastic and robust solutions.

5.1 Setting of DRO ambiguity sets

As discussed in Section 4.1, a key component of the DRO model (3) is the ambiguity set (5). When assuming a distance measure (e.g., the Wasserstein distance of norm 1 in our case), one has to define a reference distribution $\mathbb{P}_0$ and a support for the worst possible distribution $\mathbb{P}$ within the ambiguity set.

We recall that the objective of the DRO formulation is to produce investment decisions that are not sensitive to the assumed PDFs for the uncertain parameters as observed previously for the *Stochastic-U* and *Stochastic-L* solutions when using the standard stochastic approach. So, in order to demonstrate this desirable feature, we consider in our numerical experiments two ambiguity sets whose reference distributions have similar uncertain assumptions as for the *Stochastic-U* and *Stochastic-L* solutions.

More concretely, we define the first ambiguity set as

$$\mathcal{D}_{\epsilon}^{U} := \{\mathbb{P} \in \mathcal{M}_5(\Omega) : W_1(\mathbb{P}, \mathbb{P}_{\text{DRO-U}}) \leq \epsilon\} \tag{10}$$

where the reference distribution $\mathbb{P}_{\text{DRO-U}}$ corresponds to uniform distributions for the five uncertain second-stage parameters with variation ranges as reported in Table 1. Similarly, the second ambiguity set is given by

$$\mathcal{D}_{\epsilon}^{L} := \{\mathbb{P} \in \mathcal{M}_5(\Omega) : W_1(\mathbb{P}, \mathbb{P}_{\text{DRO-L}}) \leq \epsilon\}, \tag{11}$$

where the reference distribution $\mathbb{P}_{\text{DRO-L}}$ corresponds to uniform distributions for uncertain parameters with symmetrical variation ranges, i.e., the two efficiency parameters, and truncated lognormal distributions for uncertain parameters with asymmetrical variation ranges, i.e., the three cost parameters.

For the support of $\mathbb{P}$ in (10) and (11), we consider for each uncertain parameter a discrete support of three parameter values, i.e., the nominal one and its two extreme values as given in Table 1. Then we define Ω as the set of all combinations of the these values for all parameters, which results in $|\Omega| = 3^5 = 243$ possible outcomes. The set $\mathcal{M}_5(\Omega)$ is the set of all distributions with support on Ω.

In the following, we refer to DRO-U and DRO-L for DRO models with ambiguity sets (10) and (11), respectively. Each model is solved for different radius ϵ to compute DRO solutions with different levels of conservatism. We present and compare the most representative solutions, i.e, for ϵ_{min}, 0.084, 0.092, 0.108 and 1. The value $\epsilon = \epsilon_{min}$ refers to the minimum distance value for which the ambiguity set in (5) is non-empty in both models. For $\epsilon > 1$ the solutions do not change, which means that the corresponding solutions are obtained with the worst-case distributions among those with support Ω.

5.2 DRO strategic investment decisions

In this section, we present the DRO strategic investment decisions using the DRO-L and DRO-U models for different radius ϵ. The two-stage DRO algorithm was implemented in Julia 1.0.3, using the libraries of JuMP.jl and StructJuMP.jl. All solutions were obtained using a Intel Core i7-8750H CPU 2.20 GHz $\times$ 12 with 8 GB RAM.

We display in Figure 4 the DRO-L and DRO-U strategic investment decisions associated to the different radius ϵ together and the *Stochastic-L* and *Stochastic-U* solutions computed with the two-stage stochastic model.

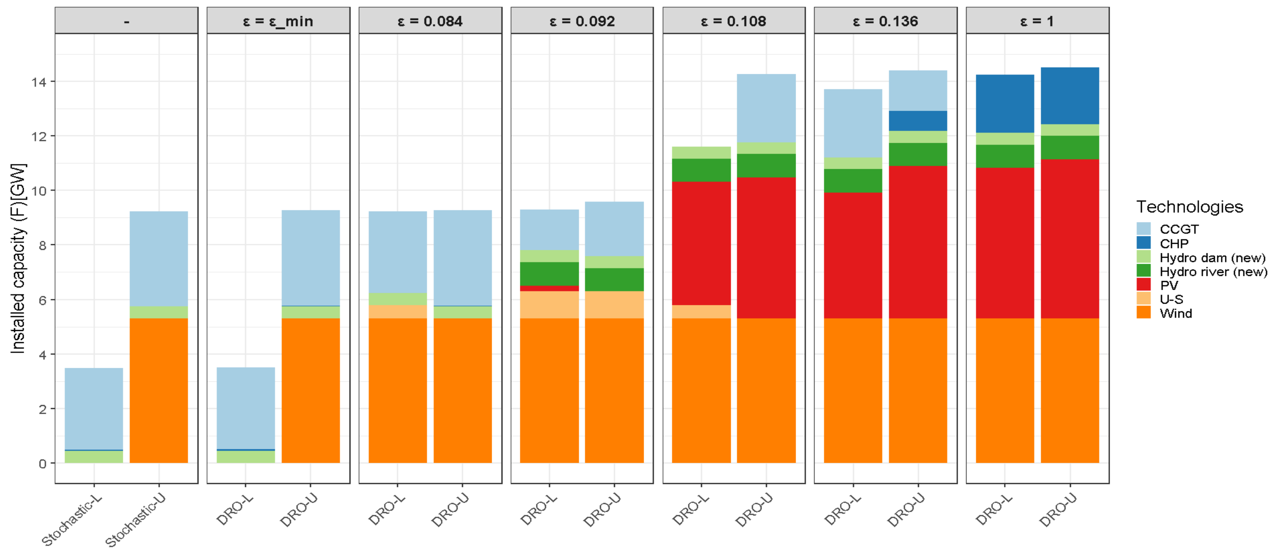

Figure 4: Installed capacity for *Stochastic-L*, *Stochastic-U*, DRO-L and DRO-U solutions.

By construction, the DRO-L and DRO-U investment solutions with ϵ_{min} are close to the *Stochastic-L* and *Stochastic-U* ones, respectively. Then, as expected, the DRO-L and DRO-U solutions are less dependent on the assumed reference distribution when ϵ increases. Except for $\epsilon = 0.108$, DRO-L and DRO-U solutions have very similar configurations. We also observe a diversification in the capacity mix as ϵ increases, which is a desirable property to reduce risk exposure—recall from the discussion in Section 4.1 that higher values of ϵ correspond to more risk-averse models. For increasing values of ϵ, we observe an increasing decarbonization of the electrical system with a mix of and efficient technologies. Indeed, the effect of high gas and coal costs makes the use of renewable and efficient technologies more competitive in a risk-averse environment.

5.3 Comparison of out-of-sample performances

To assess the economic performances of the DRO-L and DRO-U investment solutions of Figure 4, similarly to Section 3.2.2, we perform an "Out-of-Sample" simulation process assuming uniform, lognormal and triangular distributions. The simulation results are summarized in Figures 5, 6 and 7, respectively. For each solution, the figures display the boxplots for annual total cost, first-stage investment cost and second-stage operations cost.

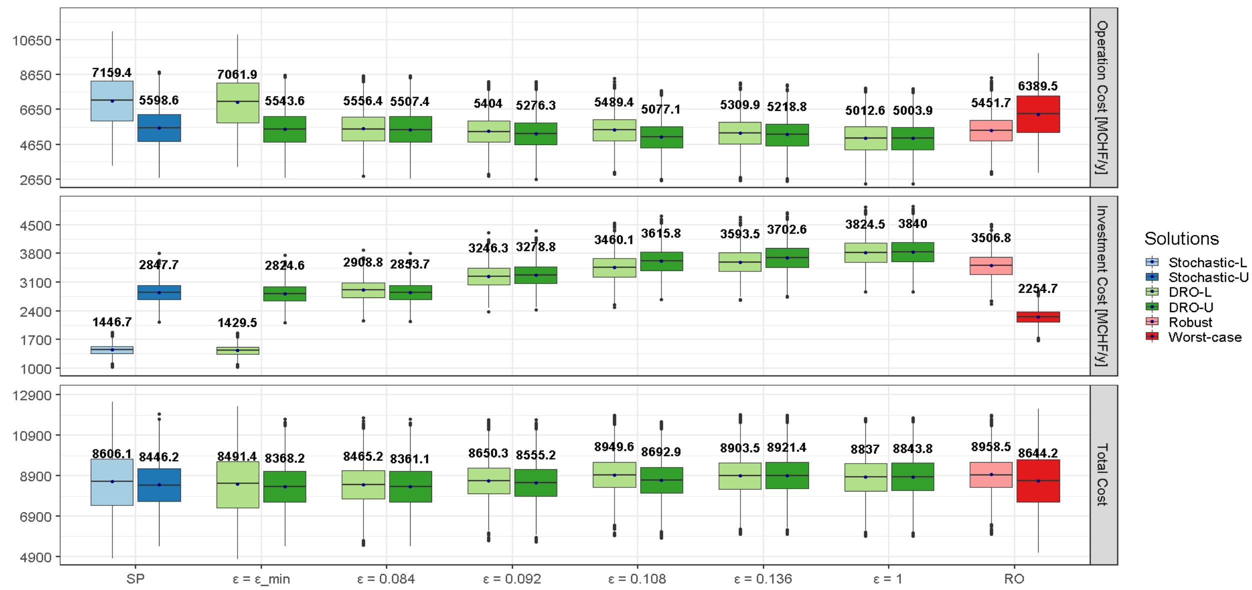

Figure 5: Boxplot of second-stage operations costs, first-stage investment costs and total cost. Simulations performed on $n_{sample} = 10.000$ scenarios generated with uniform distributions. The numbers indicate the average costs.

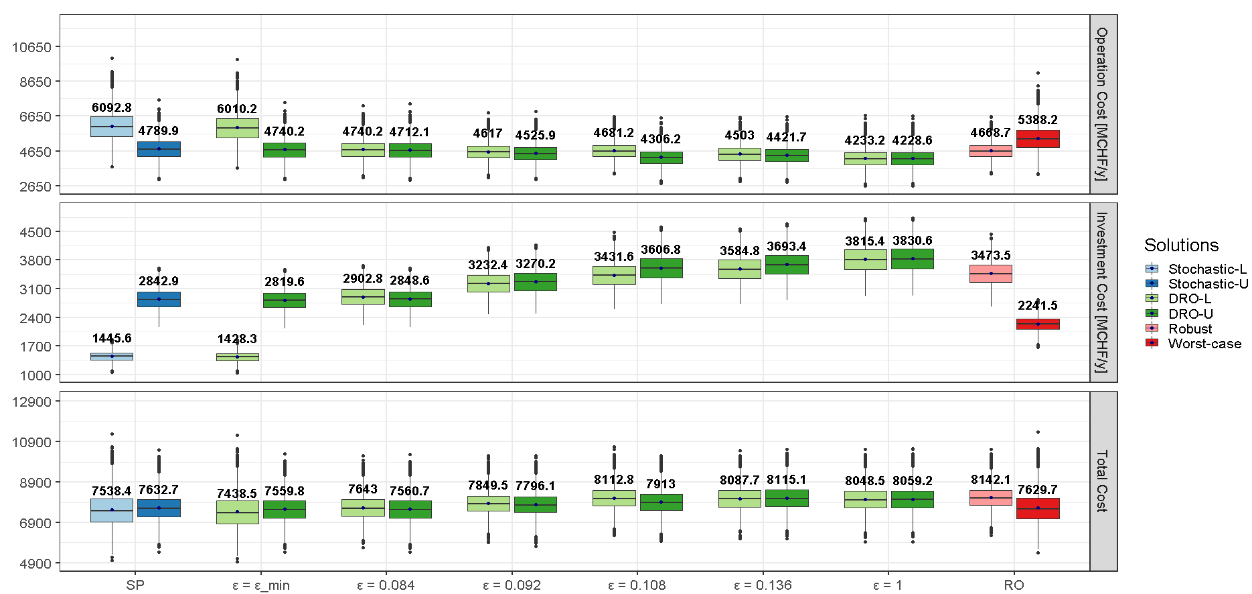

Figure 6: Boxplot of second-stage operations costs, first-stage investment costs and total cost. Simulations performed on $n_{sample} = 10.000$ scenarios generated with truncated lognormal distributions. The numbers indicate the average costs.

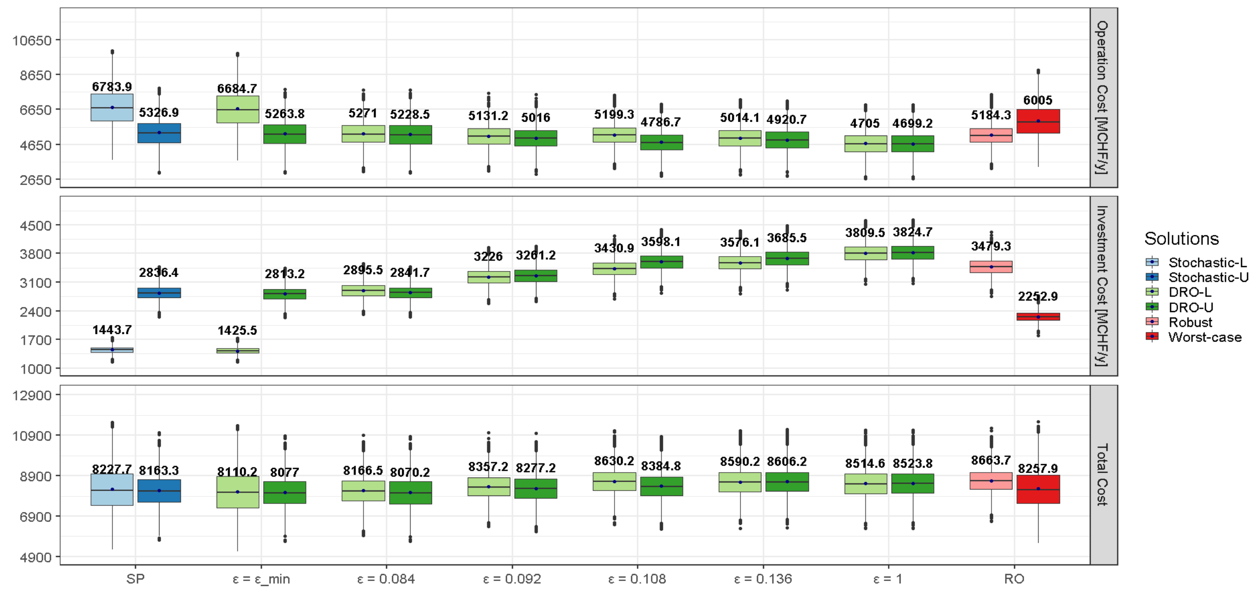

Figure 7: Boxplot of second-stage operations costs, first-stage investment costs and total cost. Simulations performed on $n_{sample} = 10.000$ scenarios generated with triangular distributions. The numbers indicate the average costs.

The main conclusion from Figures 5, 6 and 7 is that the performances of DRO solutions for a given ϵ radius are rather insensitive to the reference distribution in out-of-sample simulations. In comparison, the performances of *Stochastic-L* and *Stochastic-U* solutions are highly impacted by distribution assumptions in the three simulation processes in terms of average costs and cost dispersion. We also observe that more conservative DRO solutions are associated to higher first-stage investment costs but, at the same time, come with a small decrease of second-stage operations cost. Overall, The DRO solutions show a lower variation in the second-stage operations cost. This is of particular importance in real-world applications, in which investments are done at the beginning of the time horizon (*here-and-now* decisions); in this case, less exposure to significant variations in the second-stage operations implies more stability, and hence a lower risk of generating overcapacity in the power system, as recently showed by Moret et al. [2020b]. Total cost is quite constant among the DRO solutions.

The goal is to find a good solution that provides a balance among several desirable criteria: low total cost, low variability, independence from reference distribution and independence from the out-of-sample distribution. We see that that the standard stochastic and robust solutions fail at least on one of these criteria. On the other hand, the DRO solutions with the choice of with $\epsilon = 0.084$ seem to provide a good trade-off among those criteria, therefore it is our recommended strategy for this particular problem instance.

6 Conclusions and future work

Investment models for long-term energy planning provide important tools for strategic decision making, as they indicate which technologies are worth investing on, given the uncertainty in future costs and demand. While such models can be formulated as two-stage stochastic programs, the corresponding solutions are very sensitive to the choice of probability distributions for the uncertainty parameters in the problem, which is an enormous drawback considering that it is very difficult to assess the probability distributions of quantities far in the future.

In this paper, we have proposed a computationally tractable Distributionally Robust Optimization framework to deal with the high sensitivity of strategic investment solutions in energy planning to probability assumptions. The DRO formulation is based on the design of an ambiguous set of probability distributions (centered on a reference distribution) for a reduced number of important uncertain parameters. The selection of the important parameters—a key component of our approach, given the size of the model—is performed by solving single-scenario problems multiple times and applying machine-learning methods. Such an approach is, to the best of our knowledge, novel in the optimization and applied energy literature.

Our numerical results, obtained from experiments for a Swiss case study, show that the DRO investment strategies are quite stable regarding to variations in the underlying probability distributions, yielding in addition more diversified investments as we allow for larger ambiguity sets. As a consequence, the DRO model shows better performance in out-of-sample simulations than the standard stochastic programming and robust models.

Future research work includes extending the DRO formulation to a multi-stage model, since in most real-world energy system problems the uncertain parameters are revealed sequentially (more than two stages) and decisions must be adjusted to the uncertainty realizations. Another work direction is to develop methodologies that allow for incorporating more uncertain parameters in the ambiguity set but which are also computationally tractable and with low computational cost. Interpretation of stochastic optimization results by non-expert users is also a well-known challenge in the field [Grossmann et al., 2015]. To address this challenge, a decision-support method - similar to the "first feasibility, then optimality" approach proposed in Moret et al. [2020a] - could be developed to guide decision-makers in the choice of the most appropriate protection level ϵ, and hence the energy strategy.

Acknowledgements

First and third authors gratefully acknowledge the support provided by FONDECYT 1171145, Chile. The second author gratefully acknowledges partial support from Qatar National Research Fund under Grant Agreement no NPRP10-0212-170447 and from FONDECYT 1190325, Chile.

A Appendix

A.1 Mathematical model formulation

For interested readers, we report in this section the complete MILP model formulation as described in Moret et al. [2020a]. For the sake of simpler notations, we shorten the name of some variables.

In the following, we use the indicator function of a subset A of a set X as a function $\mathbf{1}_A : X \to \{0, 1\}$ defined as:

$$\mathbf{1}_A = \begin{cases} 1 & if \quad x \in A \\ \\ 0 & if \quad x \notin A \end{cases}$$

(I) Definition of sets.

T	:	Set of technologies	Sto	:	Set of storage units
R	:	Set of resources	EUC	:	Set of end-uses categories
P	:	Set of periods	S	:	Set of sectors
$BioFuels$	:	Set of biofuels import ($\subset R$)	L	:	Set of layers
$Export$	:	Set of exported resources ($\subset R$)	EUI	:	Set of end-uses Input
I	:	Set of infrastructure	EUT	:	Set of end-uses types
$T\text{-}EUT\{eut\}$	:	Set of technologies $\forall eut \in EUT$	$T\text{-}EUC\{euc\}$	:	Set of technologies $\forall euc \in EUC$

(II) Definition of variables

Name		Description	Stage
$\mathbf{G}_{\%Public}$	:	Ratio [0; 1] public mobility over total passenger mobility	I
$\mathbf{G}_{\%Rail}$	:	Ratio [0; 1] rail transport over total freight transport	I
$\mathbf{G}_{\%Dhn}$	:	Ratio [0; 1] centralized over total low-temperature heat	I
$\mathbf{F}_i$	:	Installed capacity with respect to main output i, $\forall i \in T$, [GW]	I
$\mathbf{Ft}_{i,t}$	:	Operation the i in each period t, $\forall i \in T \cup R$, $\forall\ t \in P$, [GW]	II
$\mathbf{Sto}^{+}_{j,l,t}$	:	Input to storage units $j \in Sto$ the $l \in L$ in period $t \in P$, [GW]	II
$\mathbf{Sto}^{-}_{j,l,t}$	:	Output from storage units $j \in Sto$ the $l \in L$ in period $t \in P$, [GW]	II
$\mathbf{Y}^{solar}_i$	:	If 1, technologies i is backup technology for decentralized solar else 0, $\forall i \in T$	I
$\mathbf{D}_{l,t}$	:	End-uses demand. Set to 0 if l $\notin$ EUT, $\forall l \in L$, $\forall t \in P$, [GW]	II
$\mathbf{N}_i$	:	Number integer of installed units i of size f^{ref}_i, $\forall i \in T$	I
$\mathbf{GWP}^{tot}$	:	Total yearly GHG emissions of the energy system, [$ktCO_2$-$eq./y$]	II
$\mathbf{GWP}^{constr}_i$	:	Technology construction GHG emissions, $\forall i \in T$, [$ktCO_2$-$eq.$]	I
$\mathbf{GWP}^{op}_r$	:	Total GHG emissions of resources, $\forall r \in R$, [$ktCO_2$-$eq./y$]	II
$\mathbf{Loss}_{eut,t}$	:	Losses in the networks (grid and DHN), $\forall eut \in EUT$, $\forall t \in P$, [GW]	II

(III) Definition of parameters

Name		Description	Unit
$eUYear_{eui,s}$	:	Annual end-uses in energy services per sector s, $\forall s \in S$,$\forall eui \in EUI$	[GWh/y]
		short name of $endUses_{year}$	
eUI_{eui}	:	Total annual end-uses in energy services eui, $\forall eui \in EUI$ $eUI_{eui} = \sum_{s \in S} eUYear_{eui,s}$ short name of endUsesInput	[GWh/y]
τ_i	:	Investment i cost annualization factor, $\forall i \in T$; $\tau_i = \frac{i_{rate}(i_{rate}+1)^{n_i}}{(i_{rate}+1)^{n_i}-1}$	
i_{rate}	:	Real discount rate	
$\underline{g}_k$, $\overline{g}_k$	:	Upper and lower limit to G_k, $\forall k \in \{\%Public, \%DHN, \%Rail\}$	
h_t	:	Time periods t duration, $\forall t \in P$	[Hour]
$\%lighting_t$	:	Yearly share (adding up to 1) of lighting end-uses, $\forall t \in P$	
$\%sh_t$	:	Yearly share (adding up to 1) of SH end-uses, $\forall t \in P$	
$f_{i,l}$	:	Input from (< 0) or output to (> 0) layers, $\forall i \in R \cup T \setminus Sto, \forall l \in L$	[GW]
f_i^{ref}	:	Reference size i with respect to main output, $\forall i \in T$	[GW]
c_i^{Inv}	:	Technology i specific investment cost, $\forall i \in T$	[MCHF/GW]
c_i^{Maint}	:	Technology i specific yearly O&M cost, $\forall i \in T$	[MCHF/GW/y]
gwp_i^{const}	:	Technology construction specific GHG emissions, $\forall i \in T$	[$ktCO_2$-*eq./GW*]
n_i	:	Technology i lifetime, $\forall i \in T$	[Year]
f_i^{min}, f_i^{max}	:	Min./max. installed size of the technology i, $\forall i \in T$	[GW]
$f_i^{min,\%}$, $f_i^{max,\%}$	:	Min./max. relative share of a technology in a layer i, $\forall i \in T$	
$avail_r$	:	Resource r yearly total availability, $\forall r \in R$	[GWh/y]
$k_{i,t}$	:	Period capacity factor of technology i in period t, $\forall i \in T, \forall t \in P$ (default 1)	
$\widehat{k}_i$	:	Yearly capacity i factor, $\forall i \in T$	
$c_{r,t}^{op}$	:	Specific cost of resources r in periods t, $\forall r \in R, t \in P$	[MCHF/GWh]
gwp_r^{op}	:	Specific GHG emissions of resources, $\forall r \in R$	[$ktCO_2$-*eq./GWh*]
$\eta_{j,l}^{+}$, $\eta_{j,l}^{-}$	:	Efficiency [0;1] of storage j input from/output to layer l. $\forall j \in Sto, \forall l \in L$	
$\%loss_{eut}$	:	Losses [0;1] in the networks (grid and DHN), $\forall eut \in EUT$	
$\%Peak_{DHN}$	:	Ratio peak/max. average DHN heat demand	
$m_{j,l,t}$	:	Auxiliary parameter $\forall i \in Sto, l \in L, t \in P$. $m_{j,l,t} = \max\left\{\frac{f_j^{max}}{\eta_{j,l}^{+} h_t}, \frac{\eta_{j,l}^{-} f_j^{max}}{h_t}\right\}$	

(IV) Model formulation

$$\min \quad \sum_{i\in T}(\tau_i \cdot c_i^{Inv} + c_i^{Maint}) \cdot \mathbf{F}_i + \sum_{r\in R}\sum_{t\in P} h_t \cdot c_{r,t}^{op} \cdot \mathbf{Ft}_{r,t} \tag{12.1}$$

s.t.

$$\mathbf{GWP}_i^{constr} = gwp_i^{constr} \cdot \mathbf{F}_i \qquad \forall i \in T, \tag{12.2}$$

$$f_i^{min} \leq \mathbf{F}_i \leq f_i^{max} \qquad \forall i \in T, \tag{12.3}$$

$$\mathbf{N}_i \cdot f_i^{ref} = \mathbf{F}_i \qquad \forall i \in T \setminus I, \tag{12.4}$$

$$\underline{g}_k \leq \mathbf{G}_k \leq \overline{g}_k \qquad \forall k \in \{\%Public, \%Rail, \%DHN\}, \tag{12.5}$$

$$\sum_{i\in T} \mathbf{Y}_i^{solar} \leq 1, \tag{12.6}$$

$$\mathbf{F}_{StoHydro} \leq f_{StoHydro}^{max} \frac{\mathbf{F}_{NewHydroDam} - f_{NewHydroDam}^{min}}{f_{NewHydroDam}^{max} - f_{NewHydroDam}^{min}} \quad , \tag{12.7}$$

$$\mathbf{F}_{DHN} \geq \sum_{i\in T\text{-}EUT\{HeatDHN\}} \mathbf{F}_i, \tag{12.8}$$

$$\mathbf{F}_{Grid} \geq 1 + \frac{9400}{c_{Grid}^{Inv}} \frac{\mathbf{F}_{Wind} + \mathbf{F}_{PV}}{f_{Wind}^{max} + f_{PV}^{max}}, \tag{12.9}$$

$$\mathbf{F}_{PowerToGas} = \max\{\mathbf{F}_{PowerToGas}; \mathbf{F}_{GasToPower}\}, \tag{12.10}$$

$$\mathbf{F}_{EFFICIENCY} = \frac{1}{1 + i_{rate}}, \tag{12.11}$$

$$\mathbf{F}_{NUCLEAR} = 0 \qquad , \tag{12.12}$$

$$\mathbf{Ft}_{i,t} \leq \mathbf{F}_i \cdot k_{i,t} \qquad \forall i \in T, \forall t \in P, \tag{12.13}$$

$$\sum_{t\in P} \mathbf{Ft}_{i,t} \cdot h_t \leq \mathbf{F}_i \cdot \widehat{k}_i \sum_{t\in P} h_t \qquad \forall i \in T, \tag{12.14}$$

$$\sum_{t\in P} \mathbf{Ft}_{r,t} \cdot h_t \leq avail_r \qquad \forall r \in R, \tag{12.15}$$

$$\sum_{i\in R\cup T\setminus Sto} f_{i,l}\mathbf{Ft}_{i,t} + \sum_{j\in Sto}(\mathbf{Sto}_{j,l,t}^{-} - \mathbf{Sto}_{j,l,t}^{+}) - \mathbf{D}_{l,t} - \mathbf{1}_A(l) \cdot \mathbf{Loss}_{l,t} = 0 \quad \forall l \in L, \forall t \in P, A = \{HeatDHN\}, \tag{12.16}$$

$$\mathbf{Ft}_{i,t} + \mathbf{Ft}_{Decsolar,t} \cdot \mathbf{Y}_i^{solar} \geq \frac{\mathbf{D}_{HeatDHN,t} + \mathbf{D}_{HeatDec,t}}{eUI_{heatSH} + eUI_{heatHW}} \sum_{t\in P} \mathbf{Ft}_{i,t} \cdot h_t \qquad \forall i \in T\text{-}EUT\{HeatDec\} \setminus \{Decsolar\}, t \in P, \tag{12.17}$$

$$\mathbf{Sto}_{j,l,t}^{+}(\lceil \eta_{j,l}^{+} \rceil - 1) = 0 \qquad \forall j \in Sto, \forall l \in L, \forall t \in P, \tag{12.18}$$

$$\mathbf{Sto}_{j,l,t}^{-}(\lceil \eta_{j,l}^{-} \rceil - 1) = 0 \qquad \forall j \in Sto, \forall l \in L, \forall t \in P, \tag{12.19}$$

$$\lceil \sum_{l\in L, \eta_{j,l}^{+}>0} \mathbf{Sto}_{j,l,t}^{+} \cdot m_{j,l,t} \rceil + \lceil \sum_{l\in L, \eta_{j,l}^{-}>0} \mathbf{Sto}_{j,l,t}^{-}/m_{j,l,t} \rceil \leq 1 \qquad \forall j \in Sto, \forall t \in P, \tag{12.20}$$

$$\mathbf{Loss}_{eut,t} = \sum_{i\in R\cup T\setminus Sto, f_{i,eut}>0} f_{i,eut} \cdot \mathbf{Ft}_{i,t} \cdot \%loss_{eut} \qquad \forall eut \in EUT, \forall t \in P, \tag{12.21}$$

$$\mathbf{GWP}_r^{op} = \sum_{t\in T} gwp_{r,t}^{op} \mathbf{Ft}_{r,t} \cdot h_t \qquad \forall r \in R, \tag{12.22}$$

$$\mathbf{GWP}^{tot} = \sum_{i\in T} \frac{\mathbf{GWP}_i^{constr}}{n_i} + \sum_{r\in R} \mathbf{GWP}_r^{op} \qquad , \tag{12.23}$$

$$\sum_{i\in T\text{-}EUT\{HeatDHN\}} \mathbf{F}_i \geq \%peak_{DHN} \max_{t\in P}\{\mathbf{D}_{HeatDHN,t}\} \qquad , \tag{12.24}$$

$$\sum_{i' \in T-EUT(eut)} f_i^{min,\%} \sum_{t \in T} \mathbf{Ft}_{i',t} h_t \leq \mathbf{Ft}_{i,t} h_t \qquad \forall eut \in EUT, \forall i \in \text{T-EUT}\{eut\}, \tag{12.25}$$

$$\sum_{i' \in T-EUT(eut)} f_i^{max,\%} \sum_{t \in T} \mathbf{Ft}_{i',t} h_t \geq \mathbf{Ft}_{i,t} h_t \qquad \forall eut \in EUT, \forall i \in \text{T-EUT}\{eut\}, \tag{12.26}$$

$$\mathbf{Ft}_{i,t} \sum_{t \in P} h_t \geq \sum_{t' \in P} \mathbf{Ft}_{i,t'} h_{t'} \qquad \forall t \in P, \forall i \in \text{T-EUC}\{MobPass\} \cup \text{T-EUC}\{MobFreight\}, \tag{12.27}$$

$$\mathbf{Sto}^+_{StoHydro,Elec,t} \leq \mathbf{Ft}_{HydroDam,t} + \mathbf{Ft}_{NewHydroDam,t} \qquad \forall t \in P, \tag{12.28}$$

$$\mathbf{Ft}_{j,t} = \mathbf{Ft}_{j,t-1} + h_t \sum_{\substack{l \in L \\ \eta^+_{j,l} > 0}} \mathbf{Sto}^+_{j,l,t} \eta^+_{j,l} - h_t \sum_{\substack{l \in L \\ \eta^-_{j,l} > 0}} \mathbf{Sto}^-_{j,l,t} / \eta^-_{j,l} \qquad \forall j \in Sto, \forall t \in P, \tag{12.29}$$

$$\mathbf{D}_{Elec,t} = \frac{eUI_{Elec}}{\sum_{t' \in P} h_{t'}} + eUI_{lighting} \frac{\%lighting_t}{h_t} + \mathbf{Loss}_{Elec,t} \qquad \forall t \in P, \tag{12.30}$$

$$\mathbf{D}_{q,t} = \left(\frac{eUI_{heatHW}}{\sum_{t' \in P} h_{t'}} + eUI_{heatSH} \frac{\%sh_t}{h_t} \right) (\mathbf{1}_B(q) + (-1)^{\mathbf{1}_B(q)} \mathbf{G}_{\%Dhn}) \qquad \forall t \in P, q \in \{HeatDHN, HeatDec\}, B = \{HeatDec\}, \tag{12.31}$$

$$\mathbf{D}_{q,t} = \frac{eUI_{passenger}}{\sum_{t' \in P} h_{t'}} (\mathbf{1}_{\{Priv\}}(q) + (-1)^{\mathbf{1}_{\{Priv\}}(q)} \mathbf{G}_{\%Public}) \qquad \forall t \in P, q \in \{Pub, Pri\}, \tag{12.32}$$

$$\mathbf{D}_{q,t} = \frac{eUI_{freight}}{\sum_{t_1 \in P} h_{t_1}} (\mathbf{1}_{\{Road\}}(q) + (-1)^{\mathbf{1}_{\{Road\}}(q)} \mathbf{G}_{\%Rail}) \qquad \forall t \in P, q \in \{Rail, Road\}, \tag{12.33}$$

$$\mathbf{D}_{HeatT,t} = \frac{eUI_{HeatT}}{\sum_{t' \in P} h_{t'}} \qquad \forall t \in P, \tag{12.34}$$

$$\mathbf{D}_{r,t} = 0 \qquad \forall t \in P, r \in R \setminus \{BioFuels \cup Export\}, \tag{12.35}$$

$$\mathbf{F}_i \geq 0 \qquad \forall i \in T, \tag{12.36}$$

$$\mathbf{Ft}_{i,t} \geq 0 \qquad \forall i \in T, \forall t \in P, \tag{12.37}$$

$$\mathbf{Sto}^+_{j,l,t}, \mathbf{Sto}^-_{j,l,t} \geq 0 \qquad \forall j \in STO, \forall l \in L, \forall t \in P, \tag{12.37}$$

$$\mathbf{Loss}_{eut,t} \geq 0 \qquad \forall eut \in EUT, \forall t \in P, \tag{12.38}$$

$$\mathbf{N}_i \in \mathbb{Z}^+ \qquad \forall i \in T, \tag{12.39}$$

$$\mathbf{Y}_i^{solar} \in \{0, 1\} \qquad \forall i \in T, \tag{12.40}$$

$$0 \leq \mathbf{G}_k \leq 1 \qquad \forall k \in \{\%Public, \%Rain, \%DHN\}, \tag{12.41}$$

$$\mathbf{GHP}^{tot} \geq 0, \tag{12.42}$$

$$\mathbf{GHP}_i^{constr} \geq 0 \qquad \forall i \in T, \tag{12.43}$$

$$\mathbf{GHP}_r^{op} \geq 0 \qquad \forall r \in R \tag{12.44}$$

(V) Definition of constraints.

Index		Description
(12.2)	:	The emissions related to the construction of each technology
(12.3)	:	Min & max limit to the size of each technology
(12.4)	:	Number of purchased technologies
(12.5)	:	% of passenger mobility, % of freight mobility and % of low-Temperature heat demand
(12.6)	:	Only one technology can be backup of solar
(12.13)	:	Relation between **Ft** and **F** via period capacity factor. This forces max monthly output
(12.14)	:	Relation between **Ft** and **F** via yearly capacity factor. This one forces total annual output
(12.15)	:	Resources availability equation
(12.16)	:	Layer balance equation with storage
(12.17)	:	Equation ensuring a constant relative use for decentralized heating technologies in each period, except for solar thermal (decsolar). This constraint is linearized as in [Moret et al., 2020a]
(12.18-12.19)	:	If the efficiency is 0 then the storage technology and the layer are incompatible
(12.20)	:	Displayed in a compact nonlinear formulation, ensures that the storage is not used as a transfer unit within a given period. This constraint is linearized as in [Moret et al., 2020a]
(12.21)	:	Calculation of losses for each end-use demand type
(12.22)	:	The emissions related to resources
(12.23)	:	Total yearly GHG emissions of the energy system, based on the technology lifetime, and the emissions related to resources
(12.29)	:	The level of the storage represents the amount of energy stored at a certain time
(12.30-12.35)	:	From annual energy demand to monthly power demand
(12.36- 12.44)	:	Nature of the variables
(12.7- 12.12) (12.24- 12.28)	:	This constraints are added to simplify the use of the model and adapt it to the specific case study of Switzerland, for a better understanding we refer to the reader see [Moret, 2017, p. 26]

B Appendix

B.1 Sample generation and optimality gap

Consider the stochastic programming problem

$$v^* = \min_{\mathbf{x}\in X} \left\{ g(\mathbf{x}) := c^T\mathbf{x} + \mathbb{E}[Q(\mathbf{x},\xi)] \right\} \tag{13}$$

where v^* is the optimal value of original problem and $g(\mathbf{x})$ is the expected value function at a given point $\mathbf{x}$ plus a constant. We will present briefly how to estimate the optimility gap using the estimates of v^* and $g(\mathbf{x})$.

In SAA, we select and fix $(\xi_i)_{i=1}^N$, all having the same distribution as ξ, and solve the following deterministic optimization problem:

$$\hat{v}_N = \min_{\mathbf{x}\in X} \left\{ g(\mathbf{x}) := c^T\mathbf{x} + \frac{1}{N}\sum_{i=1}^{N} Q(\mathbf{x}, \xi_i) \right\} \tag{14}$$

To reduce the computational effort in solving the problem (14), the ideal is to choose a small sample size N. We generate K independent random samples each of size N and solve the corresponding SAA problems (14). Let $\hat{v}_N^k$ and $\hat{x}_N^k$ be the corresponding optimal objective and optimal solutions, respectively, with $k = 1, ..., K$.

Then we can estimate v^* by

$$\bar{v}_N^K = \frac{1}{K}\sum_{k=1}^{K} \hat{v}_N^k \tag{15}$$

which represents a lower statistical bound of the original problem. Now consider a feasible solution $\hat{x} \in X$. For example, we can take $\hat{x}$ to be equal to an optimal solution $\hat{x}_N^k$ of an SAA problem. Let $g(\hat{x})$ be the true objective value the function g at the point $\hat{x}$. An unbiased estimator of $g(\hat{x})$ is given by:

$$\hat{g}_{N'}(\hat{x}) = c^T\hat{x} + \frac{1}{N'}\sum_{i=1}^{N'} Q(\hat{x}, \xi_i) \tag{16}$$

where $\xi_1, ..., \xi_{N'}$ are an independently and identically distributed random sample of $N^{'}$ realizations of random vector ξ. Since estimating the objective function $g(\hat{x})$ at a feasible point $\hat{x}$ by means of the average of $\hat{g}_{N'}(\hat{x})$ requires much less computational effort than solving the SAA problem, it makes sense to choose a very large sample size $N' \gg N$ in order to obtain an accurate estimate of the value objective $g(\hat{x})$ of an optimal solution $\hat{x}$ of the SAA problem. Consequently, since $\hat{x}$ is a feasible point of the true problem, $\hat{g}_{N'}(\hat{x})$ gives a statistical upper bound on the true optimal solution value. Using the above expressions, an estimate of the optimality gap $g(\hat{x}) - v^*$ of a candidate solution $\hat{x}$ is given by $\hat{g}_{N'}(\hat{x}) - \bar{v}_N^K$. This procedure is repeated, progressively increasing the values of K and N until a desired optimality gap is obtained. For more details on this method we suggest the reader to see [Homem-de-Mello and Bayraksan, 2014, Mak et al., 1999]. Finally, through numerical experiments of the method described above, we obtained a optimality gap of 0.3% for a sample size $N = 1500$.

References

F. Babonneau, A. Kanudia, M. Labriet, R. Loulou, and J.-P. Vial. Energy security: A robust optimization approach to design a robust european energy supply via tiam-world. *Environmental Modeling & Assessment*, 17(1):19–37, 2012. doi: 10.1007/s10666-011-9273-3.

F. Babonneau, M. Caramanis, and A. Haurie. ETEM-SG: Optimizing regional smart energy system with power distribution constraints and options. *Environmental Modeling & Assessment*, 22(5):411–430, 2017. doi: 10.1007/s10666-016-9544-0.

M. Bansal, K. Huang, and S. Mehrotra. Decomposition algorithms for two-stage distributionally robust mixed binary programs. *SIAM Journal on Optimization*, 28(3):2360–2383, 2018. doi: 10.1137/17M1115046.

A. Ben-Tal, L. El Ghaoui, and A. Nemirovski. *Robust optimization*, volume 28. Princeton University Press, 2009.

A. Ben-Tal, D. Den Hertog, A. De Waegenaere, B. Melenberg, and G. Rennen. Robust solutions of optimization problems affected by uncertain probabilities. *Management Science*, 59(2):341–357, 2013.

D. Bertsimas and M. Sim. The price of robustness. *Operations Research*, 52(1):35–53, 2004. doi: 10.1287/opre.1030.0065.

J. R. Birge and F. Louveaux. *Introduction to Stochastic Programming*. Springer Publishing Company, Incorporated, 2nd edition, 2011. ISBN 1461402360, 9781461402367.

T. Chen and C. Guestrin. Xgboost: A scalable tree boosting system. *CoRR*, abs/1603.02754, 2016. URL http://arxiv.org/abs/1603.02754.

T. Chen, T. He, M. Benesty, V. Khotilovich, Y. Tang, H. Cho, K. Chen, R. Mitchell, I. Cano, T. Zhou, M. Li, J. Xie, M. Lin, Y. Geng, and Y. Li. *xgboost: Extreme Gradient Boosting*, 2019. URL https://CRAN.R-project.org/package=xgboost. R package version 0.82.1.

Y. Chen, W. Wei, F. Liu, and S. Mei. Distributionally robust hydro-thermal-wind economic dispatch. *Applied Energy*, 173:511 – 519, 2016. ISSN 0306-2619. doi: https://doi.org/10.1016/j.apenergy.2016.04.060. URL http://www.sciencedirect.com/science/article/pii/S0306261916305153.

C. Duan, L. Jiang, W. Fang, and J. Liu. Data-driven affinely adjustable distributionally robust unit commitment. *IEEE Transactions on Power Systems*, 33(2):1385–1398, 2018. doi: 10.1109/TPWRS.2017.2741506.

J. H. Friedman. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5):1189–1232, 2001. ISSN 00905364. URL http://www.jstor.org/stable/2699986.

R. Gao and A. J. Kleywegt. Distributionally Robust Stochastic Optimization with Wasserstein Distance. *arXiv e-prints*, 2016.

A. L. Gibbs and F. E. Su. On choosing and bounding probability metrics. *International Statistical Review / Revue Internationale de Statistique*, 70(3):419–435, 2002. ISSN 03067734, 17515823. URL http://www.jstor.org/stable/1403865.

J. Goh and M. Sim. Distributionally robust optimization and its tractable approximations. *Operations Research*, 58(4-part-1):902–917, 2010. doi: 10.1287/opre.1090.0795.

I. Grossmann, R. M. Apap, B. Abreu Calfa, P. Garcia-Herreros, and Q. Zhang. Recent advances in mathematical programming techniques for the optimization of process systems under uncertainty. *Computer Aided Chemical Engineering*, 37:1–14, 2015.

Y. Guo, K. Baker, E. DallAnese, Z. Hu, and T. H. Summers. Data-based distributionally robust stochastic optimal power flowpart ii: Case studies. *IEEE Transactions on Power Systems*, 34(2):1493–1503, 2019. doi: 10.1109/TPWRS.2018.2878380.

X. Han and G. Hug. Distributionally robust generation expansion planning model considering res integrations. In *2019 IEEE Innovative Smart Grid Technologies - Asia (ISGT Asia)*, pages 1716 – 1721, Piscataway, NJ, 2019. IEEE. ISBN 978-1-7281-3520-5. doi: 10.3929/ethz-b-000348483. IEEE Innovative Smart Grid Technologies - Asia (ISGT Asia 2019); Conference Location: Chengdu, China; Conference Date: May 21-24, 2019; Conference lecture held on May 24, 2019.

G. A. Hanasusanto and D. Kuhn. Conic programming reformulations of two-stage distributionally robust linear programs over wasserstein balls. *Operations Research*, 66(3):849–869, 2018. doi: 10.1287/opre.2017.1698.

S. Hilpert, C. Kaldemeyer, U. Krien, S. Günther, C. Wingenbach, and G. Pleßmann. The open energy modelling framework (oemof) - A new approach to facilitate open science in energy system modelling. *CoRR*, abs/1808.08070, 2018. URL https://github.com/oemof/oemof.

T. Homem-de-Mello and G. Bayraksan. Monte carlo sampling-based methods for stochastic optimization. *Surveys in Operations Research and Management Science*, 19(1):56 – 85, 2014. ISSN 1876-7354. doi: https://doi.org/10.1016/j.sorms.2014.05.001. URL http://www.sciencedirect.com/science/article/pii/S1876735414000038.

M. Howells, H. Rogner, N. Strachan, C. Heaps, H. Huntington, S. Kypreos, A. Hughes, S. Silveira, J. DeCarolis, M. Bazillian, and A. Roehrl. Osemosys: The open source energy modeling system: An introduction to its ethos, structure and development. *Energy Policy*, 39(10):5850 – 5870, 2011. doi: https://doi.org/10.1016/j.enpol.2011.06.033. Sustainability of biofuels.

J. Krzemie. Application of markal model generator in optimizing energy systems. *Journal of Sustainable Mining*, 12(2):35 – 39, 2013. doi: https://doi.org/10.7424/jsm130205.

B. Li, R. Jiang, and J. L. Mathieu. Distributionally robust risk-constrained optimal power flow using moment and unimodality information. In *2016 IEEE 55th Conference on Decision and Control (CDC)*, pages 2425–2430, 2016. doi: 10.1109/CDC.2016.7798625.

G. Limpens, S. Moret, H. Jeanmart, and F. Maréchal. Energyscope td: A novel open-source model for regional energy systems. *Applied Energy*, 255:113729, 2019. ISSN 0306-2619. doi: https://doi.org/10.1016/j.apenergy.2019.113729. URL http://www.sciencedirect.com/science/article/pii/S0306261919314163.

W.-K. Mak, D. P. Morton, and R. Wood. Monte carlo bounding techniques for determining solution quality in stochastic programs. *Operations Research Letters*, 24(1):47 – 56, 1999. ISSN 0167-6377. doi: https://doi.org/10.1016/S0167-6377(98)00054-6. URL http://www.sciencedirect.com/science/article/pii/S0167637798000546.

P. Mohajerin Esfahani and D. Kuhn. Data-driven distributionally robust optimization using the wasserstein metric: performance guarantees and tractable reformulations. *Mathematical Programming*, 171(1):115–166, 2018. doi: 10.1007/s10107-017-1172-1.

S. Moret. *Strategic energy planning under uncertainty*. PhD thesis, Ecóle Polytechnique Fédérale de Lausanne, Switzerland, 2017.

S. Moret, M. Bierlaire, and F. Maréchal. Strategic energy planning under uncertainty: a mixed-integer linear programming modeling framework for large-scale energy systems. In Z. Kravanja and M. Bogataj, editors, *26th European Symposium on Computer Aided Process Engineering*, volume 38 of *Computer Aided Chemical Engineering*, pages 1899 – 1904. Elsevier, 2016. doi: https://doi.org/10.1016/B978-0-444-63428-3.50321-0.

S. Moret, V. C. Girons, M. Bierlaire, and F. Maréchal. Characterization of input uncertainties in strategic energy planning models. *Applied Energy*, 202:597 – 617, 2017. doi: https://doi.org/10.1016/j.apenergy.2017.05.106.

S. Moret, F. Babonneau, M. Bierlaire, and F. Maréchal. Decision support for strategic energy planning: A robust optimization framework. *European Journal of Operational Research*, 280(2):539–554, 2020a. ISSN 0377-2217. doi: https://doi.org/10.1016/j.ejor.2019.06.015.

S. Moret, F. Babonneau, M. Bierlaire, and F. Maréchal. Overcapacity in european power systems: analysis and robust optimization approach. *Applied Energy*, 2020b.

S. Pfenninger and B. Pickering. Calliope: a multi-scale energy systems modelling framework. *Journal of Open Source Software*, 3:825, 09 2018. doi: 10.21105/joss.00825. URL https://github.com/calliope-project/calliope.

W. B. Powell, A. George, H. Simo, W. Scott, A. Lamont, and J. Stewart. Smart: A stochastic multiscale model for the analysis of energy resources, technology, and policy. *INFORMS Journal on Computing*, 24(4):665–682, 2012. doi: 10.1287/ijoc.1110.0470.

D. Pozo, A. Street, and A. Velloso. An ambiguity-averse model for planning the transmission grid under uncertainty on renewable distributed generation. In *2018 Power Systems Computation Conference (PSCC)*, pages 1–7, 2018. doi: 10.23919/PSCC.2018.8442871.

H. Rahimian and S. Mehrotra. Distributionally Robust Optimization: A Review. *arXiv e-prints*, art. arXiv:1908.05659, 2019.

H. Rahimian, G. Bayraksan, and T. Homem-de-Mello. Identifying effective scenarios in distributionally robust stochastic programs with total variation distance. *Mathematical Programming*, 173(1): 393–430, 2019. ISSN 1436-4646. doi: 10.1007/s10107-017-1224-6. URL https://doi.org/10.1007/s10107-017-1224-6.

L. Roald, F. Oldewurtel, B. Van Parys, and G. Andersson. Security Constrained Optimal Power Flow with Distributionally Robust Chance Constraints. *arXiv e-prints*, art. arXiv:1508.06061, 2015.

A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on stochastic programming : modeling and theory*. SIAM, 2nd edition, 2014.

A. L. Soyster. Technical noteconvex programming with set-inclusive constraints and applications to inexact linear programming. *Operations Research*, 21(5):1154–1157, 1973. doi: 10.1287/opre.21.5.1154.

P. Sullivan, V. Krey, and K. Riahi. Impacts of considering electric sector variability and reliability in the message model. *Energy Strategy Reviews*, 1(3):157 – 163, 2013. doi: https://doi.org/10.1016/j.esr.2013.01.001. Future Energy Systems and Market Integration of Wind Power.

A. Velloso, D. Pozo, and A. Street. Distributionally Robust Transmission Expansion Planning: a Multi-scale Uncertainty Approach. *arXiv e-prints*, art. arXiv:1810.05212, Oct 2018.

W. Wiesemann, D. Kuhn, and M. Sim. Distributionally robust convex optimization. *Operations Research*, 62(6):1358–1376, 2014. doi: 10.1287/opre.2014.1314.

P. Xiong and C. Singh. Distributionally robust optimization for energy and reserve toward a low-carbon electricity market. *Electric Power Systems Research*, 149:137 – 145, 2017. doi: https://doi.org/10.1016/j.epsr.2017.04.008.

P. Xiong, P. Jirutitijaroen, and C. Singh. A distributionally robust optimization model for unit commitment considering uncertain wind power generation. *IEEE Transactions on Power Systems*, 32(1):39–49, 2017. doi: 10.1109/TPWRS.2016.2544795.

G. Xu and S. Burer. A data-driven distributionally robust bound on the expected optimal value of uncertain mixed 0-1 linear programming. *Computational Management Science*, 15(1):111–134, 2018. ISSN 1619-6988. doi: 10.1007/s10287-018-0298-9.