



UNIVERSIDAD ADOLFO IBÁÑEZ

FACULTAD DE INGENIERÍA Y CIENCIAS
DOCTORADO EN INGENIERÍA INDUSTRIAL E INVESTIGACIÓN DE
OPERACIONES

**STRATEGIC PLANNING MODELS FOR ENERGY SYSTEMS UNDER
UNCERTINTY**

A dissertation submitted in partial fulfilment of the requirements for the degree of PhD in
Industrial Engineering and Operations Research

PH.D. STUDENT: ESNIL JOSUE GUEVARA MENDOZA

Advisor: Tito Homem-de-Mello.
Co-Advisor: Frédéric Babonneau.

Santiago, Chile.
March, 2022

This thesis is dedicated to:

To God who has kept me and has given me strength to overcome discouragement and also creativity in the moments when I was not able to move forward.

To my wife Diana Pino who has been a special and loving partner, who has helped me find a balance in my life. Also, God used her to bring me back to Chile to finish my doctorate.

To my parents Pedro and Vilma who with their love and effort taught me to walk with honesty and in the fear of God.

To my brothers Jesus and Ruth for their unconditional love and support, they were with me all this time. To all my family, my mother-in-law and brothers and sisters in faith who supported me with their prayers.

ACKNOWLEDGMENTS

I want to express my gratitude to God, since He has given me so much in these years, I have never lacked anything since I left home and my country together with my wife.

To the Universidad Adolfo Ibáñez, to the entire Faculty of Engineering and Sciences, both the authorities and the staff, and to all my professors, who with their teachings and valuable knowledge made me continue to grow in my professional life.

To all my friends, to those who helped me financially, to Osmelia Tochón and Mrs. Luzbanny for having welcomed me in their homes when I arrived in Chile, also thanks to my classmates: Orlando Rivera, Sebastián Arpón, Hernán Lespay, Gianpiero Canessa, who helped me to integrate into the university.

Finally, I would like to express my greatest and most sincere gratitude to Professors Tito Homem-de-Mello and Frédéric Babonneau, my main tutors, whose guidance, knowledge, teaching and collaboration allowed the development of this work. I would especially like to thank Professor Tito Homem-de-Mello, because in addition to being a great mentor, he was an instrument that God used to give me the best advice when I needed it most.

Contents

ACKNOWLEDGMENTS.	ii
List of Figures	iv
List of Tables	vi
ABSTRACT	viii
RESUMEN	ix
Introduction	x
0.1 Problem Statement and Motivation	x
0.2 Outline of this Dissertation	xii
1 A machine learning and distributionally robust optimization framework for strategic energy planning under uncertainty	1
1.1 Introduction	1
1.2 Methodologies	3
1.2.1 Strategic energy model	3
1.2.2 The classical two-stage stochastic approach	4
1.2.3 A Machine Learning and Distributionally Robust Optimization Framework	5
1.3 Results and Discussion	10
1.3.1 Model uncertainty	10
1.3.2 Assessing the impact of distribution assumptions in stochastic solutions	11
1.3.3 Numerical experiments with DRO	15
1.4 Conclusions and future work	22
2 Modeling uncertainty processes for multi-stage optimization of strategic energy planning : An auto-regressive and Markov chain formulation	24
2.1 Introduction	24
2.1.1 Modeling uncertainty	25
2.1.2 Solutions methods for multi-stage stochastic programming	26
2.2 Multi-stage strategic energy planning	28
2.2.1 A deterministic multi-stage model	28
2.2.2 A stochastic multi-stage model	30
2.3 Uncertainty modelling through auto-regressive processes	31
2.3.1 The goals of the model	31
2.3.2 The construction of the AR model	33
2.4 SDDP and Markov Chain	36
2.4.1 Finite state Markov-chain approximations	36

2.5 Application to the Swiss energy System	39
2.5.1 Implementation details	39
2.5.2 Numerical results	40
2.6 Conclusions and Future Work	45
3 General Conclusion	47
Bibliography.	50
A Two-stage stochastic programming	58
A.1 Mathematical model formulation	58
B Sample generation and optimality gap	63
B.1 The sample average approximation method	63
C Multi-stage stochastic programming	65
C.1 Mathematical model formulation	65
C.2 Objective function	69
C.3 Constraints	69
C.3.1 Short-term operating equations	70
C.4 Additional constraints	72
C.4.1 Long-term investment equations	72
C.4.2 Short-term operating equations	72
D Assumptions for uncertain parameters	74
D.1 Uncertain parameters	74
E Modeling the price of resources and demands	77
E.1 First-order autoregressive process	77
F Markov chains for electricity and natural gas prices	79
F.1 Transition probability matrix of a three-state Markov chain	79

List of Figures

1.1	Electricity capacity mix (Full capacity F and available capacity F_{cp}) for the five investment strategies: <i>Deterministic</i> , <i>Robust</i> , <i>Worst-case</i> , <i>Stochastic-U</i> and <i>Stochastic-L</i> . (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))	12
1.2	Information gains to the improvement of the XGBoost models stacked by parameters.	16
1.3	Boxplot visualizes the ranking statistics for the most important parameters over 50 runs.	17
1.4	The effect of radius ϵ on the installed capacity of DRO-L and DRO-U solutions compared to the installed capacity of stochastic solutions (<i>Stochastic-L</i> and <i>Stochastic-U</i>).	19
1.5	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.	20
1.6	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.	20
1.7	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.	21
2.1	A small example of four stages with three periods of investment, followed by operating decisions conditioned on revealed uncertainty, which shows the dynamic process of the model.	31
2.2	In this figure $\mu_1, \mu_2, \dots, \mu_T$ are the nominal values, R_t^- and R_t^+ for $t = 1, \dots, T$ are the lower and upper bounds of the ξ_t parameter, respectively. ξ_0 is deterministic and represents the initial realization of the process. The figure on the left represents the evolution of the processes in the deterministic case, while the figure on the right, a possible realization of the stochastic process is shown.	33
2.3	The figure represents the possible relations of $\{\xi_t\}$ in red lines. The first stage covers 100% of the range, while the following stages always cover 75% of the range. It also shows a realization of process.	34
2.4	10,000 realizations of $\{\xi_t\}$, $t = 1, 2, 3$ for two different values of α .	35
2.5	An example with a 5-state Markov chain located at the base of the uniform distribution. The uniform distribution is represented with different colors. The colors represent the intervals formed with the selected step length.	38
2.6	Boxplot of total investment, maintenance, operation and total costs. The red dot represents the sample mean. For each out-of-sample simulation, 30000 scenarios were generated from the original continuous AR process with $\alpha \in \{0.75, 0.50, 0.25\}$.	41
2.7	Installed capacities for the SDDP-IND and SDDP- α policies for the maximum (MAX), minimum (MIN), nominal (Nominal) scenarios.	43
2.8	Power generation for the SDDP-IND and SDDP- α policies for the maximum (MAX), minimum (MIN), nominal (Nominal) scenarios.	44

2.9	Installed capacity for the heating sector for the SDDP-IND and SDDP- α policies for the maximum (MAX), minimum (MIN), nominal (Nominal) scenarios.	45
C.1	Diagram of sets and indices used in the MSLP model (source: Moret et al. [2017]). Abbreviations: space heating (SH), hot water (HW), temperature (Temp.), mobility (MOB), Low Temperature (LT), High Temperature (HT), Freight (FR.), Passenger (PASS.).	65
E.1	30,000 scenarios generated from the original AR processes with $\alpha \in 0.75, 0.50, 0.25$ for demand of ELECTRICITY, HEAT_HIGH_T. and LIGHTING.	77
E.2	30,000 scenarios generated from the original AR processes with $\alpha \in 0.75, 0.50, 0.25$ for price of COAL, ELECTRICITY and NG.	78
F.1	Markov Chains for electricity prices for different values of α	79
F.2	Markov Chains for Natural Gas prices for different values of α	80

List of Tables

1.1	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 (1.1) through the elements c, e, d, T and W.	11
1.2	Comparison of problem sizes and solution time for deterministic, robust and stochastic solutions.	13
1.3	Comparing the mean, half-width and standard deviation of the investment, operation and total costs obtained from stochastic, robust and deterministic models through different out-of-sample distributions.	14
1.4	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.	15
1.5	Performances of the XGBoost models on the testing dataset through several statistical indices.	16
1.6	DRO problem sizes and comparison of solution statistics for DRO-L and DRO-U models with different radios.	18
2.1	Characteristics and detailed numerical results for the SDDP-IND and SDDP- α models.	40
2.2	Costs associated to the SDDP-IND and SDDP- α policies for the maximum (MAX), minimum (MIN), nominal (Nominal) scenarios.	42
A.1	Variables for the first-stage problem	58
A.2	Variables for the second-stage problem	59
1	Nominal, min and max values for end-uses demands in heating, electricity and mobility sectors. Abbreviations: Temperature (T), Hot Water (HW), Space Heating (SH). Parameter $eUI_{year,EUIdemands} [GWh]$	74
2	Nominal, min and max values for resource costs. Abbreviations: Natural Gas (NG), Synthetic Natural Gas (SNG), Carbon Capture and Storage (CCS). Parameter $C_{year,resources,}^{Op} [MCHF/GWh]$	75

Strategic Planning Models for Energy Systems Under Uncertainty

Ph.D. student: Esnil Josue Guevara Mendoza

ABSTRACT

Long-term energy planning is the process by which the objectives and investment strategies (national or regional) are derived from quantitative analysis of energy sector scenarios, often supported by optimization-based energy models. In current energy planning practices, most of these energy models are “deterministic”, i.e. it does not consider uncertainty. Uncertainty in an optimization problem refers to the fact that some or all of the parameters of the problem are unknown at the time of solving it. Stochastic programming (SP) and robust optimization (RO) have often been used to solve energy planning optimization problems under uncertainty; however, they have usually been limited to considering only a handful of the uncertain input data. Based on an energy model of a real Swiss energy system reported in the literature, whose modeling framework allows considering hundreds of uncertain parameters, in this thesis we focus on two aspects; in the first objective, we study how the choice of stochastic approaches and distributional assumptions influence strategic investment decisions in energy planning problems. We present a two-stage stochastic programming model and assuming different probability distributions for the parameters of the second stage. Our first results showed that there is a significant discrepancy between the stochastic solutions associated with the assumed distributions. To remedy this sensitivity problem, we apply a combined machine learning (ML) and distributionally robust optimization (DRO) approach that produces more robust and stable investment decisions concerning distribution assumptions. The ML is used to restrict DRO to a set of most important parameters and ensure computational tractability of the resulting model. The results show that DRO investment strategies are more stable concerning variations in the underlying probability distributions. The second objective addresses the modeling of stochastic processes in multi-stage energy planning problems when little information is available on the degree of uncertainty of these processes. In the literature, works related to energy planning, in which several investment periods have been considered, most assume that the processes are independent between stages (or stage-wise independent, i.e., the information revealed at stage t of a random variable does not depend on past information). However, the scenarios generated by assuming stage-wise independent also involve the appearance of zigzag scenarios that do not occur in the real world. We start by making simple estimates of the ranges of variation of uncertain parameters for each stage, such as demands and resource costs, and the temporal correlation of these parameters is modeled by autoregressive (AR) models using as input data the ranges of variation and a parameter controlling the degree of temporal correlation. The AR model resulting from our method has the property that the random variables corresponding to each period are uniformly distributed within those pre-specified ranges, and also allows balancing the occurrence of zigzag and extreme scenarios using an α control parameter. We show that our ARs are suitable to be represented in the stochastic dual dynamic programming (SDDP) algorithm, since the errors meet the assumption of stage-wise independent. In addition, we also analyze the advantages of considering stage-wise dependent in multistage stochastic programming problems. For this purpose, we approximate the continuous AR(1) processes associated with the resource costs to a finite-state Markov chain. This is to preserve the convexity of the problem without losing the stage-wise dependent on the resource costs. We show that out-of-sample costs perform better when considering time dependence in the SDDP, especially in extreme scenario realizations. Moreover, the case where uncertainty is assumed to be independent in all periods does not allow for adaptation.

Strategic Planning Models for Energy Systems Under Uncertainty

Ph.D. student: Esnil Josue Guevara Mendoza

RESUMEN

La planificación energética a largo plazo es el proceso por el cual los objetivos y las estrategias de inversión (nacionales o regionales) se derivan del análisis cuantitativo de los escenarios del sector energético, a menudo con el apoyo de modelos energéticos basados en la optimización. En las prácticas actuales de planificación energética, la mayoría de estos modelos energéticos son “deterministas”, es decir, no tienen en cuenta la incertidumbre. La incertidumbre en un problema de optimización significa que algunos o todos los parámetros del problema no se conocen en el momento en que hay que resolverlo. La programación estocástica (SP) y la optimización robusta (RO) se han utilizado a menudo para resolver problemas de optimización bajo incertidumbre; sin embargo, normalmente se han limitado a considerar sólo un puñado de datos de entrada inciertos. Basándonos en un modelo energético de un sistema energético real suizo reportado en la literatura, cuyo marco de modelamiento permite considerar cientos de parámetros inciertos, en esta tesis nos centramos en dos aspectos; en el primer objetivo, estudiamos cómo la elección de los enfoques estocásticos y los supuestos de distribución influyen en las decisiones estratégicas de inversión en los problemas de planificación energética. Presentamos un modelo de programación estocástica en dos etapas y asumimos diferentes distribuciones de probabilidad para los parámetros de la segunda etapa. Nuestros primeros resultados mostraron que existe una discrepancia significativa entre las soluciones estocásticas asociadas a las distribuciones asumidas. Para remediar este problema de sensibilidad, aplicamos un enfoque combinado de aprendizaje automático (ML) y de optimización robusta en distribución (DRO) que produce decisiones de inversión más robustas y estables en relación con los supuestos de distribución. La ML se utiliza para restringir el DRO a un conjunto de parámetros más importantes y garantizar la trazabilidad computacional del modelo resultante. Los resultados muestran que las estrategias de inversión DRO son más estables respecto a las variaciones de las distribuciones de probabilidad subyacentes. El segundo objetivo aborda la modelización de procesos estocásticos en problemas de planificación energética multietapa cuando se dispone de poca información sobre el grado de incertidumbre de estos procesos. En la literatura, los trabajos relacionados con la planificación energética, en los que se han considerado varios períodos de inversión, la mayoría asumen que los procesos son independientes entre etapas, es decir, la información revelada en la etapa t de una variable aleatoria no depende de la información pasada. Sin embargo, los escenarios generados asumiendo la independencia por etapas también implican la aparición de escenarios en zigzag que no se dan en el mundo real. Comenzamos haciendo estimaciones simples de los rangos de variación de los parámetros inciertos para cada etapa, como las demandas y los costos de los recursos, y la correlación temporal de estos parámetros es modelada mediante modelos autorregresivos (AR) utilizando como datos de entrada los rangos de variación y un parámetro que controla el grado de correlación temporal. El modelo AR resultante de nuestro método tiene la propiedad de que las variables aleatorias correspondientes a cada período se distribuyen uniformemente dentro de esos rangos preestablecidos, y además permite equilibrar la aparición de escenarios zigzag y extremos mediante un parámetro de control α . Mostramos que nuestros AR son adecuados para ser representados en el algoritmo de programación dinámica dual estocástica (SDDP), ya que los errores cumplen el supuesto de independencia entre etapas. También analizamos las ventajas de considerar la dependencia por etapas en los problemas de programación estocástica multietapa. Para ello, aproximamos los procesos continuos AR(1) asociados a los costos de los recursos a una cadena de Markov de estado finito. Esto es para preservar la convexidad del problema sin perder la dependencia temporal en los costos de los recursos. Demostramos que los costos por fuera de la muestra tienen un mejor desempeño cuando se considera la dependencia temporal en el SDDP, especialmente en las realizaciones de escenarios extremos. Además, el caso en el que se supone que la incertidumbre es independiente en todos los períodos no permite la adaptación.

Introduction

0.1 Problem Statement and Motivation

Energy system planning offers an ideal investment strategy in efficient technologies and resources that will be required to ensure an affordable, secure, sustainable, and dependable future energy supply. Optimization models applied to energy systems are widely used to generate knowledge to inform energy and environmental policy. Common purposes include minimizing the investment and operating costs of the system, assessing the economic and political implications, reducing dependence on fossil fuels, increasing the share of renewable energies and efficient technologies, and its sensitivity to future uncertainties.

Analyses in these models usually cover a period of several decades, i.e., 20 to 50 years. Some optimization models based on energy systems have been formulated in their entirety as a linear programming (LP) problem such as **DIETER** [Zerrahn and Schill, 2015] and **ENTIGRIS** [Senkpiel et al., 2016]. Tools such as **MARKAL/TIMES** [Krzemień, 2013], **OSeMOSYS** [Howells et al., 2011], **ETEM** [Babonneau et al., 2017], **MESSAGE** [Sullivan et al., 2013] and **SMART** [Powell et al., 2012] enable system planning with multiple investment periods, and time resolution of typical days or finer. In addition, they can incorporate different energy products, such as heating, gas, and transportation networks, in addition to electricity. For more details of the existing models we refer the reader to Ringkjøb et al. [2018].

Most of these energy models are “deterministic”, this means that it is based on long-term forecasts of key parameters such as fuel prices and demand. However, one of the intrinsic characteristics of these models is the absence of reliable data and the prevalence of measurement errors in the input parameters, which has made these systems extremely difficult to analyze. For example, for power system planning, uncertainty arises mainly in demand, efficiency, technology costs, and energy resource prices. Tools such as **ETEM** and **MARKAL/TIMES** have been recognized to incorporate uncertainty in models, but these have been limited to considering only some uncertain parameters. When the uncertainty associated with the prediction of such parameters is considered in the model, this implies the construction of large and complex models, which require advanced optimization algorithms to obtain results that are optimal for the largest number of possible scenarios.

Stochastic Programming (SP) has been widely used to deal with optimization models under uncertainty [Prékopa, 1995]. Under this approach, one or more uncertain parameters can be modeled through random variables. In particular, the two-stage (TS) or multi-stage (MS) SP allows sequential decisions to be made conditional on information revealed at any point in time. In a standard two-stage stochastic programming model, the decision variables are divided into two groups; namely, the first stage variables and the second stage variables. The first-stage variables (or decisions) are made with the data available at the time, before observing the realization of the random variable (“here and now”); in the second stage, once the information of the uncertain parameter is revealed (“wait and see”), corrective actions are taken, this ability to act is known as recourse. SP applied to energy planning problems, has demonstrated its superiority over deterministic models [Usher and Strachan, 2012]. However, the limitations present in SP are: knowledge of the true distribution and its high computational cost, which increases linearly with the number of scenarios and the number of stages. An important issue that is often overlooked in MSSPs is how uncertainty is incorporated into the model. Sometimes simple form statistical forecasting techniques, such as first-order autoregressive time series models, are used to extract scenarios and incorporate them into the optimization model; however,

it is not always possible to estimate or calibrate the model parameters due to the limited information available in the historical data.

On the other hand, robust optimization (RO) is another approach used to deal with optimization models under uncertainty. Its growing interest in recent years was motivated by the work of [Ben-Tal and Nemirovski \[1998\]](#), and unlike SP, it does not require a specific distribution for each uncertain parameter. Instead, the uncertain data are assumed to vary over a deterministic uncertainty set, which has been specified by the user. RO adopts a min-max approach that addresses uncertainty such that feasibility can be guaranteed for any realization of the uncertain parameters within the uncertainty set. However, it is known that RO can generate very conservative solutions, which naturally increases the cost of solutions [\[Bertsimas and Sim, 2004\]](#). In the literature, there are several assumptions to be imposed on the uncertainty set structure in order to have computationally tractable problems and less conservative solutions. Among them, we can mention ellipsoidal and polyhedral uncertainty sets [\[Ben-Tal et al., 2015\]](#). Another approach that has recently gained much popularity is called distributionally robust optimization (DRO). The DRO combines the ideas of RO and SP, but instead of constructing an uncertainty set from uncertain data as in RO, the DRO assumes that the underlying probability distribution is unknown and lies in an ambiguity set of probability distributions. As in robust optimization, this approach hedges against the ambiguity in probability distribution by taking a worst-case approach, to provide the model with protection against the worst distribution [\[Wiesemann et al., 2014\]](#). Many of the existing studies on DRO have focused on how to construct the ambiguity set and how to transform the resulting DRO into computationally tractable models such as MILP, LP, among others. The most popular ambiguity sets that have been proposed are the following: Moment-based and Distance-based [\[Mohajerin Esfahani and Kuhn, 2018\]](#).

A robust model in the field of energy planning was developed by [Moret et al. \[2020d\]](#), applied to a real Swiss energy system. In this model, the authors propose planning for 2035, in which decisions are made “here and now” with the parameters projected in the last planning period. Compared to the models that have been presented (from the **MARKAL/TIMES** family of models), which consider hourly time intervals and multiperiod investment plans, the [Moret et al. \[2020d\]](#) model is a more simplified representation with a single planning period, multienergy, which aims to find the optimal investment strategy to meet future energy needs at minimum (investment and operating) cost. On the one hand, the authors manage to incorporate a large number of uncertain parameters in the model, both in the objective function and in the right-hand side of the constraints, and in the coefficients of the variables thanks to their simplified version, but on the other hand, this model lacks adaptability, i.e., the decision variables cannot self-adjust once a certain part of the uncertain parameter information is revealed. However, there are some promising methods, such as adaptive or adjustable robust optimization (ARO), which allows sequential decision making and control of the level of conservatism.

The authors of [Moret et al. \[2020d\]](#) demonstrate that uncertainty drastically influences strategic energy planning decisions. In fact, the deterministic solution, i.e., the energy strategy defined without taking uncertainty into account, changes dramatically as soon as uncertainty is considered, also stresses that, in order to obtain sound and firm investment decisions, it is crucial to consider key uncertainties in energy models. On the other hand, requiring decision sequences to adapt to the realization of uncertainty is very relevant in energy planning problems, which can be achieved by stochastic optimization. However, seen under the stochastic approach, this involves the task of assuming or assigning a probability distribution to each uncertain parameter reported by [Moret et al. \[2020d\]](#), a task that is prone to errors. Therefore, a number of questions arise, which form the basis of our work: *what would be the impact of distributional assumptions on investment decisions?* *what are the solutions according to the stochastic approaches used?* *how do determine the key uncertainties of the problem?* *what could be the impact of considering time dependence in strategic energy planning?* *how to model the uncertainty of the uncertain parameters with little statistical information, so that the evolution of the processes over the planning horizon provides a more realistic representation of the problem?*

0.2 Outline of this Dissertation

Two papers are provided in this dissertation, one in each chapter. The first paper was published in August 2020 in Applied Energy [Guevara et al., 2020], while the second paper was submitted to EJOR [Guevara et al., 2022].

In Chapter 1, we present a two-stage stochastic programming model with a single investment period based on the energy modeling framework by Moret et al. [2020d], where we analyze how the choice of statistical approaches and distributional assumptions can impact investment solutions. We start by assuming different distributions over the ranges of variation of the second stage parameters, making sure that the assumed distributions over the uncertain parameters are consistent with the support and its associated nominal value. Nominal values not only indicate the most likely value of the ranges but also indicate that the data is distributed 50% above and 50% below the range, which provides additional information on the asymmetry of the ranges. The output solutions of the TSSP models show that there is a large discrepancy between the stochastic solutions associated with the assumed distribution and other robust solutions found in Moret et al. [2020d]. In order to avoid this sensitivity problem in investment solutions related to the assumed distribution, we combine machine learning with DRO, to restrict the ambiguity set to a set of most important parameters and thus preserve the computational tractability of the resulting problem. There are few two-stage DRO proposals in the literature, however, we found that the most suitable for our two-stage DRO model is the TSDR-LPs algorithm described by Bansal et al. [2018] since in our model randomness can be present in all elements of the second stage problem. Moreover, its formulation as LP allows the use of existing solvers and the decomposition of the original problem. A total of three distance-based DRO models were developed, whose reference distribution has similar distribution assumptions to those used in the TSSP models. Results obtained from out-of-sample simulations reveal that strategic DRO solutions are more stable to variations in the underlying probability distributions.

In Chapter 2, we present an extension of two-stage to multistage stochastic programming, where we address the modeling of stochastic processes in multistage long-term energy planning problems when little information is available about the degree of uncertainty of these processes. In the literature, works related to energy planning [Ioannou et al., 2019, Lara et al., 2019, Rebennack, 2014], in which several investment periods have been considered, most assume that the processes are independent between stages (or stage-wise independent, i.e., the information revealed at stage t of a random variable does not depend on past information). However, the scenarios generated by assuming stage-wise independent also involve the appearance of zigzag scenarios that do not occur in the real world. Therefore, we start by making simple estimates of the ranges of variation of uncertain parameters for each stage, as the demands and price of resources, and then we modeled the temporal correlation of these parameters by (first-order) autoregressive (AR) time-series models using the ranges of variation as input data. The AR model resulting from our method has the property that the random variables corresponding to each period are uniformly distributed within those pre-specified ranges, and also allows balancing the occurrence of zigzag and extreme scenarios using an α control parameter. We also show that our ARs are suitable to be represented in the Stochastic Dual Dynamic Programming (SDDP) algorithm [Pereira and Pinto, 1991]. The SDDP is an efficient technique to solve large-scale multistage stochastic linear programs, also based on Benders decomposition, incorporating Monte Carlo sampling to approximate the original problem using only a subset of the total scenarios at each iteration. This algorithm assumes that the MSSP has a relatively complete resource, and that stochastic processes are stage-wise independent, but this last assumption may be unrealistic. To preserve the stage-wise dependent and maintain the convexity of the stochastic problem, we discretize the AR models associated with the cost parameters involved in the objective function by Markov chains. The resulting formulation is then solved with an advanced SDDP algorithm available in the literature that handles finite-state Markov chains. Next, to gain insight into the advantages of using time dependence in energy planning problems, we propose four versions of multistage stochastic programming, where the first version considers the demands and cost of resources to be stage-wise independent, and the rest consider stage-wise dependent under three different levels of correlation. The out-of-sample simulation results show that the average costs of all models are very similar. However, by considering stage-wise dependent in the SDDP, better performance and protection

against extreme scenarios are obtained. Moreover, the case where uncertainty is assumed to be independent in all periods does not allow for adaptation: although investment decisions may vary in each period, they are not adapted to the specific scenarios observed in each period.

Conclusions and possible future research directions are provided in Chapter 3.

Chapter 1

A machine learning and distributionally robust optimization framework for strategic energy planning under uncertainty

1.1 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. Based on a recent review work [Limpens et al., 2019], the most commonly used optimization energy models are MARKAL/TIMES [Krzemień, 2013], OSeMOSYS [Howells et al., 2011], ETEM [Babonneau et al., 2017], MESSAGE [Sullivan et al., 2013], SMART [Powell et al., 2012], while more recent and promising options are 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. Gabrielli et al. [2019] focus on an urban (decentralized) energy system in which the weather conditions, the demand for energy (electricity and heat) and the price of the energy are considered uncertain. 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]. This robust optimization framework is also used in Moret et al. [2020c] to analyze the issue of overcapacity in Europe. 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 (ML) 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].

In our work, we build an ambiguity set in which we assume that the true PDFs are close (using a Wasserstein distance) to a given reference distribution. To the best of our knowledge, ours is the first work to use DRO tools in the context of strategic long-term energy planning. A distinguishable feature of our model, compared to the aforementioned works that apply DRO in other energy settings, is the large dimension of the

underlying uncertainty—there are more than 70 uncertainty parameters in the model. Such large dimension creates enormous computational challenges for the DRO approach. To circumvent this issue, we use machine learning tools (henceforth denoted ML for short) 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 1.2, we present a compact formulation of the EnergyScope model and we establish our focus of analysis. Also, we introduce the two-stage stochastic programming formulation. Then, we present the novel combination of DRO with ML that we implement to produce a tractable robust dynamic planning energy model. In Section 1.3, we describe the uncertainty of the model parameters and define different PDFs for the most important uncertain parameters. The stochastic solutions obtained by using different PDFs are then compared to robust solutions obtained from the literature. Then, we discuss the experimental results and main findings using the ML-DRO framework. We show that our approach produces robust and stable strategic solutions in relation to the assumptions of the reference probability distributions of uncertain parameters. Finally, concluding remarks are presented in Section 1.4.

1.2 Methodologies

1.2.1 Strategic energy model

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.

1.2.1.1 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. Due to a decision of the Swiss government of phase out nuclear power plants at the end of their useful life, the country is defining its future energy strategy. Therefore, the model considers the long-term planning of the energy system until 2035 with a monthly time resolution and 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

¹The version of the EnergyScope model used in this paper is an exact reproduction of the Swiss energy system model presented by Moret [2017]: the model is described in detail in chapter 1 of that thesis, while the data are documented in detail in Appendix A of the same thesis. The code is publicly available at <https://github.com/energyscope/EnergyScope/tree/v1.0>

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:

$$\min \quad c^T \mathbf{x} + e^T \mathbf{y} \quad (1.1a)$$

s.t.

$$A\mathbf{x} \leq b, \quad (1.1b)$$

$$T\mathbf{x} + W\mathbf{y} \geq d, \quad (1.1c)$$

$$\mathbf{x} \in X, \quad (1.1d)$$

$$\mathbf{y} \in Y, \quad (1.1e)$$

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 is a subset of $\mathbb{R}_+^{n_2}$.

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 (1.1b) 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 (1.1c) 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. On the other hand, taxes and subsidies are not accounted for in the objective function; in fact, being internal exchanges within the energy system boundaries, they do not contribute to the total cost.

1.2.2 The classical two-stage stochastic approach

Problem (1.1) 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:

$$\begin{aligned} \min_{\mathbf{x} \in X} \quad & c^T \mathbf{x} + \mathbb{E}[Q(\mathbf{x}, \xi)] \\ \text{s.t.} \quad & A\mathbf{x} \geq b, \end{aligned} \quad (1.2)$$

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

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

and $\xi := (e, T, W, d)$ indicates that the uncertainty can be present in any of the coefficients of the second-stage problem. In the two-stage model formulation, the corresponding variables and constraints are allocated into the first-stage problem, as shown in constraints (A.2)-(A.12) in Appendix A. The operating variables and their corresponding constraints are placed in the second-stage, as shown in constraints (A.14)-(A.32) in Appendix

A. 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 (1.2) 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 Problem (1.2) 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 the approach we implement to generate a reduced set of 1,500 samples that yields an acceptable optimality gap and thus an acceptable approximation of the expected value.

1.2.3 A Machine Learning and Distributionally Robust Optimization Framework

In this section, we introduce a Distributionally Robust Optimization (DRO) approach and the challenges presented by the model when considering a large number of uncertain parameters. 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. Then, we introduce Machine Learning (ML) tools to identify a reduced subset of the most significant uncertain parameters that will be considered in the construction of the ambiguity set, and thus alleviate the computational time of the DRO models.

1.2.3.1 Distributionally robust optimization for two-stage models

The DRO formulation of the Problem (1.2) can be written as follows:

$$\min_{\mathbf{x} \in X} c^T \mathbf{x} + \max_{\mathbb{P} \in \mathcal{D}} \mathbb{E}_{\mathbb{P}}[Q(\mathbf{x}, \xi)], \quad (1.3)$$

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 -divergence² [\[Ben-Tal et al., 2013\]](#), Wasserstein distance [\[Mohajerin Esfahani and Kuhn, 2018\]](#) and total

²The 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.

variation distance [Rahimian et al., 2019a]. 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 (1.3) can be written as a risk-averse version of Problem (1.2) 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 Problem (1.3) exactly is in general very challenging and its tractable reformulation will depend on the ambiguity set chosen. Using the Wasserstein distance, different reformulations of the Problem (1.3) 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 the cost vector e and in the technology matrix T . They show that, under proper assumptions and with a 2-norm Wasserstein distance centered on a discrete reference distribution, the two-stage DRO problem 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 TSDR-LPs algorithm 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 as an LP allows for the use of existing solvers 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, not only in the vectors e and d but also in the matrices W and T . Since in our model the randomness can be present in all of those elements, 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.

1.2.3.2 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

$$W_p(\mathbb{P}_1, \mathbb{P}_2) := \left(\inf_{\Pi \in \Gamma_m(\mathbb{P}_1, \mathbb{P}_2)} \mathbb{E}_{\Pi}[\|\xi - \xi'\|^p] \right)^{1/p}, \quad (1.4)$$

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}_ϵ 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:

$$\mathcal{D}_\epsilon := \{\mathbb{P} \in \mathcal{M}_m(\Omega) : W_1(\mathbb{P}, \mathbb{P}_0) \leq \epsilon\}. \quad (1.5)$$

That is, the ambiguity set \mathcal{D}_ϵ 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 (1.2), whereas a value of $\epsilon = \infty$ (in practice, a large value of ϵ) corresponds to solving a robust version of Problem (1.2) 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.

1.2.3.3 Model formulation and algorithm

In this section, we review the algorithm presented in [Bansal et al. \[2018\]](#) to solve Problem (1.3), which we enhance through the addition of feasibility cuts. The algorithm is a Benders decomposition method which uses a distribution separation algorithm. An important assumption we make is that (i) the support Ω has a finite number L of points, and (ii) the reference distribution \mathbb{P}_0 also has finite support. Part (i) of this assumption is enforced by restricting the distributions \mathbb{P} in (1.5) to those with finite number L of points, whereas for part (ii) we replace the original reference distribution \mathbb{P}_0 with an empirical distribution corresponding to N samples drawn from \mathbb{P}_0 . Consequently, the ambiguity set \mathcal{D}_ϵ is defined by a polytope with a finite number of extreme points. Notice initially that the Problem (1.3) cannot be solved directly by a general purpose optimization solver, since this is a “min-max-min” problem. Therefore, Problem (1.3) is divided into three subproblems: a first-stage Master problem, a second-stage subproblem and distribution separation problem. The methodological details are described below. Problem (1.3) can be formulated as:

$$\begin{aligned} & \min_{\mathbf{x} \in X} \quad c^T \mathbf{x} + \theta \\ & \text{s.t.} \\ & \max_{\mathbb{P} \in \mathcal{D}_\epsilon} \{\mathbb{E}_{\mathbb{P}}[Q(\mathbf{x}, \xi)]\} \leq \theta. \end{aligned} \tag{1.6}$$

Since the probability distribution \mathbb{P} supported in Ω is finite, the constraint of the above problem can be expressed as:

$$\sum_{l=1}^L \mathbb{P}(\xi_l) Q(\mathbf{x}, \xi_l) \leq \theta \quad \forall \mathbb{P} \in \mathcal{D}_\epsilon. \tag{1.7}$$

Through this formulation, [Bansal et al. \[2018\]](#) propose an cutting-plane approach whereby the constraints (1.7) are generated sequentially. More specifically, given a particular solution $\bar{\mathbf{x}}$, the following separation problem is solved

$$\max \left\{ \sum_{l=1}^L \mathbb{P}(\xi_l) Q(\bar{\mathbf{x}}, \xi_l) : \mathbb{P} \in \mathcal{D}_\epsilon \right\}. \tag{1.8}$$

To simplify the notation, hereinafter we shall write p_l to denote $\mathbb{P}(\xi_l)$. The cuts for the Master problem are given by the following inequality constraint:

$$\sum_{l=1}^L \bar{p}_l \{(\bar{\mu}_l)^t (d_l - T_l \mathbf{x})\} \leq \theta,$$

where the \bar{p}_l for $l = 1, \dots, L$ are obtained by solving the distribution separation problem (1.8) and $\bar{\mu}_l$ for $l = 1, \dots, L$ are optimal dual multipliers corresponding to constraint set $W_l \mathbf{y} = d_l - T_l \bar{\mathbf{x}}$. Before obtaining $\{\bar{p}_l\}_{l=1}^L$ from the distribution separation problem, the first stage decision variable $\bar{\mathbf{x}}$ is required to be feasible for all second-stage problems $Q(\bar{\mathbf{x}}, \xi)$. In case that $\bar{\mathbf{x}}$ is infeasible for some l , with $l = 1, \dots, L$, we add the following cut (also called “*feasibility cut*” in the Benders decomposition method):

$$(\nu_l)^t (d_l - T_l \mathbf{x}) \leq 0$$

to the Master problem to restrict movements in that direction, where ν_l is an extreme ray associated with the dual formulation of the second-stage problem.

To summarize, the Master problem has the following structure:

$$\begin{aligned}
 (\text{Master}) \quad & \min_{\mathbf{x} \in X} \quad c^T \mathbf{x} + \theta \\
 & \text{s.t.} \\
 & \text{Optimality Cuts : } \sum_{l=1}^L \bar{p}_{j,l} \{(\bar{\mu}_{j,l})^t (d_l - T_l \mathbf{x})\} \leq \theta \quad \forall j = 1, \dots, c_1, \\
 & \text{Feasibility Cuts : } (\bar{\nu}_i)^t (\bar{d}_i - \bar{T}_i \mathbf{x}) \leq 0 \quad \forall i = 1, \dots, c_2
 \end{aligned} \tag{1.9}$$

The formulation of the second stage problem is given by

$$(\text{second-stage}) \quad Q(\mathbf{x}, \xi_i) = \min_{\mathbf{y} \in Y} \quad e^T \mathbf{y} \tag{1.10}$$

and the distribution separation problem for a given $\mathbf{x} \in X$ is formulated as

$$\begin{aligned}
 (\text{Distribution separation}) \quad & \max_{\pi, p} \quad \sum_{l=1}^L p_l Q(\mathbf{x}, \xi_l) \\
 & \text{s.t.} \\
 & \sum_{i=1}^L \sum_{j=1}^N \|\xi_i - \xi_j^*\|_1 \pi_{ij} \leq \epsilon, \\
 & \sum_{j=1}^N \pi_{ij} = p_i \quad i = 1, \dots, L, \\
 & \sum_{i=1}^L \pi_{ij} = p_j^* \quad j = 1, \dots, N, \\
 & \sum_{i=1}^L p_i = 1 \quad , \\
 & \pi \geq 0,
 \end{aligned} \tag{1.11}$$

where $\xi \in \Omega$, $\xi^* \in \Omega_0$, $L = |\Omega| > 0$, $N = |\Omega_0| > 0$ and $\sum_{i=1}^L p_i = \sum_{j=1}^N p_j^* = 1$. The pseudocode of the algorithm that solves the Problem (1.3) is presented in Algorithm (1). For a better understanding of the estimation of the upper and lower bounds, we refer the reader to [Bansal et al. \[2018\]](#).

Algorithm 1: Two-stage distributionally robust method from [Bansal et al. \[2018\]](#), enhanced with feasibility cuts

```

1 Initialization:  $k = 0$ ,  $LB = -\infty$ ,  $UB = \infty$ ,  $\epsilon = \epsilon^*$ ,  $\delta = \delta^*$ ,  $c_1 = c_2 = 0$ 
2 while  $\frac{|UB-LB|}{UB} \geq \delta^*$  or  $k \leq maxIter$  do
3   feasible  $\leftarrow$  TRUE
4   Get  $(\mathbf{x}^k, \theta^k)$  solving MILP Master problem (1.9)
5   for  $l = 1 : L$  do
6     Solve LP  $Q(\mathbf{x}^k, \xi_l)$ , (second-stage (1.10))
7     if second-stage “ $l$ ” is infeasible (i.e. its dual is unboundeda) then
8       Get  $\nu_l \leftarrow$  extreme ray of second-stage (1.10)
9       Derive Feasibility Cut:  $\bar{\nu}_{c_2+1} = \nu_l$ ,  $\bar{d}_{c_2+1} = d_l$ ,  $\bar{T}_{c_2+1} = T_l$ 
10      Add Feasibility Cut:  $c_2 = c_2 + 1$ 
11      feasible  $\leftarrow$  FALSE
12     $Q(\mathbf{x}^k, \xi_l) \leftarrow$  optimal solution value
13   $LB \leftarrow c^t \mathbf{x}^k + \theta^k$ 
14  if feasible then
15    Solve distribution separation problem (1.11) using  $Q(\mathbf{x}^k, \xi_l)$ , to get  $p_l$ ,  $l = 1, \dots, L$ 
16    if  $UB > c^t \mathbf{x}^k + \sum_{l=1}^L p_l Q(\mathbf{x}^k, \xi_l)$  then
17       $UB \leftarrow c^t \mathbf{x}^k + \sum_{l=1}^L p_l Q(\mathbf{x}^k, \xi_l)$ 
18      if  $\frac{|UB-LB|}{UB} \leq \delta^*$  then
19         $\mathbf{x}^* \leftarrow \mathbf{x}^k$ 
20        Go to line (26)
21     $\mu_l \leftarrow$  optimal dual multipliers obtained by solving  $Q(\mathbf{x}^k, \xi_l) \forall l = 1, \dots, L$ 
22    Derive Optimality Cut:  $\bar{p}_{c_1+1,l} = p_l$ ,  $\bar{\mu}_{c_1+1,l} = \mu_l$ ,  $\forall l = 1, \dots, L$ 
23    Add Optimality Cut:  $c_1 = c_1 + 1$ 
24     $\mathbf{x}^* \leftarrow \mathbf{x}^k$ 
25   $k \leftarrow k + 1$ 
26 return  $x^*, LB$ 

```

^aWe assume that the dual problem is feasible.

As discussed in Section 1.2.2, the number L of possible outcomes can grow exponentially with the number m of uncertain parameters of the model. Moreover, since we use an empirical distribution to approximate the original reference distribution, it follows from well-known results in probability theory (see, e.g., [Dudley 1969](#)) that the number of samples required to obtain a given precision grows exponentially with m . Hence, it is impractical to have random vectors ξ even of moderate dimension, especially considering that the separation problem (1.8) 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.

1.2.3.4 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]. XGBoost also has some advantages over other ML algorithms, since it is able to parallelize the computation to construct trees, and handle missing data efficiently. Moreover, it is shown in the literature that XGBoost has the maximal performance among all algorithms within the category boosting [Memon et al., 2019]. To further confirm these results from the literature, we have compared the performance of the random forest, regression trees, gradient boosting machines and leaps (regression subset selection) methods with XGBoost and found that the latter was the best algorithm based on the metrics discussed below.

The first step consists in generating a large sample of random parameter scenarios and in solving a large number of deterministic Problems (1.1) (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 (installed capacity size). 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. In addition, we use the *information gain* metric of XGBoost as a measure to rank each parameter.

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:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \tilde{y}_i)^2} \quad (1.12)$$

$$R^2 = 1 - \frac{\sum_i^n (y_i - \tilde{y}_i)^2}{\sum_i^n (y_i - \bar{y}_i)^2} \quad (1.13)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \tilde{y}_i| \quad (1.14)$$

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

1.3 Results and Discussion

1.3.1 Model uncertainty

As discussed in Moret et al. [2017], uncertain parameters in Problem (1.1) 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 1.1 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 of Problem (1.1).

Parameters	Min %	Max %	Element in Problem (1.1)
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 1.1: 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 (1.1) through the elements c, e, d, T and W.

Here and throughout the paper, the uncertainty ranges have been defined based on values in the literature, forecasts, historical data, etc., following the methodology and data presented in previous work [[Moret et al., 2017](#)].

1.3.2 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 in Section 1.2.2 and generate scenarios assuming different probability distributions. Then, we analyze the resulting stochastic solutions and compare them with robust solutions reported in the literature.

1.3.2.1 Uncertainty assumptions

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 Problem (1.2). 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 Section 1.1. The two stochastic solutions are defined as follows:

- *Stochastic-U*: The first stochastic solution is obtained by solving Problem (1.2) and assuming, as in [Moret et al. \[2020a\]](#), uniform distributions for all uncertain second-stage parameters with variation ranges as reported in Table 1.1.
- *Stochastic-L*: The second stochastic solution is obtained by solving Problem (1.2) 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.

Note that, in these stochastic models, the uncertain first stage parameters that appear in the objective function (e.g., investment costs) take their nominal value, since the expected value of a random variable corresponds to its nominal value. As discussed in Section 1.2.2, we adopt a sample average approximation approach to solve the problems, using 1,500 samples. The resulting 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. Based on the strong duality theorem, the equivalent robust counterpart of the uncertain constraints is formulated as a set of linear constraints. The resulting robust model belongs thus to the realm of linear programming and can be solved directly with a linear solver without need of a decomposition algorithm [\[Bertsimas and Sim, 2004\]](#). We refer the reader to [Moret et al. \[2020a\]](#) for more details.

1.3.2.2 An empirical assessment

Figure 1.1 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 F_{cp} for production, i.e., taking into account the yearly available factor cp of the technologies.

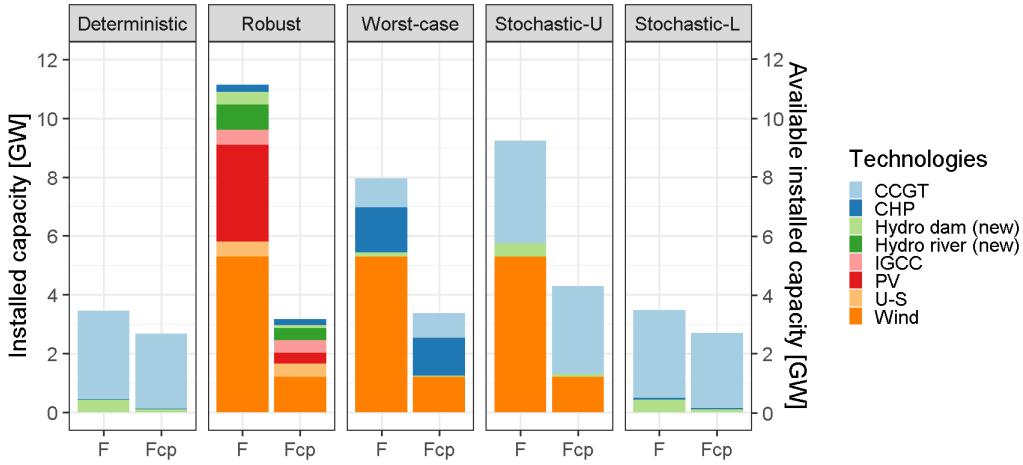


Figure 1.1: Electricity capacity mix (Full capacity F and available capacity F_{cp}) for the five investment strategies: *Deterministic*, *Robust*, *Worst-case*, *Stochastic-U* and *Stochastic-L*. (Acronyms: Photovoltaics (PV), Combined Cycle Gas Turbine (CCGT), Cogeneration of Heat and Power (CHP), 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.

The implementation and computational details of the Deterministic, Robust, Worst-case and Stochastic (presented as a very large deterministic model) models used in this first experiments, are shown in the Table 1.2. All the problems in the Table 1.2 are MILP and can be solved by CPLEX without using decomposition techniques, except for the stochastic models which require decomposition techniques. The size of the problems Robust and the Worst-case are the same, the difference lies in the value of the Γ parameter, which controls the number of parameters that take their worst value.

	Deterministic	Robust ^c	Worst-case ^c	Stochastic models ^a
Type	MILP	MILP	MILP	MILP
Variables (cont.)	1,469	19,646	19,646	2,071,500
Variables (bin.)	118	118	118	72,070
Variables (int.)	56	56	56	56
Constraints	2,236	4,868	4,868	3,031,500
Optimizer	CPLEX-12.8	CPLEX-12.7	CPLEX-12.7	CPLEX-12.8
CPU (s)	0.17	5.00	1.97	2,095.87/1,907.99 ^b

^a We can express Problem (1.2) in a deterministic form by introducing a different second-stage \mathbf{y} variable for each scenario. This formulation is called the deterministic equivalent.

^b There are two computational time obtained by the Benders decomposition method, corresponding to the *Stochastic-L* and *Stochastic-U* solutions respectively.

^c The robust and worst-case solutions have been computed in [Moret et al. \[2020b\]](#) on a different machine using CPLEX-12.7. As the CPU times are quite low and solution times are not the focus of our paper, we did not reproduce those results on our machine.

Table 1.2: Comparison of problem sizes and solution time for deterministic, robust and stochastic solutions.

1.3.2.3 “Out-of-Sample” simulation process

To assess and compare the economic performance of the five solutions in Figure 1.1, 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 1.4. In Table 1.3, 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 \pm half-width/std		
Distributions		Investment cost	Operations cost	Total
Uniform	<i>Deterministic</i>	1,406.4 \pm 2.34/119.7	7,137.4 \pm 30.00/1,530.5	8,543.8 \pm 30.12/1,536.7
	<i>Robust</i>	3,506.8 \pm 5.80/296.3	5,451.7 \pm 16.48/841.2	8,958.5 \pm 17.34/885.0
	<i>Worst-case</i>	2,254.7 \pm 3.60/183.8	6,389.5 \pm 26.08/1,330.8	8,644.2 \pm 26.57/1,355.4
	<i>Stochastic-U</i>	2,847.7 \pm 4.85/247.8	5,598.5 \pm 20.94/1,068.3	8,446.2 \pm 21.56/1,100.0
	<i>Stochastic-L</i>	1,446.7 \pm 2.38/121.5	7,159.4 \pm 29.09/1,484.1	8,606.1 \pm 29.21/1,490.4
Lognormal	<i>Deterministic</i>	1,405.2 \pm 2.32/118.4	6,048.3 \pm 17.00/866.8	7,453.5 \pm 17.18/876.7
	<i>Robust</i>	3,473.5 \pm 5.73/292.3	4,668.6 \pm 8.87/452.0	8,142.1 \pm 10.56/539.0
	<i>Worst-case</i>	2,241.5 \pm 3.55/181.3	5,388.2 \pm 14.57/743.5	7,629.7 \pm 15.17/774.3
	<i>Stochastic-U</i>	2,842.8 \pm 4.88/249.0	4,789.9 \pm 11.82/603.1	7,632.7 \pm 12.81/654.0
	<i>Stochastic-L</i>	1,445.6 \pm 2.35/120.3	6,092.8 \pm 16.35/834.4	7,538.4 \pm 16.55/844.7
Triangular	<i>Deterministic</i>	1,402.9 \pm 1.63/83.2	6,740.2 \pm 21.86/1,115.6	8,143.1 \pm 21.94/1,119.3
	<i>Robust</i>	3,479.3 \pm 4.05/206.9	5,184.3 \pm 11.42/583.0	8,663.6 \pm 12.00/612.2
	<i>Worst-case</i>	2,252.9 \pm 2.47/126.4	6,005.0 \pm 18.74/956.2	8,257.9 \pm 19.11/975.3
	<i>Stochastic-U</i>	2,836.4 \pm 3.41/174.0	5,326.9 \pm 15.15/773.3	8,163.3 \pm 15.55/793.5
	<i>Stochastic-L</i>	1,443.7 \pm 1.65/65.6	6,783.9 \pm 21.10/1,076.9	8,227.6 \pm 21.18/1,080.6

Table 1.3: Comparing the mean, half-width and standard deviation of the investment, operation and total costs obtained from stochastic, robust and deterministic models through different out-of-sample distributions.

First, we observe in Table 1.3 that the estimates for the mean costs 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.

Distributions	Infeasibility		Elec. Imports
	% of simulations	Demand shortage	in GWh/h
Uniform	<i>Deterministic</i>	53.7%	0.84%
	<i>Robust</i>	0.03%	0.22%
	<i>Worst-case</i>	0%	0%
	<i>Stochastic-U</i>	0.19%	0.12%
	<i>Stochastic-L</i>	0.19%	0.14%
Lognormal	<i>Deterministic</i>	54.0%	0.86%
	<i>Robust</i>	0.03%	0.05%
	<i>Worst-case</i>	0%	0%
	<i>Stochastic-U</i>	0.19%	0.10%
	<i>Stochastic-L</i>	0.18%	0.12%
Triangular	<i>Deterministic</i>	55.2%	0.58%
	<i>Robust</i>	0%	0%
	<i>Worst-case</i>	0%	0%
	<i>Stochastic-U</i>	0%	0%
	<i>Stochastic-L</i>	0%	0%

Table 1.4: 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 1.4 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.

1.3.3 Numerical experiments with DRO

In this section, we solve the two-stage DRO model considering the 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.

1.3.3.1 Variable selection

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). We performed the ML analysis as described in Section 1.2.3.4 with predictor variables corresponding to the second-stage uncertain parameters of Table 1.2 and 8 target variables as described above. To do so, we used the *xgboost* R package by [Chen et al. \[2019\]](#).

In Table 1.5, we report the performance measures for each of the XGBoost models, introduced in Section 1.2.3.4, using the test samples (30% dataset).

Indices	Target variables							
	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 1.5: Performances of the XGBoost models on the testing dataset through several statistical indices.

We observe in Table 1.5 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.

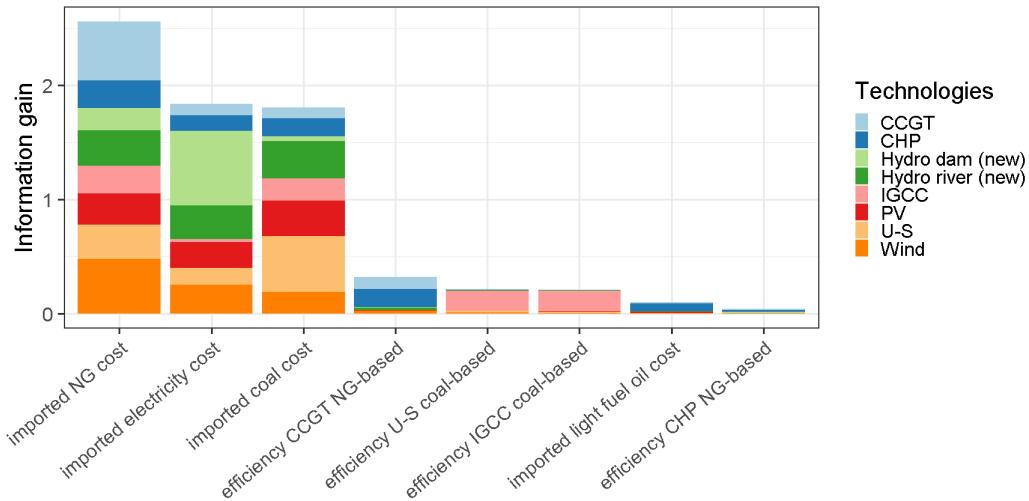


Figure 1.2: Information gains to the improvement of the XGBoost models stacked by parameters.

Figure 1.2 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 1.3 for the eight most important parameters.

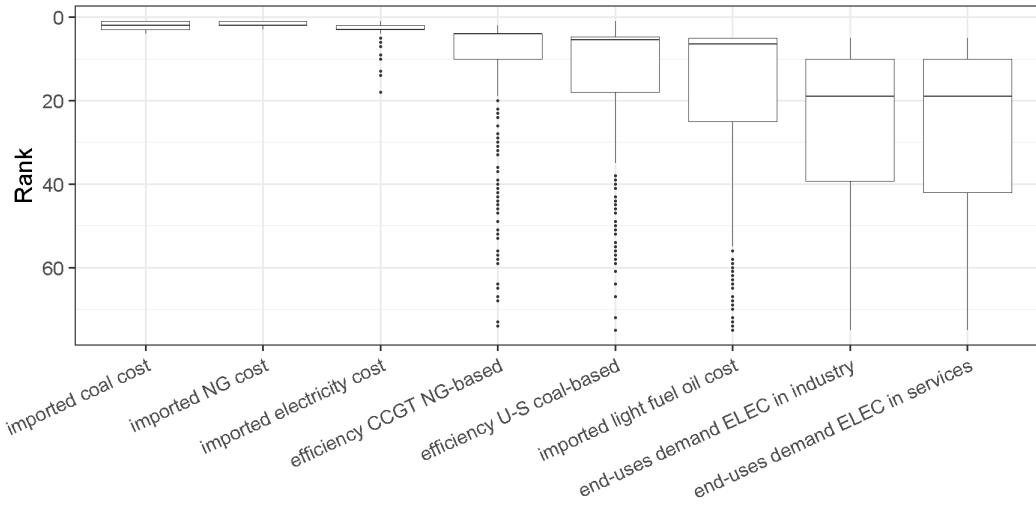


Figure 1.3: Boxplot visualizes the ranking statistics for the most important parameters over 50 runs.

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 1.2 and 1.3.

1.3.3.2 Setting of DRO ambiguity sets

As discussed in Section 1.2.3.1, a key component of the Problem (1.3) is the ambiguity set (1.5). 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

$$\mathcal{D}_\epsilon^U := \{\mathbb{P} \in \mathcal{M}_5(\Omega) : W_1(\mathbb{P}, \mathbb{P}_{\text{DRO-U}}) \leq \epsilon\} \quad (1.15)$$

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 1.1. Similarly, the second ambiguity set is given by

$$\mathcal{D}_\epsilon^L := \{\mathbb{P} \in \mathcal{M}_5(\Omega) : W_1(\mathbb{P}, \mathbb{P}_{\text{DRO-L}}) \leq \epsilon\}, \quad (1.16)$$

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. Note that in order to use the algorithm described in Section 1.2.3.3 we need to assume that the reference distribution has finite support; thus, we approximate the uniform and lognormal distributions with respective empirical distributions corresponding to 1,000 samples.

For the support of \mathbb{P} in the ambiguity sets (1.15) and (1.16), 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

1.1. 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 (1.15) and (1.16), 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, 0.136 and 1. The value $\epsilon = \epsilon_{min}$ refers to the minimum distance value for which the ambiguity set in (1.5) 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 .

1.3.3.3 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.

In Table 1.6, the scale and computational time of the DRO-L and DRO-U problems are presented, applying Algorithm 1. The Master problem is formulated as a MILP problem, while the subproblem and the distribution separation problem are formulated as a Linear Programming (LP). Ds is associated with the distribution separation problem. Note that CPU times of DRO problems increase exponentially with the number of uncertain parameters. With more than seven uncertain parameters, the problems become intractable. This confirms the importance of the variable selection procedure.

	<i>Master</i> ^a		Second-stage		Ds	
	Type	MILP	LP	LP	LP	LP
Variables (cont.)	88		2,125		243,243	
Variables (bin.)	11		0		0	
Variables (int.)	44		0		0	
Optimizer	CPLEX-12.8		Gurobi-8.01		CPLEX-12.8	
	$\epsilon = \epsilon_{min}$	$\epsilon = 0.084$	$\epsilon = 0.092$	$\epsilon = 0.108$	$\epsilon = 0.136$	$\epsilon = 1$
DRO-L	Constraints	3,960	4,027	5,303	6,421	5,453
	Total CPU (s)	346.62	425.58	497.14	715.50	677.97
DRO-U	Constraints	3,366	3,864	4,260	6,277	7,151
	Total CPU (s)	324.71	359.72	400.83	751.61	515.64
						497.25

^a The size of the Master problem corresponding to the last iteration.

Table 1.6: DRO problem sizes and comparison of solution statistics for DRO-L and DRO-U models with different radios.

We display in Figure 1.4 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.

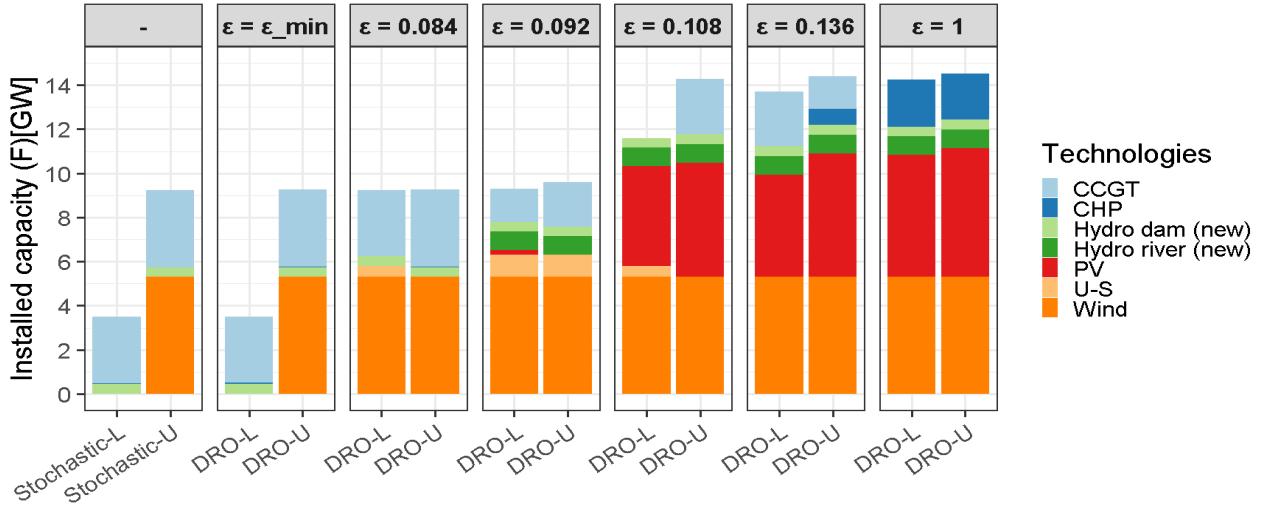


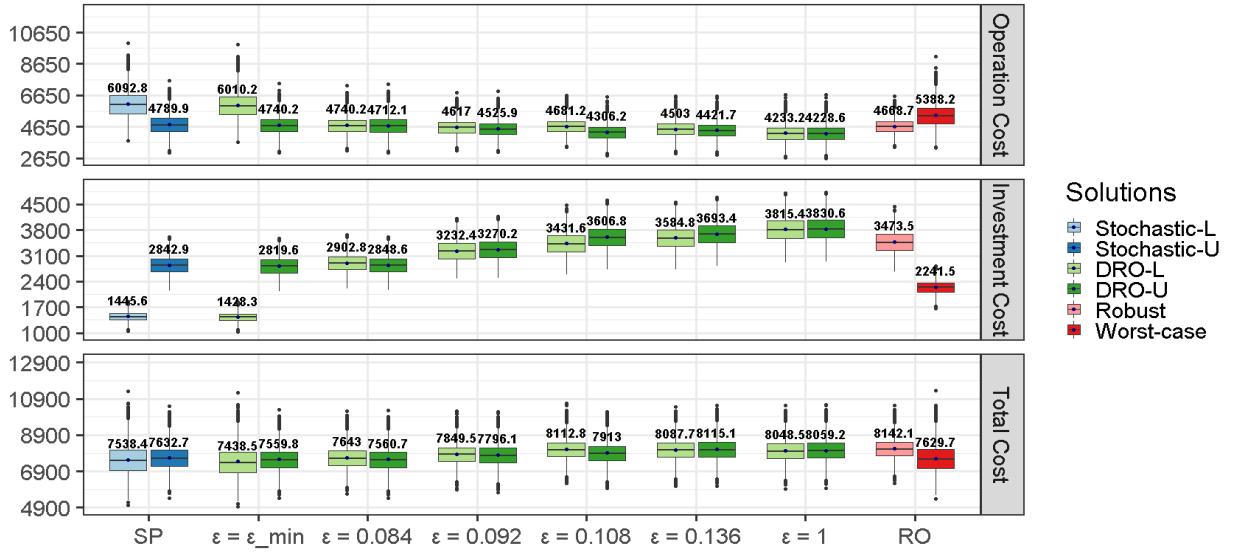
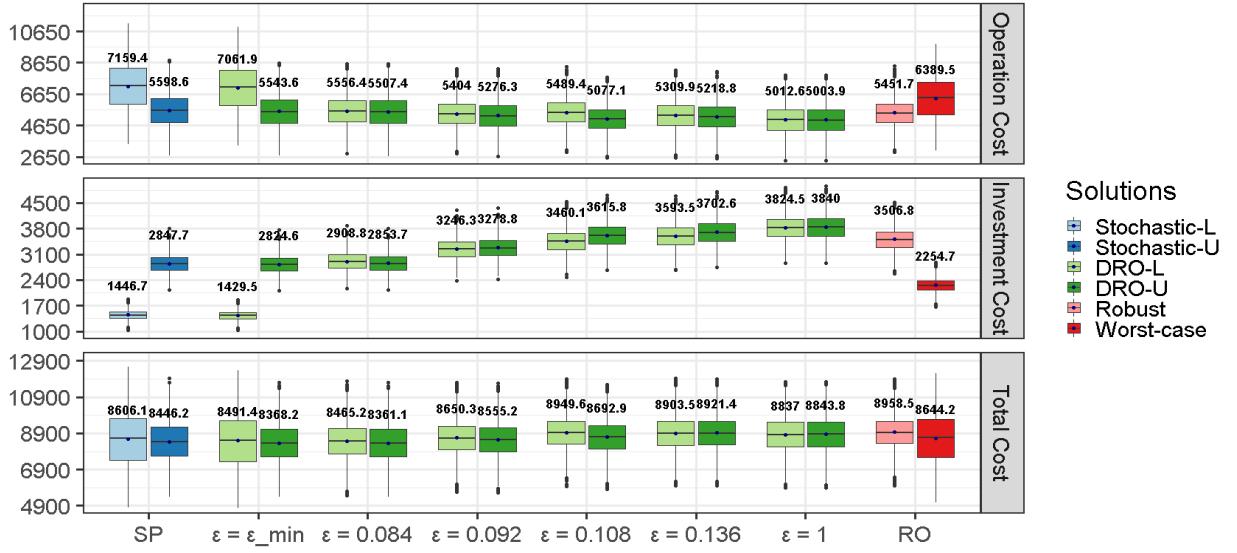
Figure 1.4: The effect of radius ϵ on the installed capacity of DRO-L and DRO-U solutions compared to the installed capacity of stochastic solutions (*Stochastic-L* and *Stochastic-U*).

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 1.2.3.1 that higher values of ϵ correspond to more risk-averse models. This desirable diversification effect is rather common when dealing with uncertainty. In particular, Bertsimas and Sim [2004] and Nicolas [2016] observe similar results in their numerical experiments using robust optimization formulations. 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.

1.3.3.4 Comparison of out-of-sample performances

To assess the economic performances of the DRO-L and DRO-U investment solutions of Figure 1.4, similarly to Section 1.3.2.3, we perform an “Out-of-Sample” simulation process assuming uniform, lognormal and triangular distributions. The simulation results are summarized in Figures 1.5, 1.6 and 1.7, respectively. For each solution, the figures display the boxplots for annual total cost, first-stage investment cost and second-stage operations cost.



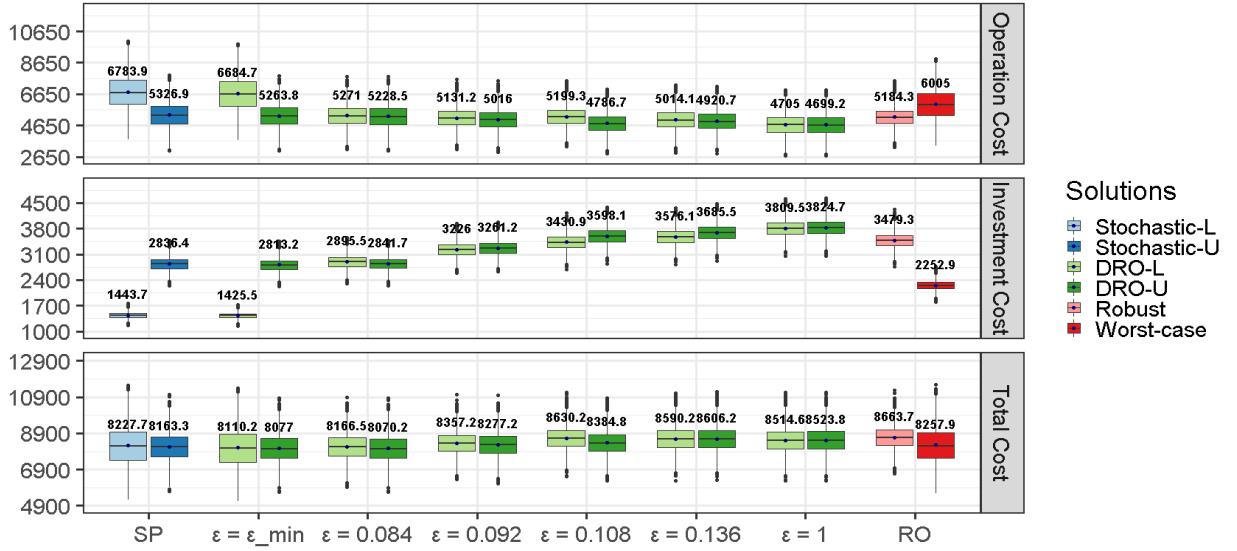


Figure 1.7: Boxplot of second-stage operations costs, first-stage investment costs and total cost. Simulations performed on $n_{sample} = 10.000$ scenarios generated with triangular distributions. The numbers indicate the average costs.

The main conclusion from Figures 1.5, 1.6 and 1.7 is that the performances of DRO solutions for a given ϵ radius are rather insensitive to the reference distribution in out-of-sample simulations. For example, for $\epsilon = 0.084$, the DRO-L and DRO-U yield very similar performances both in terms of investment and operations costs. Total costs for both DRO-L and DRO-U solutions are about 8400, 7500 and 8100 k\$ with uniform, truncated lognormal and triangular distributions, respectively.

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, producing the most extremes performances. The *Stochastic-L* solution has always a very low investment cost (with reduced dispersion) but compensated with a high operations cost while the *Stochastic-U* solution looks more balanced between the two cost components. The two stochastic solutions perform correctly (but still producing the worst total costs) only when assuming truncated lognormal distributions (Figure 1.6) which generates parameter scenarios with the lowest dispersion around the nominal value. In other words, as expected Stochastic programming is not efficient to protect against extreme realizations. We also see that the performances of *Robust* and *Worst-case* solutions are very different regardless the choice of the out-of-sample distribution. The *Robust* solution leads to higher total costs (average and extreme) than DRO solutions as already discussed in Section 1.3.2.3, while the *Worst-case* solutions have higher variability.

Based on these simulation results, the DRO approach appears to provide a good trade-off between the min-max approach adopted in Robust Optimization and the expected criterion of Stochastic Programming that relies on a specific probability distribution. DRO permits to overcome the drawbacks of the two alternatives while generating robust strategic investments decisions.

We also observe that more conservative DRO solutions corresponding to higher ϵ values 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 $\epsilon = 0.084$ seem to provide a good trade-off among those criteria, therefore it is our recommended strategy for this particular problem instance.

1.4 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 not only in the applied energy literature but also more generally in the stochastic optimization literature; indeed, we believe this is one of the contributions of our work.

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—that is, the performance of the solution does not change much when different distributions are used in the out-of-sample evaluation procedure—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.

We believe that the framework we propose in this paper can be helpful to other energy models, in several aspects: first, our results show that uncertainty should be considered in energy planning models, something that today happens very rarely. As seen in our results, the inclusion of uncertainty dramatically impacts/changes the deterministic solution. Thus, energy modelers should account for uncertainty from the early stages of the model development. A major challenge to include uncertainty factors in optimization models, of course, is the requirement of estimating the corresponding probability distributions. Our framework can also be very helpful in that regard, as we have demonstrated that the use of DRO tools can mitigate the effect of sensitivity of solutions with respect to the chosen distributions, even when the model involve a large number of uncertainty parameters.

We view our work as a first step in the use of DRO and machine learning tools for strategic energy planning. Thus, there is room to improve upon the limitations of our models. For instance, it would be helpful to have a method that could guide an *a priori* choice of the size of the ambiguity set. Some methods for that purpose have been proposed in the literature in the context of data-driven problems, whereby the level of ambiguity is determined based on the number of data points (see, e.g., [Mohajerin Esfahani and Kuhn \[2018\]](#) and [Blanchet et al. \[2019\]](#)); however, such techniques do not apply to our context, where the distributions are not necessarily obtained from data. One possibility, left for future work, could be to adopt the approach proposed by [Rahimian et al. \[2019b\]](#), who relate the size of the ambiguity set to a region of scenarios that are *critical* to the problem in a well-defined sense.

Other possibilities for future research work include 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. We also plan to apply the models studied in this paper to data from other countries; a study for the case of Chile is underway.

Chapter 2

Modeling uncertainty processes for multi-stage optimization of strategic energy planning : An auto-regressive and Markov chain formulation

2.1 Introduction

Energy plays a fundamental role in our lives as it is an essential contribution not only for daily tasks but also for carrying out various productive activities. Due to the constant transformation that the energy sector is undergoing, it is essential to plan long-term strategies that will guarantee our energy supply in the future. The work of [Rojas-Zerpa and Yusta, 2014, p. 67] indicates that “*energy planning implies finding a set of sources and conversion equipment that optimally satisfy the energy demand of all activities*”. In other words, a strategic energy plan establishes the resources and technologies that will be needed to meet our future energy needs. The planning horizon is generally long enough, i.e., 20 to 50 years, to identify profitable and robust developments according to technical, economic, social and environmental criteria.

Optimization models are generally used to find an optimal strategy to minimize the investment costs of new technologies and the operating and maintenance costs of the entire system while satisfying a set of underlying conditions. These models can be divided into two broad categories. In the first one, the system is optimized for a target year without looking at the transition. Models in this category, such as [Hilpert et al., 2018], Calliope [Pfenninger and Pickering, 2018] and EnergyScope [Limpens et al., 2019], therefore have only one investment and operation period. In the second category, we find multi-period planning models that optimize the transition of the energy system over a given horizon such as MARKAL/TIMES [Krzemien, 2013], OSeMOSYS [Howells et al., 2011], ETEM [Babonneau et al., 2017], MESSAGE [Sullivan et al., 2013] and SMART [Powell et al., 2012]. In general, energy planning models are large-scale optimization models and integrate all energy sectors that are mutually dependent on each other, such as electricity, heating or transportation. In addition, these models combine different time scales in order to model both short-term operations and long-term investment decisions. Energy planning solutions are often based on deterministic demand forecasts, fuel prices, technological efficiencies, resource availability, etc., assuming that all input parameters are known over the entire planning horizon. However, if some of the future forecasts deviate from the estimated values, the planned strategic investment decisions are likely to be sub-optimal. In addition, when formulating a deterministic problem, rare events and special cases are often excluded from optimization.

Given these limitations, stochastic formulations have been proposed to obtain robust solutions that protect the system against future uncertainties. For example, in [Powell et al., 2012], resource costs, renewable

generation and demand growth are considered uncertain in the SMART model. In [Babonneau et al., 2017], the authors deal with uncertainty on resource costs, renewable generation and demand growth in the ETEM model. See [Moret et al., 2020b] for a review of applications.

There are currently many ways to deal with uncertainty in optimization problems, Stochastic Programming (SP) and Robust Optimization (RO) being the most widely used in energy models. Nevertheless, it is well-known that these techniques come with numerical and modeling complexities and, as a consequence, most of the studies in the literature are limited either to two-stage problems and/or simplistic uncertainty models. This paper aims to fill this gap by proposing a realistic modeling of uncertain parameters in a multi-stage optimization problem. It is a continuation of two recent papers [Moret et al., 2020b, Guevara et al., 2020] that we discuss next.

A two-stage mixed-integer linear optimization model, EnergyScope, for the Swiss real energy system has been formulated under these two types of approaches. The deterministic model was first introduced by Moret et al. [2016] and formulated in a complete Robust Optimization framework in Moret et al. [2020b].

The model considers a single investment period and a monthly time scale for operations, which allows for capturing the seasonal storage of hydroelectric dams. As described in [Moret et al., 2017], this model has a large number of uncertain parameters, which were grouped into 21 categories (see Moret et al. [2017], Table 1), such as resources costs, investment costs of technologies, maintenance costs of technologies, lifetimes of technologies, technology efficiencies and end-use energy demands to mention a few. All these uncertain parameters are thus taken into consideration in [Moret et al., 2020b] under a unified robust but static formulation, i.e. no adaptation of the second-stage operation decisions to the revealed uncertainty is possible. More recently, the authors in [Guevara et al., 2020] have formulated EnergyScope as a Two-Stage Distributionally Robust Optimization (DRO) model to generate robust strategic investment decisions that are insensitive to the probability distribution functions of the uncertain parameters, which are considered as 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]. However, in both models mentioned above the dynamics of data process and sequential investment decisions are not considered.

Multistage Stochastic Linear Programming (MSLP) is a modeling framework that allows decisions to be made sequentially under uncertainty (see, for instance, Birge and Louveaux [2011]). In its classic version, the uncertainty is represented by a stochastic process, with a known probability distribution, where the objective is to seek a sequence of decisions (or policy) for each stage that optimizes an objective function of an expected value form. More specifically, a policy is a sequence of functions—one for each stage—that map the observed outcomes up to the current stage into a decision that satisfies the constraints of that stage. Usually, the random data of the processes are assumed to have continuous distribution, so when calculating the implied expected value (multivariate integrals), a standard approach is to discretize the random process by generating a finite number of possible realizations, called *scenarios*. These scenarios can be represented by the corresponding scenario tree that branches at each stage. Therefore, in order to use a dynamic model that provides a more realistic representation of the problem including multiple investment periods, it is necessary to define stochastic processes that represent adequately the evolution of the uncertain parameters of the model.

2.1.1 Modeling uncertainty

Modeling uncertain processes in energy models correctly is a complex task. In forecasting models, it is often necessary to have adequate knowledge of the past behavior of the data in order to be able to forecast what will happen in the future. Time series forecasting methods are well-designed to take into account a spatial-temporal relationship. There are many ways to model time series, such as the Autoregressive model (AR), the Autoregressive Moving Average model (ARIMA) and Moving-Average model (MA), in which a linear function of past observations is used to predict future values [Box et al., 2015]. Some works in energy planning incorporate more complex modeling techniques to have a better representation of uncertainty. For example, in Jin et al. [2011] it is assumed that annual electricity demand and natural gas prices can be

represented as a GBM (Geometric Brownian Motion) since both are modeled with an annual growth rate relative to previous years, and the annual geometric growth rate is uncorrelated in different years. Such an approach however requires large volumes of data that are often not available, and a series of transformations and hypothesis tests to validate that it corresponds to the proposed model [Marathe and Ryan, 2005].

Another way to model uncertainty is through scenario trees. In a scenario tree, each node represents the state of the system in a given time period. The arcs connecting the nodes between successive stages represent the possible transitions from one state at time period t to another state at time period $t + 1$, where each of the states has an associated probability of occurrence of the given uncertain parameter. A path from the root node to the leaf node represents a scenario. There are methods that construct scenario trees given a set of scenarios by approximating an appropriate metric over the probability distribution space [Heitsch and Roemisch, 2009]. Stochastic processes are often discretized to generate scenario trees, which are then used to solve multistage stochastic problems.

Regardless of the many possible ways of constructing a scenario tree, the problem of dimensionality and exponential growth of the problem size is unavoidable especially when the input data involves multiple stochastic processes, thereby making the optimization of such models very difficult even when the number of stages is small. This issue, called “Curse of dimensionality”, is mainly due to the fact that each node must have a unique predecessor, i.e., each node uniquely mapped into the entire history of the process. Although this framework is quite flexible in that it allows for modeling arbitrary processes, in some cases it may not be necessary to consider the entire history of a scenario. For instance, a finite-state Markov chain may be used in such cases, allowing each node to have multiple predecessors in each stage. In a Markov chain it is assumed that the future state evolves probabilistically as a function of the current state. When a real-valued stochastic process satisfies this property, a finite state Markov chain can be obtained by discretization techniques [Tauchen, 1986].

2.1.2 Solutions methods for multi-stage stochastic programming

We now discuss the impact of uncertainty modeling on solution methods for MSLP problems and on their efficiency. As discussed earlier, multistage stochastic programs usually approximate the stochastic process $\{\xi_t\}_{t=1}^T$ with a finite number of scenarios exhibiting a tree structure. This usually leads to large-scale Linear programming (LP) or Mixed-integer linear programming (MILP) reformulations that can be solved by decomposition methods. In particular, energy planning formulations can yield significant computational challenges due to the number of scenarios, the number of stages and/or the presence of integer decision variables. Singh et al. [2009] present a multistage stochastic MILP model for planning capacity expansion of an electricity distribution network in New Zealand. Only the future demand is considered uncertain and is modeled by a scenario tree. The deterministic-equivalent MILP problem is solved using Dantzig-Wolfe decomposition. Another multistage stochastic MILP model is formulated in [Park and Baldick, 2016] to solve the generation expansion problem, where the uncertain parameters of load and wind availability are considered as random variables. A scenario tree is generated (using the Gaussian copula method), and then reduced for improved computation performance when using a rolling-horizon method. Ioannou et al. [2019] formulate the power generation expansion planning as a multistage stochastic LP model in which ten power generation technologies are considered as alternatives for new power plants. The authors propose a hybrid uncertainty modeling approach, in which the setting of scenarios tree (e.g. uncertainty of energy demand and capital cost of renewable energy technologies) is combined with a Monte Carlo simulation method. However, fuel prices are modeled as time-independent random variables and the number of time periods is limited. Liu et al. [2018] propose a multistage stochastic investment planning for a centralized power system model with five generic generation technologies, where the long-term uncertainty (e.g. investment, fuel-cost and demand-growth rate) is explicitly modeled through a scenario tree and the short-term uncertainty (e.g. hourly demand and renewable generation) is modeled by means of time-series approaches. To solve the resulting problem, the authors decompose the model by scenarios through a progressive hedging algorithm to reduce computational time.

As the number of the uncertain parameters realizations and the number of periods in the planning horizon

make exponential growth of the tree inevitable [Shapiro and Nemirovski, 2005], usually leading to numerical difficulties, the papers above are limited either to few uncertainties, few periods or do not consider the whole energy system. In other words, the tree approach is applicable when the number of stages is small, or when combined with scenario reduction techniques (such as backward reduction or forward selection).

To deal with the rapidly growing scenario tree issue, Stochastic Dual Dynamic Programming (SDDP) was introduced first in the context of the hydrothermal scheduling problem by Pereira and Pinto [1991]. This algorithm assumes that the MSLP has relatively complete recourse (i.e., for any feasible solutions at stages $t = 0, \dots, T - 1$, there exists a feasible solution to any realized stage $t + 1$ subproblem with probability one), and that the stochastic process is *stage-wise independent*, i.e., the information revealed at stage t of the random variable ξ_t does not depend on the past information of the variables $\xi_{t-1}, \xi_{t-2}, \dots, \xi_2, \xi_1$ of the previous stages. Although these assumptions are helpful in easing the computational burden for the solution method, thereby making SDDP one of the few algorithms available to solve large-scale problems, they can be considered as very unrealistic. To circumvent such limitations, the SDDP algorithm has been extended to integrate correlation or non-continuous variables and applied in several publications [Lara et al., 2018, 2019, Rebennack, 2014, Zou et al., 2019, Thomé et al., 2019]. SDDP relies on the nested decomposition algorithm of Benders [Birge, 1985] and thus avoids solving all the possible scenarios of the multi-stage scenarios tree.

In the literature, it has been assumed that the uncertainties in generation expansion planning addressed by the SDDP are stagewise independent. Rebennack [2014] solves a generation planning problem for hydro-thermal power systems via Benders decomposition, where the master problem decides investment decisions “here and now” without knowledge of uncertainty, while operational decisions are made under a “wait and see” approach. The underlying uncertainty of hydro inflows is considered in the operational decisions and is solved using a standard SDDP algorithm. Similar problems where the SDDP is combined with a master problem to solve generation and transmission expansion planning are presented in Campodónico et al. [2003] and Thomé et al. [2019]. An extension of SDDP to multistage integer programming models (SDDiP proposed by Zou et al. [2019]) that addresses the long-term planning of electric power infrastructure under multiscale uncertainty is presented by Lara et al. [2019]. In that work it is assumed that the cost of fuels, the availability of renewable generation and the load demand are stage-wise independent in the scenario tree, and the decisions of investments in generation and storage units are integer variables. However, the time dependence governing the original stochastic process is not considered within SDDiP.

In this paper, we present a multi-stage multi-sector strategic energy planning model in which both the end-use demand for energy services and resource costs are considered uncertain. These two sets of uncertain parameters were identified in [Guevara et al., 2020] as the ones with the most significant impact on the optimal investment strategy. The main contribution of our work is a novel methodology to model the time-dependence of these uncertain parameters. Our methodology takes, as its input, ranges defined in the literature for the uncertain parameters (as well as their nominal values), and constructs auto-regressive (AR) models with the property that the random variables corresponding to each period are—conditionally on the value of the previous observation—distributed *uniformly* within those pre-specified ranges. The latter property creates new challenges since the noise of the AR process becomes dependent on the variables of the previous stage; to circumvent this issue, we introduce a coefficient α in the AR construction for controlling the zigzag effects in the evolution of uncertain processes, and show that the resulting process can be represented in the SDDP model with stage-wise independent variables. As it turns out, the coefficient α can be used to control the likelihood of extreme scenarios. To preserve the convexity of the stochastic problem, we discretize the AR models associated with the cost parameters involved in the objective function using Markov chains. The resulting formulation is then solved with an advanced SDDP algorithm available in the literature that handles finite-state Markov chains. Finally, we perform numerical experiments on a Swiss case study to assess the benefit of considering time-dependence in stochastic processes.

The rest of the paper is organized as follows. In Section 2.2, we describe the deterministic formulation of the multi-stage strategic energy planning problem from the literature, which does not take uncertainties into account, and we introduce a stochastic multi-stage formulation. In Section 2.3 we present our methodology for modeling uncertainty based on auto-regressive (AR) models. Section 2.4 discusses the implementation of these AR models within an SDDP framework and how to approximate an AR process by finite state-space

Markov chains (in discrete time). Section 2.5 is dedicated to a numerical application to the Swiss energy system, and in Section 2.6 we provide summary results and conclusions for future research.

2.2 Multi-stage strategic energy planning

2.2.1 A deterministic multi-stage model

In this section, we present the deterministic formulation of the multi-stage strategic energy planning model which is an extension of the one-period model, EnergyScope, first introduced by Moret et al. [2017] and applied to the Swiss energy system. EnergyScope is a multi-sector and multi-energy model that is driven by end-user demand in energy services (electricity, heating and transportation), by the efficiency and cost of technologies (generation and storage) and by the cost and availability of resources (imported and local). In this model, heating demand is split into centralized, decentralized or industrial. The first two heating demands include end-use demand for space heating and hot water and belongs to the category of low-temperature heat. The third is the end-use demand for industrial process heating and belong to the category of high-temperature heat. Transport end-use demand is also divided into passenger mobility (public and private) and freight mobility (rail and road). The model seeks an optimal policy that minimizes investment and operating costs over the entire planning horizon, i.e. 2020-2050. The operating decisions are taken on a monthly basis to represent long-term energy storage through hydroelectric dams, which is relevant for Switzerland. Investment decisions are related to the installation of new technologies, which are affected by strategic decisions. Strategic decisions concern to (i) the proportion of low-temperature heat demand to be dedicated to decentralized heat (district heating network (DHN)) and centralized heat demand, (ii) the proportion of passenger mobility to be dedicated to private and public sector demand, and (iii) the proportion of freight mobility to be dedicated to road and rail freight demand.

A simplified formulation of the model is presented below. A detailed formulation is presented in Appendix C.

$$\begin{aligned} & \min_{\mathbf{C}_t^{INV}, \mathbf{C}_t^{SV}, \mathbf{C}_t^{O\&M}, \mathbf{x}_{t,p}, \\ & \quad \mathbf{y}_{t,p}, \mathbf{z}_{t,r,m}, \mathbf{Sto}_{t,u,l,m}^+, \\ & \quad \mathbf{Sto}_{t,u,l,m}^-, \mathbf{D}_{t,l,m}, \mathbf{Loss}_{t,l,m}, \\ & \quad \mathbf{GWP}_{t,r}^{OP}} \sum_{t \in \mathcal{T}} \tau(t) \left(\mathbf{C}_t^{Inv} - \mathbf{C}_t^{SV} + \mathbf{C}_t^{O\&M} \right) \\ & \text{s.t.} \end{aligned} \tag{m.1}$$

$$\mathbf{C}_t^{Inv} = \sum_{p \in \mathbf{P}} c_{t,p}^{Inv} \cdot \mathbf{y}_{t,p} \quad \forall t \in \mathcal{T}, \tag{m.2}$$

$$\mathbf{C}_t^{O\&M} = \varsigma_t \cdot \left(\sum_{p \in \mathbf{P}} c_{t,p}^{Maint} \cdot \mathbf{x}_{t,p} + \sum_{r \in \mathbf{RES}} \sum_{m \in \mathbf{M}} c_{t,r}^{Op} \cdot \mathbf{z}_{t,r,m} \cdot h_m^{Op} \right) \quad \forall t \in \mathcal{T}, \tag{m.3}$$

$$\mathbf{C}_t^{SV} = \sum_{p \in \mathbf{P}} sv_{t,p} \cdot c_{t,p}^{Inv} \cdot \mathbf{y}_{t,p} \cdot \mathbb{1}_{\{t+lt_p > T+1\}}(t) \quad \forall t \in \mathcal{T}, \tag{m.4}$$

$$\mathbf{x}_{0,p} = res_{0,p}, \quad \forall p \in \mathbf{P}, \tag{m.5}$$

$$\mathbf{x}_{t,p} = \mathbf{x}_{t-1,p} + \mathbf{y}_{t,p} - rest_{t-1,p} + rest_{t,p} - \mathbf{y}_{t-lt_p,p} \cdot \mathbb{1}_{\{t > lt_p\}}(t) \quad \forall p \in \mathbf{P}, \forall t \in \mathcal{T}, \tag{m.6}$$

$$f_{t,p}^{min} \leq \mathbf{y}_{t,p} \leq f_{t,p}^{max} \quad \forall p \in \mathbf{P}, \forall t \in \mathcal{T}, \tag{m.7}$$

$$\bar{f}_p^{min} \leq \mathbf{x}_{t,p} \leq \bar{f}_p^{max} \quad \forall p \in \mathbf{P}, \forall t \in \mathcal{T}, \tag{m.8}$$

$$\mathbf{z}_{t,p,m} \leq \mathbf{x}_{t,p} \cdot k_{p,m} \quad \forall p \in \mathbf{P}, \forall m \in \mathbf{M}, \forall t \in \mathcal{T}, \tag{m.9}$$

$$\sum_{m \in \mathbf{M}} \mathbf{z}_{t,p,m} \cdot h_m \leq \mathbf{x}_{t,p} \cdot \hat{k}_p \sum_{m \in \mathbf{M}} h_m \quad \forall p \in \mathbf{P}, \forall t \in \mathcal{T}, \tag{m.10}$$

$$\sum_{m \in \mathbf{M}} \mathbf{z}_{t,r,m} \cdot h_m \leq avail_r \quad \forall r \in \mathbf{RES}, \forall t \in \mathcal{T}, \tag{m.11}$$

$$\mathbf{z}_{t,u,m} = \mathbf{z}_{t,u,m-1} + h_m \sum_{\substack{l \in \mathbf{L} \\ \eta_{u,l}^+ > 0}} \mathbf{Sto}_{t,u,l,m}^+ \eta_{u,l}^+ - h_m \sum_{\substack{l \in \mathbf{L} \\ \eta_{u,l}^- > 0}} \mathbf{Sto}_{t,u,l,m}^- / \eta_{u,l}^- \quad \forall u \in \mathbf{STO}, \forall m \in \mathbf{M}, \forall t \in \mathcal{T}, \tag{m.12}$$

$$\sum_{i \in \mathbf{P} \cup \mathbf{RES} \setminus \mathbf{STO}} f_{i,l} \cdot \mathbf{z}_{t,i,m} = - \sum_{u \in \mathbf{STO}} (\mathbf{Sto}_{t,u,l,m}^- - \mathbf{Sto}_{t,u,l,m}^+) + \mathbf{D}_{t,l,m} + a_l \cdot \mathbf{Loss}_{t,l,m} \quad \forall l \in \mathbf{L}, \forall m \in \mathbf{M}, \forall t \in \mathcal{T}, \tag{m.13}$$

$$\mathbf{Loss}_{t,l,m} = \sum_{\substack{i \in \mathbf{RES} \cup \mathbf{P} \setminus \mathbf{STO}, f_{i,l} > 0}} f_{i,l} \cdot \mathbf{z}_{t,i,m} \cdot \%loss_l \quad \forall m \in \mathbf{M}, \forall l \in \mathbf{L} \setminus \mathbf{RES}, t \in \mathcal{T}, \tag{m.14}$$

$$\mathbf{GWP}_{t,r}^{Op} = \sum_{m \in \mathbf{M}} gwp_{r,m}^{Op} \cdot \mathbf{z}_{t,r,m} \cdot h_m \quad \forall r \in \mathbf{RES}, \forall t \in \mathcal{T} \tag{m.15}$$

The objective function of the energy problem minimizes the total discounted system cost (i.e., investment and operation costs) minus the salvage value of the residual life of installed technologies at the end of the planning horizon (m.1), where $\tau(t)$ is the discount factor at stage t , with $t \in \mathcal{T} = \{1, \dots, T\}$. The investment cost \mathbf{C}_t^{Inv} in period $t \in \mathcal{T}$ is the sum of the product of the investment cost of technology $p \in \mathbf{P}$ with the new capability of its corresponding technology (m.2). The O&M cost $\mathbf{C}_t^{O\&M}$ in period t is equal to the amount of the system's maintenance and operating costs multiplied by the annualization factor ς_t , where $r \in \mathbf{RES}$

represents the energy resource and $m \in \mathbf{M}$ represents the months of the year (m.3). The discounted salvage \mathbf{C}_t^{SV} in period t assigns a non-zero value to the technologies still available at the end of the planning horizon (m.4).

The set of restrictions (m.5)-(m.6) refers to the initial design of the capacity and its subsequent expansion, which considers the new capacity installed and the decommissioning plan. The available capacity $\mathbf{x}_{t,p}$, as shown in (m.6), is continuously updated by balancing the accumulated capacity $\mathbf{x}_{t-1,p}$ in period $t-1$, the new capacity installed $\mathbf{y}_{t,p}$, the elimination of unused capacity $rest_{t,p} - rest_{t-1,p}$ at the beginning of the period t and the retirement capacity of technologies that were installed for optimization but have reached their useful life. When $t = 0$, the initial capacity is the existing infrastructure at that moment as shown in (m.5). Once a technology is installed, it can be operated until its end-of-life. The indicator function $\mathbb{1}_{\{condition\}}(t)$ takes the value 1 when the *condition* is met and the value 0 when it is not met. Constraint (m.7) sets the upper and lower bounds to the new capacity installed for each period t and technology p . For the available capacity, we also use the upper and lower bounds over the entire planning horizon as shown (m.8). The decision variable $\mathbf{z}_{t,\cdot,m}$ determines how resources, the unit's storage and technologies are used in each period. Constraints (m.9)-(m.10) link the total capacity available of a technology to its actual use in each month and period via two capacity factors: a capacity factor for each month $k_{p,m}$ depending on resource availability (e.g. renewable) and a yearly capacity factor \hat{k}_p accounting for technology downtime and maintenance. The constraint (m.11), is used to limit the total use of resources by their availability during the period *avail*. Constraint (m.12) is the balancing equation for storage units $u \in \mathbf{STO}$. In this constraint when $m = 1$ then the initial condition $\mathbf{z}_{t,u,0}$ is equal to $\mathbf{z}_{t,u,12}$, $\forall t \in \mathcal{T}$ and $\forall u \in \mathbf{STO}$.

The variables $(\mathbf{Sto}_{t,u,l,m}^+, \mathbf{Sto}_{t,u,l,m}^-)$ determine the inputs to or outputs from the storage units. Layers \mathbf{L} are defined as the elements in the system that need to be balanced in each month, such as resources and end-use demand. For example, electricity imported or produced on the system (as well as end-use demand for electricity) can be stored in hydroelectric dams or used as inputs to other energy conversion technologies (such as heat pumps) to meet the end-use demand for electricity. Constraint (m.13) expresses the balance for each layer: all outputs from resources and technologies (including storage) are used to satisfy the end-uses-demand $\mathbf{D}_{t,l,m}$ or as inputs to other resources and technologies. The matrix $f \in \mathcal{M}_{|\mathbf{PURES} \setminus \mathbf{STO}| \times |\mathbf{L}|}(\mathbb{R})$ defines for all technologies and resources outputs to (positive) and inputs from (negative) layers. Losses $\mathbf{Loss}_{t,l,m}$ are considered for the electricity grid and for the DHN (District heating network). Constraint (m.14) calculates the amount of electricity that is lost from both produced and imported electricity in the corresponding layers (a_l is equal to one for these corresponding layers). The total emissions of resources $\mathbf{GWP}_{t,m}^{Op}$ is calculated as the sum of the use $\mathbf{z}_{t,r,m}$ over different months multiplied by the month duration h_m and the emissions of the resource $gwp_{r,m}^{Op}$ (m.15). Note that in our model there are no emission targets, but this expression allows us to estimate the amount of emissions produced by the system in each period t .

2.2.2 A stochastic multi-stage model

The problems of real world energy systems are usually of a dynamic nature, where uncertain parameters are revealed sequentially and decisions must be adjusted to the recent realization. Multistage Stochastic Linear Programming captures well the nature of sequential decision-making under uncertainty, where the uncertainty is modeled by the overall stochastic process. Let T be a time horizon that spans several years, then consider the following T -stage stochastic linear programming model, written in a recursive form:

$$\begin{aligned} & \min_{(y_1(\xi_1), z_0(\xi_1)) \in \mathcal{X}_1(y_0(\xi_0), \xi_1)} f_1(y_1(\xi_1), z_0(\xi_1), \xi_1) + \mathbb{E}_{\xi_2 | \xi_1} \left[\min_{(y_2(\xi_{[2]}), z_1(\xi_{[2]})) \in \mathcal{X}_2(y_1(\xi_1), \xi_2)} f_2(y_2(\xi_{[2]}), z_1(\xi_{[2]}), \xi_2) + \right. \\ & \quad \left. \mathbb{E}_{\xi_3 | \xi_{[2]}} \left[\dots + \mathbb{E}_{\xi_T | \xi_{[T-1]}} \left[\min_{(y_T(\xi_{[T]}), z_{T-1}(\xi_{[T]})) \in \mathcal{X}_T(y_{T-1}(\xi_{[T-1]}), \xi_T)} f_T(y_T(\xi_{[T]}), z_{T-1}(\xi_{[T]}), \xi_T) \right] \right] \right]. \quad (2.2) \end{aligned}$$

In the above formulation, ξ_t represents a random vector containing some of (or all) the stochastic parameters at stage t , $\xi_{[t]} = (\xi_1, \xi_2, \dots, \xi_t)$ denotes the history of the stochastic process up to time t and $\mathbb{E}_{\xi_t | \xi_{[t-1]}}$ denotes the conditional expectation operation in stage t with respect to ξ_t given the information $\xi_{[t-1]}$ of stage $t-1$.

We distinguish two sets of decision variables in each stage, namely, the state variable, denoted by $y_t(\xi_{[t]}) = [\mathbf{y}_{t,p}, \mathbf{x}_{t,p}]$, which links successive stages (e.g. capacity installed, available capacity), and the local or stage variable, denoted by $z_t(\xi_{[t]}) = [\mathbf{C}_t^{INV}, \mathbf{C}_t^{O\&M}, \mathbf{C}_t^{SV}, \mathbf{z}_{t,..m}, \mathbf{Sto}^+_{t,u,l,m}, \mathbf{Sto}^-_{t,u,l,m}, \mathbf{D}_{t,l,m}, \mathbf{Loss}_{t,l,m}, \mathbf{GWP}_{t,p}^{Op}]$, which is only contained in the subproblem at stage t . To ease the notation, we will henceforth use y_t and z_t as a shorthand for $y_t(\xi_{[t]})$ and $z_t(\xi_{[t]})$ respectively. The function f_t and the set \mathcal{X}_t denote the objective function and set of feasible decisions associated with stage t , respectively. The function f_1 and the set \mathcal{X}_1 are deterministic (i.e. ξ_1 is a known value). y_0 represent the initial state of the system. In addition, we assume that the objective function f_t is linear of the form $f_t(y_t, z_{t-1}, \xi_t) = c_t^\top(\xi_t)y_t + d_t^\top(\xi_t)z_{t-1}$ which represent the cost function associated with stage t , and the \mathcal{X}_t set is of the form

$$\mathcal{X}_t(y_{t-1}, \xi_t) := \{(y_t, z_{t-1}) : B_t(\xi_t)y_{t-1} + A_t(\xi_t)y_t + C_t(\xi_t)z_{t-1} = b_t(\xi_t), y_t \geq 0, z_{t-1} \geq 0\}, \forall t = 2, \dots, T \quad (2.3)$$

where $\mathcal{X}_t(\cdot, \cdot)$ describes the feasible set for decisions to be made at stage t . We introduce two different time scales, a long-term time scale for investment decisions and a short-term time scale for operations decisions. As an example, the dynamics between investment and operation decisions using these two time scales are illustrated in the Figure 2.1.

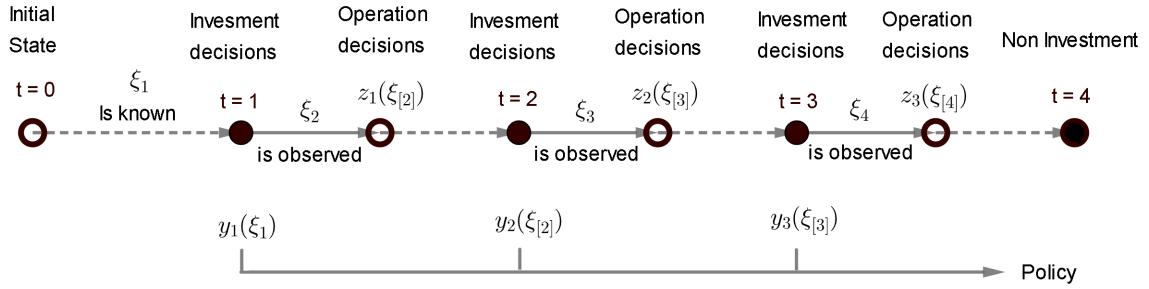


Figure 2.1: A small example of four stages with three periods of investment, followed by operating decisions conditioned on revealed uncertainty, which shows the dynamic process of the model.

As shown in the figure above, $t = 0$ represents the initial state of the system, that is, the total available capacity, which is considered as input data. Uncertainty is sequentially revealed over time. Investment decisions are made at the beginning of each stage before the uncertainties are revealed. On the other hand, operational decisions are made during the next stage after the uncertainties are realized. Here $\{y_t, z_t\}_{t=1}^T$ describe a policy, a solution to the MSLP. We assume relatively complete recourse for any policy y_t, z_t . $\forall t = 2, \dots, T$. Such an assumption is reasonable in our context since it is equivalent to assuming that one can always import energy (albeit at a high cost) to satisfy demand.

2.3 Uncertainty modelling through auto-regressive processes

2.3.1 The goals of the model

Multi-stage models under uncertainty imply a sequence of decisions to be made at each period, considering the state of the system and some uncertain parameters that have a serial correlation. Since in our model (2.2), the investment and operational decisions are made sequentially based on the information available on the random variables up to stage t and the state variables, we need a model that describes the evolution of stochastic processes in each stage. [Moret et al., 2017] only defines the ranges of variation of the uncertain parameters, but does not describe how the process evolves over the planning horizon. Assuming that the uncertain parameters are independent across stages may be unrealistic. Therefore, to describe the evolution of the processes, we use the ranges of variation of the uncertain parameters reported by [Moret et al., 2017]

to construct several auto-regressive processes. In this section, we assume that the end-use demand for energy services and the costs of all resources are time-dependent stochastic processes.

Suppose that a stochastic process ξ_t is driven by an auto-regressive process of order 1, denoted by AR(1), and has the following representation

$$\xi_{t+1} = a_t \cdot \xi_t + \epsilon_{t+1}, \quad \forall t = 0, \dots, T-1, \quad (2.4)$$

where the AR coefficient a_t and the standard deviation of the innovation (or error term) ϵ_{t+1} are both allowed to depend on time t ; ϵ_t is independent across t . The initial realization ξ_0 may be deterministic. This AR model tries to explain the future value of ξ_{t+1} using the response variable of the previous stage ξ_t as a predictor. Forecasting models such as the one in (2.4) are often used to model inflows in multi-stage models for hydroelectric planning (see, e.g., [Shapiro et al. \[2013\]](#), [Löhndorf and Shapiro \[2019\]](#)).

In this study, we want to use AR models to estimate the future energy demand and resource costs, but due to the lack of information from historical data, approaches such as maximum likelihood estimation or least-squares estimation cannot be applied directly to estimate each parameter a_t , and therefore we cannot know the distributions of the error term ϵ_t . To estimate the a_t parameters and the distribution of the errors ϵ_t , we use the ranges of variation and the nominal values of the uncertain parameters found in [Moret et al. \[2017\]](#). A range of variation is defined by assigning a lower bound R_t^- and an upper bound R_t^+ to the process ξ_t at stage t , and the nominal value μ_t is the most probable realization corresponding to the range $[R_t^-, R_t^+]$. Nominal values, in addition to providing the most probable value of the range at stage t , also indicate that the data is distributed 50% above and below it. When a range is symmetric (mean equals nominal value), we construct the corresponding AR process so that it has uniform distribution within each range $[R_t^-, R_t^+]$ (conditionally on the value of the process in stage $t-1$), while for an uncertain parameter with an asymmetric range (e.g., cost of resources), we use a stochastic process ξ'_t defined as:

$$\xi'_t = \begin{cases} \mu_t \cdot (1 + \xi_t) & \text{if } \xi_t \geq 0, \\ \mu_t / (1 - \xi_t) & \text{otherwise,} \end{cases}$$

where $\{\xi_t\}$ is the process drawn from (2.4).

In summary, in order to be able to build a continuous AR stochastic process of the form (2.4), we assume that the ranges of variation $[R_t^-, R_t^+]$ (or $[-R'_t, R'_t]$), nominal values μ_t of the uncertain parameters and the starting point of the processes at stage $t=0$ are given. Figure 2.2a represents the input data necessary to estimate the parameters of the AR model. We want to construct a sequence of $\{\xi_t\}_{t=0}^T$ (or $\{\xi'_t\}_{t=0}^T$) using the following strategy: when the stochastic process realization is near the upper bound of the range in step t , then in the next stage, the process realization cannot be near the lower bound of the range; similarly, when the stochastic process realization is near the lower bound in stage t , then in the next stage, the process realization cannot be near the upper bound. This proposal tries to avoid the zigzag shape that a process presents when it is considered stage-wise independent. Figure 2.2b shows the ranges of variation and the nominal values of each process $\{\xi_t\}_{t=1}^T$ through black and blue dots, respectively. An arbitrary realization of the process $\{\xi_t\}_{t=0}^T$ is represented by red dots.

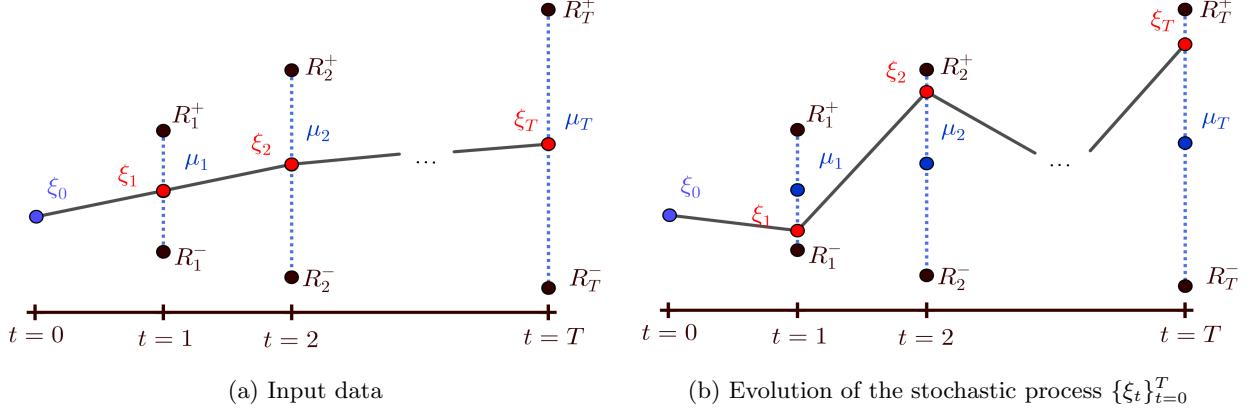


Figure 2.2: In this figure $\mu_1, \mu_2, \dots, \mu_T$ are the nominal values, R_t^- and R_t^+ for $t = 1, \dots, T$ are the lower and upper bounds of the ξ_t parameter, respectively. ξ_0 is deterministic and represents the initial realization of the process. The figure on the left represents the evolution of the processes in the deterministic case, while the figure on the right, a possible realization of the stochastic process is shown.

2.3.2 The construction of the AR model

We start by defining a model that describes the evolution of the process without considering the error term ϵ_t . Then we introduce randomness in the initial model to finally derive an AR process with which the degree of dependency can be controlled, thereby avoiding the zigzag shape in the stochastic process realizations.

2.3.2.1 The deterministic process

Without loss of generality suppose that the range $[R_t^-, R_t^+]$ represents all symmetric ranges (including $[-R_t', R_t']$). Since ξ_0 is a given constant, we define ξ_1 as the midpoint of $[R_1^-, R_1^+]$. The idea of our construction is to transport each realization of ξ_t , $t = 1, \dots, T - 1$ to an equivalent place within the range of the next stage while keeping proportions. More specifically, let $t \geq 1$ be the current stage and let $Z(\xi_t)$ be the equivalent place of ξ_t within the range of the next stage. We impose that the ratio between $(R_t^+ - R_t^-)$ and $(R_{t+1}^+ - R_{t+1}^-)$ be proportional to the ratio between $(\xi_t - R_t^-)$ and $(Z(\xi_t) - R_{t+1}^-)$, i.e.,

$$\frac{R_t^+ - R_t^-}{R_{t+1}^+ - R_{t+1}^-} = \frac{\xi_t - R_t^-}{Z(\xi_t) - R_{t+1}^-}. \quad (2.5)$$

Thus, we obtain the model without the random factor:

$$\xi_{t+1} := Z(\xi_t) := R_{t+1}^- + \frac{R_{t+1}^+ - R_{t+1}^-}{R_t^+ - R_t^-} (\xi_t - R_t^-). \quad (2.6)$$

It is easy to see that in the case of deterministic processes each element ξ_t , $t = 1, \dots, T$ is equal to the midpoint of the range corresponding to stage t , which is also the nominal value μ_t (see Figure 2.2a).

2.3.2.2 The stochastic process

Our goal now is to add a random factor to the equation (2.6). To do so, we need to construct a random variable ω_t that allows us to move $Z(\xi_t)$ without leaving the range corresponding to the stage t . For that purpose, we introduce for $t \geq 1$ a *variability-controlling* parameter $\alpha \in [0, 1]$ that limits the support of $Z(\xi_t) + \omega_t$ to a proportion α of the range $[R_{t+1}^-, R_{t+1}^+]$. More specifically, we define ξ_1 as a uniform random variable on

$[R_1^-, R_1^+]$; then, for $t \geq 1$, given the value of $Z(\xi_t)$ computed in (2.6), we add a random perturbation ω_t to $Z(\xi_t)$ as follows:

$$\xi_{t+1} := Z(\xi_t) + \omega_t, \text{ with } \omega_t \sim \text{Uniform}(-\alpha_t \cdot \rho_t, \alpha_t \cdot (1 - \rho_t)), \quad (2.7)$$

where $\alpha_t := \alpha(R_{t+1}^+ - R_{t+1}^-)$, and

$$\rho_t := \frac{Z(\xi_t) - R_{t+1}^-}{R_{t+1}^+ - R_{t+1}^-}.$$

It follows that

$$\xi_{t+1} \sim \text{Uniform}(\alpha R_{t+1}^- + (1 - \alpha)Z(\xi_t), \alpha R_{t+1}^+ + (1 - \alpha)Z(\xi_t)), \quad (2.8)$$

so we see that, for any given value of ξ_t , the support ξ_{t+1} is contained in $[R_{t+1}^-, R_{t+1}^+]$ and has length $\alpha(R_{t+1}^+ - R_{t+1}^-)$ as desired. Figure 2.3 illustrates the basic idea for the construction of $\{\xi_t\}$, with $\alpha = 0.75$ being used in this example.

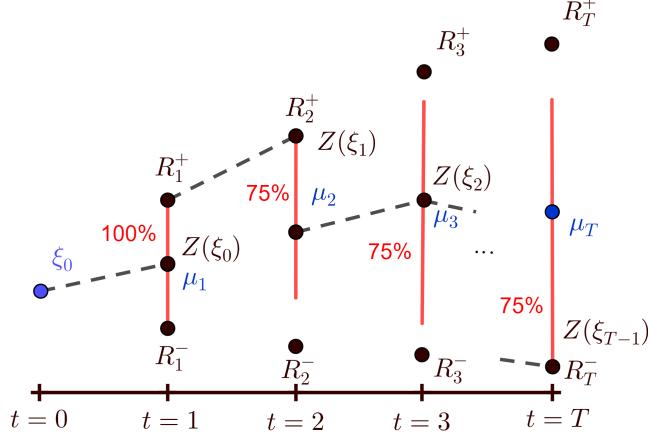


Figure 2.3: The figure represents the possible relations of $\{\xi_t\}$ in red lines. The first stage covers 100% of the range, while the following stages always cover 75% of the range. It also shows a realization of process.

Equations (2.6)-(2.8) make the role of α clear. It allows the decision maker to control the variability of the underlying stochastic processes based on his/her beliefs, but at the same time α has a number of properties, listed below:

- (i) If $\alpha = 0$, then for all $t \geq 2$ we have that ξ_t is a linear transformation of the uniform random variable ξ_1 with a positive multiplicative constant; thus, ξ_t is (unconditionally) uniformly distributed, has correlation one with ξ_1 , and by applying an inductive argument in (2.6) we see that the support of ξ_t is the range $[R_t^-, R_t^+]$.
- (ii) If $\alpha = 1$, the process $\{\xi_t\}$ is independent across periods and each ξ_t is uniformly distributed on $[R_t^-, R_t^+]$.
- (iii) If $\alpha \in (0, 1)$, then the unconditional distribution of ξ_t is symmetric but not uniformly distributed, as it is defined as the sum of uniformly distributed random variables.
- (iv) As α decreases, the probability that either all ξ_t , $t = 1, \dots, T$ are close to their maximum values or all ξ_t are close to their minimum values, increases.

Property (iv) above—which is a direct consequence of the fact that the correlation among the ξ_t s increase as α decreases—is particularly important in the context of strategic energy planning, since the scenarios with high investment and operations costs and high demand clearly are the scenarios with the highest system cost; conversely, those scenarios with low costs and low demand clearly are the scenarios with the lowest system cost. Thus, by choosing lower values of α , the decision-maker can make sure that scenarios that are close to the ones with extreme costs are generated more often in the course of the algorithm. To illustrate property (iv) above, Figure 2.4 depicts 10,000 realizations of a process $\{\xi_t\}_{t=1}^3$ defined with $[R_t^-, R_t^+] = [3, 4], [5, 6], [7, 8]$, $t = 1, 2, 3$ respectively, for $\alpha = 0.25$ and $\alpha = 0.75$. We see that in the case of lower α the correlation among the ξ_t is high and indeed the extreme scenarios are generated more often.

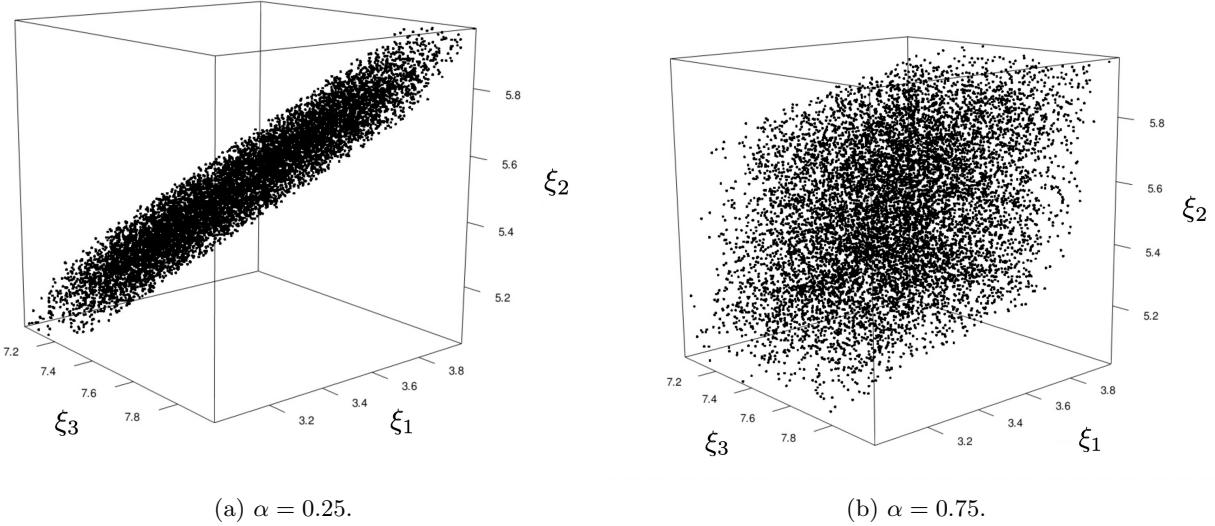


Figure 2.4: 10,000 realizations of $\{\xi_t\}$, $t = 1, 2, 3$ for two different values of α .

Finally, we write the AR model in the form (2.4). For that, we need to rewrite (2.7), since in the latter formulation the random “noise” term ω_t depends on the previous values of the process.

One way to avoid this is by using the inverse transform method according to which we can write

$$\text{Uniform}(a, b) = a + (b - a) \cdot U, \text{ with } U \sim \text{Uniform}(0, 1).$$

It follows that

$$\xi_{t+1} = Z(\xi_t) - \alpha_t \cdot \left(\frac{Z(\xi_t) - R_{t+1}^-}{R_{t+1}^+ - R_{t+1}^-} \right) + \alpha_t \cdot U, \text{ with } U \sim \text{Uniform}(0, 1). \quad (2.9)$$

Then, by substituting $Z(\xi_t)$ from (2.6) into (2.9) and doing some simplifications in the latter equation, we obtain the following expression

$$\begin{aligned} \xi_{t+1} &= \left(\frac{R_{t+1}^+ - R_{t+1}^- - \alpha_t}{R_t^+ - R_t^-} \right) \xi_t + \left(-\frac{R_{t+1}^+ - R_{t+1}^- - \alpha_t}{R_t^+ - R_t^-} \right) R_t^- + R_{t+1}^- + \alpha_t \cdot U \\ &= \underbrace{\left(\frac{(1-\alpha)\Delta_{t+1}}{\Delta_t} \right)}_{a_t} \xi_t + \underbrace{R_{t+1}^- - \frac{(1-\alpha)\Delta_{t+1}}{\Delta_t} R_t^- + \alpha \Delta_{t+1} \cdot U}_{\epsilon_{t+1}}, \end{aligned} \quad (2.10)$$

where $\Delta_t := R_t^+ - R_t^-$.

2.4 SDDP and Markov Chain

We discuss now solution methods for the multi-stage stochastic model described in Section 2.2, where the uncertainty components are modeled using the techniques presented in Section 2.3. As discussed in Section 2.1, we use the Stochastic Dual Dynamic Programming (SDDP) algorithm for that purpose. The SDDP algorithm performs a series of iterations, each consisting of one step forward and one step backward through the stages. It decomposes the problem into stages to construct an approximately optimal policy using Benders' cuts. In addition, it incorporates Monte Carlo sampling to approximate the original process using only a subset of the scenarios at each iteration, which makes it a method capable of solving large-scale problems. When the stochastic process is stage-wise independent the expected future cost function is independent of the realization of the process, and therefore the cuts that are generated for a subproblem are also valid cuts for other subproblems in the same stage. This property of sharing cuts is fundamental for an efficient implementation of the SDDP. Despite its advantages, an important limitation of the SDDP is related to the modeling of the dynamics of the state variables.

When the stage-wise independent condition is relaxed, then the expected future cost functions $\mathcal{Q}_t(y_t, \xi_{t+1})$ also depend on past realizations of stochastic process. The most common way to preserve dependence within the SDDP is to assume that the stochastic process is modeled as an AR process of order 1 or AR(1). An AR(1) is incorporated into the SDDP model by defining the stochastic process as a state variable, the noise or innovation ξ_t is considered stage-wise independent and the recursive equation (as given by equation (2.4)) is added to the model. Note however that, if the new state variable multiplies any other decision variable, bilinear terms destroy the convexity of the future cost function. Therefore, to maintain the mathematical properties that make the method efficient, only the processes that are on the right-hand side of the constraints can be expressed as an affine function of the errors. The price of such reformulation is an increase in the number of state variables.

Another way to preserve the dependence in the SDDP without destroying the convexity of the problem is to model the stochastic process as a finite-state Markov chain [Löhndorf and Shapiro, 2019, Philpott and de Matos, 2012]. A stochastic process $\{\xi_t\}$ is called Markovian if the conditional distribution of ξ_t given $\xi_{[t-1]}$ is the same as that of ξ_t given ξ_{t-1} for $t = 1, \dots, T$. When a data process is assumed to be Markovian, a future cost function must be enumerated for each value taken by the state variable, which increases the dimensionality of the problem. It is the main computational disadvantage of this approach. Also, if the process is multi-dimensional, the number of Markov states needed to obtain a good enough solution becomes prohibitive.

In the previous section, we constructed a general AR(1) process to model the energy demands and resource costs. Given the ways in which the uncertainty of stochastic processes can be included in the SDDP, we use the state-variable approach for the uncertainty in the demand, which appears only in the right-hand side of the constraints; for the uncertainty in costs, we apply the aforementioned strategy to approximate the AR(1) process of the costs by a Markov chain to maintain the convexity of the model. Therefore, in the next section, we explain how to construct Markov Chains with finite state space from AR(1) processes. As this process is multidimensional, we reduced the dimensionality of the process by choosing two uncertain parameters, specifically the cost of fossil fuel (all fossil fuels costs are assumed to be correlated) and electricity, which turn out to be the parameters with the most significant impact on investment decisions [Guevara et al., 2020]. Furthermore, we assume that the two stochastic processes are independent of each other, which allows us to perform the discretization of each one separately.

2.4.1 Finite state Markov-chain approximations

It is well known that Markov chains can be used to approximate AR processes of order 1. When an AR(1) process is stationary, this means that the transition probability distribution is constant for all time instants. Different methods have been proposed to discretize a stationary process AR(1). Among them, we can mention Tauchen [1986], Tauchen and Hussey [1991], Kopecky and Suen [2010] y Rouwenhorst [1995]. An extension of the Tauchen method (Tauchen [1986]) to a non-stationary process AR(1) was proposed by Fella et al. [2019],

for a problem class where the variance increases with time and the persistence component is non-stationary. The main difference between the extension and its stationary counterpart is that the range of the equidistant state space can vary over time and therefore so do the transition probabilities.

Fella et al. [2019] consider an AR(1) process of the following form

$$\eta_t = \rho_t \eta_{t-1} + \epsilon_t, \quad \epsilon_t \sim \text{Normal}(0, \sigma_{\epsilon t}), \quad (2.11)$$

where the standard deviation of the innovation $\sigma_{\epsilon t}$ and the AR coefficient ρ_t are both allowed to depend on time t , and where the unconditional standard deviation of η_t is σ_t . In addition, it is not required that $|\rho_t| < 1$ (as in the stationary case). Then, they approximate the non-stationary AR(1) process by a Markov chain with N states in each stage t , but with a time-dependent state space $\Upsilon_t^N = \{\bar{\eta}_t^1, \dots, \bar{\eta}_t^N\}$ (uniformly-spaced) and transition matrix Π_t^N . The time-varying grid points are defined as

$$\bar{\eta}_t^N = -\bar{\eta}_t^1 = \Omega \sigma_t, \quad (2.12)$$

where Ω is a positive constant. The step size $h_t = 2\Omega\sigma_t/(N-1)$ and the transition probabilities for any $i, j = 1, \dots, N$ are

$$\pi_t^{ij} = \begin{cases} \Phi\left(\frac{\bar{\eta}_t^j - \rho \bar{\eta}_{t-1}^i + h_t/2}{\sigma_{\epsilon t}}\right) & \text{if } j = 1, \\ 1 - \Phi\left(\frac{\bar{\eta}_t^j - \rho \bar{\eta}_{t-1}^i - h_t/2}{\sigma_{\epsilon t}}\right) & \text{if } j = N, \\ \Phi\left(\frac{\bar{\eta}_t^j - \rho \bar{\eta}_{t-1}^i + h_t/2}{\sigma_{\epsilon t}}\right) - \Phi\left(\frac{\bar{\eta}_t^j - \rho \bar{\eta}_{t-1}^i - h_t/2}{\sigma_{\epsilon t}}\right) & \text{otherwise.} \end{cases}$$

where Φ denotes the cumulative distribution function for the standard normal distribution. Our goal is to use this method to discretize the AR(1) processes constructed in Section 2.3 to model the evolution of costs; however, since the AR(1) processes in (2.10) involve uniform distributions instead of normal ones, we must adapt this method to our setting, as we explain below.

Discrete approximations of AR(1) processes

In this section, we explain how to adapt the Fella et al. [2019] method to discretize a one-dimensional AR process, where innovations are independent and identically distributed non-Gaussian random variables.

Let d_t be the process that describes the costs of natural gas (or electricity). Before discretizing d_t , it is necessary to specify an N -state Markov chain in each stage t . Let \hat{d}_t be the discrete-valued process that approximates d_t in each stage t . Let $\Upsilon_t^N = \{\bar{d}_t^1, \bar{d}_t^2, \dots, \bar{d}_t^N\}$ be the finite set of possible realizations of \hat{d}_t in time t , where $\bar{d}_t^1 < \bar{d}_t^2 < \dots < \bar{d}_t^N$. To determine the time-varying points of the grid, we set \bar{d}_t^N and \bar{d}_t^1 to be equal to the maximum and minimum values of the range $[-R'_t, R'_t]$ associated with the process ξ_t . The other values are calculated using the formula

$$\bar{d}_t^i = -R'_t + (i-1) \cdot s_t \quad (2.13)$$

where $s_t = (2R'_t)/(N-1)$. It is the first modification we make to the method described in Fella et al. [2019] since we rely directly on the range of the parameter. Note that for any t , $\bar{d}_t^1 = -R'_t$ and $\bar{d}_t^N = R'_t$. Another important change in the method is the assumption of the uniform distribution instead of the normal distribution to describe the error distribution ϵ_t . However, the change in the distribution causes the original step length h_t to lie outside the domain of the distribution at extreme points. To fix this problem, we choose a step length that constructs intervals of equal size but within the domain of the chosen distribution.

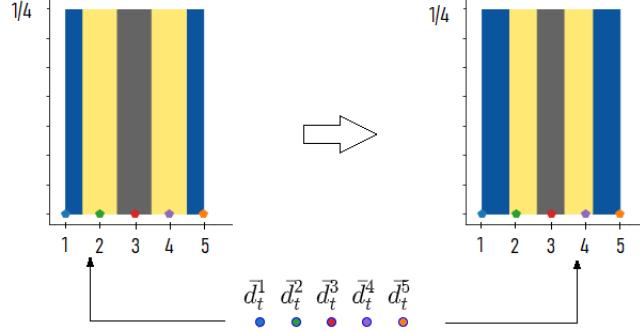


Figure 2.5: An example with a 5-state Markov chain located at the base of the uniform distribution. The uniform distribution is represented with different colors. The colors represent the intervals formed with the selected step length.

Figure 2.5 shows two ways of assigning probabilities to a 5-state Markov chain using different step lengths h_t . Both plots represent the density of the uniform distribution on the interval $[1,5]$. Each state variable d_t^i is mapped to the support of ϵ_t and is represented with the same initial color but with a different shape. Basically, the discretization of a process consists of assigning a probability to \hat{d}_t of falling in the interval $(\bar{d}_t^j - h_t, \bar{d}_t^j + h_t)$ conditionally on $\hat{d}_{t-1} = \bar{d}_{t-1}^i$. In addition, the colored bars represent probabilities formed by the intervals. Note that if the step length $h_t = \frac{1}{2}|\bar{d}_t^N - \bar{d}_t^1|/(N-1)$ for $N \leq 5$, then the method tends to give more probability to the central values than to the extreme values. For example, in the figure on the left side, the probability that \hat{d}_t falls in the interval $(\bar{d}_t^4 - h_t, \bar{d}_t^4 + h_t)$ is much higher than the probability that \hat{d}_t falls in the interval $(\bar{d}_t^5 - h_t, \bar{d}_t^5 + h_t)$ since half of the interval is out of distribution support. But by redistributing intervals of equal size within the domain of the uniform distribution, now the probability of falling into the extremes will be slightly higher. The figure on the right also shows that the step length at the extreme states of the Markov chain is the same, whereas the step lengths around the other Markov states are different.

Given that the step length is not constant for all t and that the parameters of the uniform distribution depend on the time t , we denote \mathcal{G}_t as the cumulative distribution function of ϵ_t in time t . So for any two states i and j , the associated transition probability is

$$\pi_t^{ij} = \begin{cases} \mathcal{G}_t(\bar{d}_t^j - a_t \bar{d}_{t-1}^i + h_t) & \text{if } j = 1, \\ 1 - \mathcal{G}_t(\bar{d}_t^j - a_t \bar{d}_{t-1}^i - h_t) & \text{if } j = N, \\ \mathcal{G}_t(\bar{d}_t^j - a_t \bar{d}_{t-1}^i + h_t^{j+}) - \mathcal{G}_t(\bar{d}_t^j - a_t \bar{d}_{t-1}^i - h_t^{j-}) & \text{otherwise,} \end{cases} \quad (2.14)$$

where $h_t = (d_t^N - d_t^1)/N$, $h_t^{j+} = |(j-1) \cdot (d_t^2 - d_t^1) - j \cdot h_t|$, and $h_t^{j-} = |(j-1) \cdot (d_t^2 - d_t^1) - (j-1) \cdot h_t|$.

We can simplify further the expressions in (2.14) by observing that, as seen in (2.10), we have

$$\epsilon_t \sim R_t^- - \frac{(1-\alpha)\Delta_t}{\Delta_{t-1}} R_{t-1}^- + \alpha \Delta_t \cdot \text{Uniform}(0, 1).$$

By using the fact that (with $U \sim \text{Uniform}(0, 1)$ and $b > 0$)

$$P(a + b \cdot U \leq z) = P\left(U \leq \frac{z-a}{b}\right) = \begin{cases} 0 & \text{if } z < a \\ \frac{z-a}{b} & \text{if } a \leq z < a+b \\ 1 & \text{if } z > a+b, \end{cases}$$

it follows that we can write

$$\mathcal{G}_t(z) = \begin{cases} 0 & \text{if } z < R_t^- - \left(\frac{(1-\alpha)\Delta_t}{\Delta_{t-1}} \right) R_{t-1}^- \\ \frac{z - R_t^- + \left(\frac{(1-\alpha)\Delta_t}{\Delta_{t-1}} \right) R_{t-1}^-}{\alpha\Delta_t} & \text{if } R_t^- - \left(\frac{(1-\alpha)\Delta_t}{\Delta_{t-1}} \right) R_{t-1}^- \leq z < R_t^- - \left(\frac{(1-\alpha)\Delta_t}{\Delta_{t-1}} \right) R_{t-1}^- + \alpha\Delta_t \\ 1 & \text{if } z > R_t^- - \left(\frac{(1-\alpha)\Delta_t}{\Delta_{t-1}} \right) R_{t-1}^- + \alpha\Delta_t. \end{cases} \quad (2.15)$$

Now, by using the expression for $\mathcal{G}_t(z)$ derived above in (2.14), it is straightforward to calculate the transition probabilities π_t^{ij} for each t .

We have now a complete description of our model and the method used to solve it. We call our approach “SDDP- α ”, to emphasize the importance of the parameter α in the uncertainty modeling, as discussed in Section 2.3. In the next section, we present an application of our approach.

2.5 Application to the Swiss energy System

In this section, we apply the proposed time-dependent stochastic SDDP- α approach to the Swiss energy system in which resource costs and demands are uncertain. The objective is twofold. First, we assess the benefit of considering time dependence in the stochastic processes by comparing the optimal solutions and the out-of-sample performances of the SDDP- α model with those obtained with the classical stage-wise independent SDDP approach, called SDDP-IND. In the latter, one optimizes the model assuming that all processes are stage-independent and uncorrelated. Secondly, we analyze in more detail the sensitivity of the α setting on the SDDP- α solutions. Note that the coefficient $\alpha \in (0, 1)$ is used to control the amplitude of the zigzag effect in the AR(1) processes. When α is small, only modest variations in time of the stochastic processes are allowed, thus leading to limited zigzag effects. An α close to 1 will on the contrary generate strong variations over time of the values of the uncertain parameters. Note that the particular case of $\alpha = 1$ corresponds to the SDDP-IND model. As discussed in Section 2.3, small values of α make the generation of extreme scenarios more likely. In our experiments, we consider three intermediate values for α , i.e., 0.25, 0.5 and 0.75.

The uncertainty on the demands (electricity, lighting, and heating), and on the cost of fossil fuels and electricity is modeled through AR(1) processes using the α coefficient. We assume a correlation one among the different fossil fuel costs (gas, coal and diesel). Then the two AR(1) models for fossil fuels and electricity costs are discretized each one into a three-state Markov Chain at each stage, yielding a total of nine states per stage. Assumptions on uncertain parameters (i.e., nominal, min and max values) and the resulting Markov chains are given in Appendices D and F, respectively. Appendix E illustrates potential trajectories predicted by the AR processes as defined in Section 2.3 for the realization of the stochastic processes.

Existing capacities for generation and end-use technologies have been calculated based on information from [Moret et al., 2016]. The time horizon of the study [2010,2050] is decomposed into five 10-year periods with four investment periods from 2020. For interested reader, all data can be found in Moret et al. [2017]¹ and on the website <https://www.energyscope.ch/>.

2.5.1 Implementation details

All stochastic models are written in Julia 1.4.0 and solved with the open-source library SDDP.jl [Dowson and Kapelevich, 2020]. This library implements a Stochastic Dual Dynamic Programming algorithm with possible extensions to AR models and Markov Chain models. We use the single-cut version of the algorithm and only one scenario in each forward pass. The sampling method applied to generate the sample space Ω_t for each stage was the Latin Hypercube Sampling (LHS) [McKay et al., 1979], since it has proven to yield more accurate results in the SDDP context [Homem-de Mello et al., 2011] as it covers the uncertain space

¹The model is described in detail in chapter 1 of that thesis, while the data are documented in detail in Appendix A of the same thesis. The code and data are publicly available at <https://github.com/energyscope/EnergyScope/tree/v1.0>

more uniformly than Monte Carlo sampling. To obtain an upper bound estimator (UB) on the models arising from SDDP-IND and SDDP- α models, we used the methodology presented by de Matos et al. [2017]. The formula of the optimality gap for the SDDP algorithm is the one proposed by Ding et al. [2019]. The first policy evaluation is done at iteration 500 and then every 50 iterations until the desired optimality gap is reached.

We solve each instance using a single core of Intel Core i7-8750H CPU 2.20 GHz \times 12 with 32 GB RAM and CPLEX(V12.9) to solve linear programming subproblems. For a fair comparison between SDDP-IND and SDDP- α results, we use the same number of realizations and random vector samples (or scenarios) in all methods.

Our goal is to obtain optimal policies, where a policy is a rule that describes what decisions must be made at each stage for each realization of the stochastic process at that point in time. More specifically, a policy is a sequence of functions $\{\Pi_t\}_{t=1}^T$ yielding, for any state y_t and observation of the random variable ξ_t , a feasible control z_t . Because the implemented methods are an approximation of the real problem, it is necessary to adopt an evaluation procedure to assess the quality of the policies. Therefore, we generated from the original AR(1) processes (resource demands and costs) 30,000 out-of-sample(α) scenarios with $\alpha \in \{0.75, 0.50, 0.25\}$. To evaluate SDDP- α policies in all out-of-sample(α) scenarios, we search for a Markov Chain scenario in some sense closer to a considered scenario of the continuous process, to substitute Markov Chain states for the considered sample. For all simulations, we employ the same scenario sequence for the forward pass.

2.5.2 Numerical results

All stochastic models have been solved with a 2% optimality gap objective. Table 2.1 reports the model characteristics, the final objective function values, the CPU times for the training and the upper bound evaluations and the number of upper bound evaluations performed during the optimization process (first one at iteration 500 and then every 50 iterations).

	SDDP-IND	SDDP-0.75	SDDP-0.50	SDDP-0.25
# of subproblems by stage	40	40	40	40
Objective value (MCHF)	175,478	172,995	171,482	168,753
Upper bound (MCHF)	177,166	174,917	172,952	168,977
Standard deviation (MCHF)	28,926	33,995	41,503	56,561
Optimality gap (%)	1.55	1.81	1.72	1.33
Training-CPU time (sec)	18,721	25,932	59,651	14,503
# of iterations	550	650	1,850	3,250
CPU time for UB evaluations (sec)	1,334	3,019	16,294	11,526
# of UB estimations	2	4	28	56

Table 2.1: Characteristics and detailed numerical results for the SDDP-IND and SDDP- α models.

As more zigzag effect is permitted, i.e., for SDDP-IND and higher values of α , we observe that the total cost of the solution (i.e. the objective value) increases. We also notice that the number of iterations increases when α decreases while the total CPU time does not change monotonically with α . CPU times for training vary from 4 hours for SDDP-0.25 to more than 16 hours for SDDP-0.50.

2.5.2.1 Out-of-sample performances

We now compare the performances of the different stochastic solutions using samples of 30,000 scenarios generated from the original AR processes with $\alpha \in \{0.75, 0.50, 0.25\}$. Figure 2.6 summarizes the cost performances decomposed among its different components, i.e., Investment, Maintenance, Operation and Total costs.

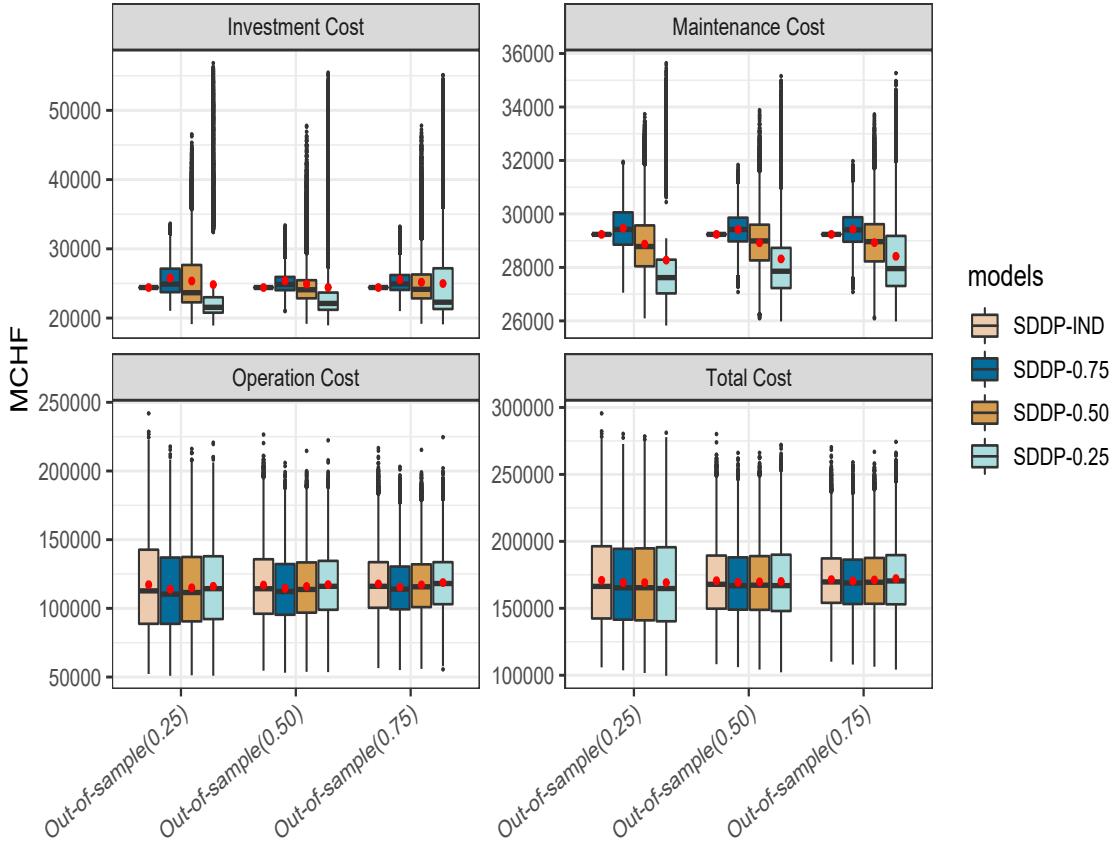


Figure 2.6: Boxplot of total investment, maintenance, operation and total costs. The red dot represents the sample mean. For each out-of-sample simulation, 30000 scenarios were generated from the original continuous AR process with $\alpha \in \{0.75, 0.50, 0.25\}$.

At first sight, the four stochastic solutions lead to similar performance in terms of total costs although the SDDP-IND policy costs seem to have a higher variability with higher extreme total costs. The decomposition of the total cost into its different components is however completely different. It turns out that the SDDP-IND solution does not adapt investment decisions over time to the realizations of the uncertain parameters. Since the model structure assumes temporal independence, the SDDP-IND model results in sequence a two-stage investment/operations problems in each period that are independent. This leads, as it can be seen on Figure 2.1 and then for investment decisions on Figures 2.7 and 2.9, to a unique investment and maintenance cost and a larger variation in operations costs.

SDDP- α solutions produce much more realistic investment strategies that adapt to the realizations of uncertainties. We notice that, when α decreases, the investment and maintenance costs have a greater dispersion but a much lower average value. We will see later that the adaptability of the SDDP- α policy allows for investing in efficient technologies according to the degree of correlation with which they have been trained. The smaller the α , the greater the investment diversification, a very desirable result to reduce the risks associated with uncertainty.

2.5.2.2 Scenario analysis

We now investigate in more detail the role of the α coefficient on its degree of protection against extreme scenarios of parameter realizations. Note that in the sets of 30,000 scenarios generated above with the continuous AR processes, extreme scenarios have a very low probability to appear, i.e., a maximum (or minimum) scenario implies that the sequence of the AR processes takes the maximum (or minimum) value at each stage, which is unlikely. We therefore study specifically the SDDP- α and SDDP-IND policies on the two extreme scenarios (MIN and MAX) together with the nominal scenario and we discuss the resulting costs (Table 2.2), installed capacities (Figures 2.7 and 2.9) and system operations (Figure 2.8). The MIN (MAX) scenario corresponds to the scenario with the lowest (highest) possible realizations of costs and demands in all stages.

Scenarios	Policies	Inv.	Maint.	Op.	Total
MAX	SDDP-IND	24,412	29,235	341,875	395,521
MAX	SDDP-0.75	32,643	32,059	300,867	365,569
MAX	SDDP-0.50	47,803	34,074	273,114	354,991
MAX	SDDP-0.25	58,124	35,589	259,174	352,887
MIN	SDDP-IND	24,412	29,235	34,563	88,210
MIN	SDDP-0.75	22,534	28,625	34,153	85,312
MIN	SDDP-0.50	22,090	28,093	34,947	85,129
MIN	SDDP-0.25	20,402	26,933	35,810	83,145
Nominal	SDDP-IND	24,412	29,235	112,333	165,979
Nominal	SDDP-0.75	23,210	28,575	113,517	165,302
Nominal	SDDP-0.50	21,023	27,218	116,964	165,204
Nominal	SDDP-0.25	20,079	26,485	118,650	165,214

Table 2.2: Costs associated to the SDDP-IND and SDDP- α policies for the maximum (MAX), minimum (MIN), nominal (Nominal) scenarios.

First, we remark in Table 2.2 that the extreme scenarios lead to more extreme total costs than those observed in the previous simulations. This confirms our suspicion that they are not present in the samples. We also observe that the SDDP- α policies perform better than the SDDP-IND one, since the total costs in all scenarios are lower than those of the SDDP-IND, especially on the extreme ones. Indeed, considering the time dependencies of the stochastic process yields a 3%-8% cost reduction on the extreme scenarios. On the one hand, the non-adaptability of the SDDP-IND policy causes the power system to have excess capacity when minimum scenarios are forecast. On the other hand, the power system is not sufficiently protected when maximum scenarios are forecast; indeed, the power system with the SDDP-IND policy has the worst operating cost for the MAX scenario. The operating and investment costs of the SDDP- α policies behave appropriately, i.e., when there is an increase in capacity, the operating costs become lower, but when there is a low capacity, the operating costs become higher. We can notice that the SDDP-0.25 policy performs better than all other policies. It confirms that with a small α the SDDP- α model is trained on a scenario sample that is more likely to include extreme scenarios as it avoids the zigzag effect. Finally, in the nominal scenario, all policies follow a similar behavior as mentioned above and do not show much difference in the actual total cost.

Figure 2.7 plots installed capacities resulting from the SDDP-IND and SDDP- α policies for the maximum (MAX), minimum (MIN), nominal (Nominal) scenarios. For the first 20 years, policies invest little in new capacity. Then, when the nuclear plants reach the end of their useful life in 2030, all policies begin to determine which technologies and resources will meet future energy needs.

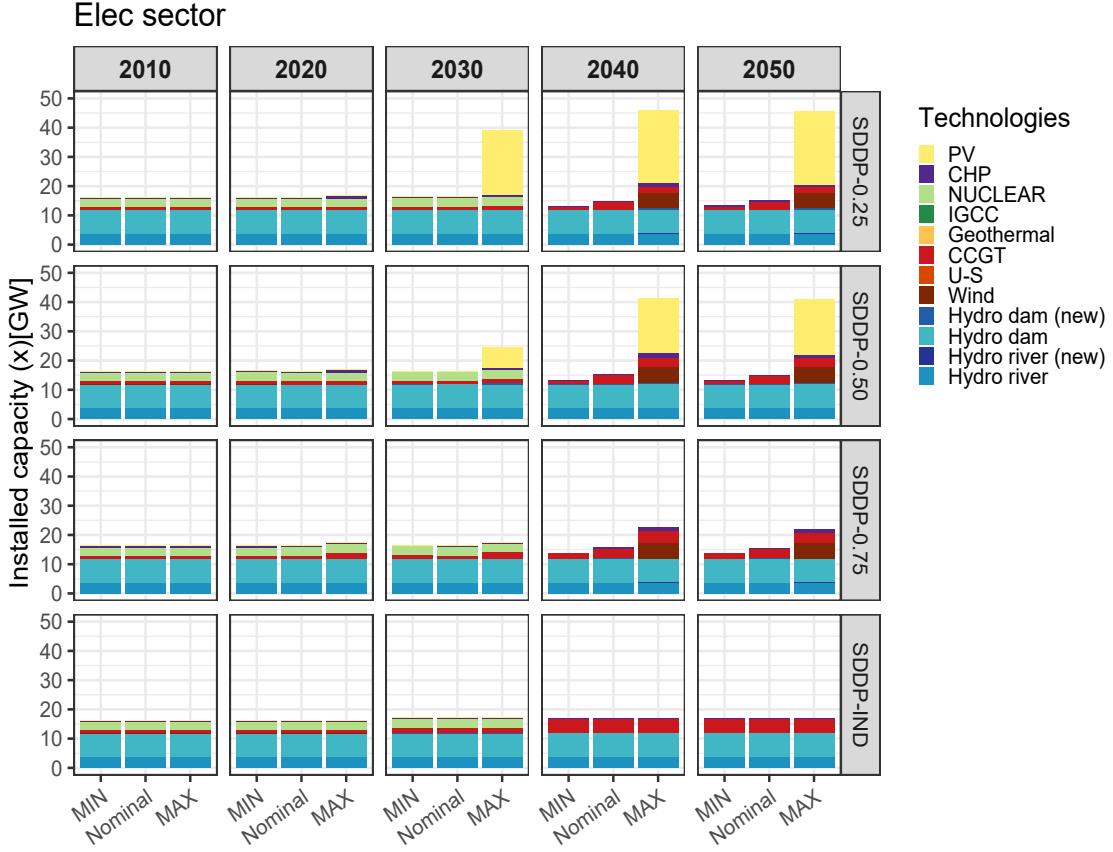


Figure 2.7: Installed capacities for the SDDP-IND and SDDP- α policies for the maximum (MAX), minimum (MIN), nominal (Nominal) scenarios.

When the demands and resource costs are considered stage-wise independent, the optimal policy solution of SDDP-IND prefers conventional technologies (e.g., based on natural gas) to renewables (e.g., photovoltaic, wind) or cogeneration (CHP). In contrast, the SDDP- α policies will have a more diversified set of technologies by 2050, including PV, Cogeneration, Wind and CCGT at the MAX scenario. In the MIN scenario, the solution will only invest in CCGT and Cogeneration.

Figure 2.8 shows the power generation by technology in the electricity sector. As observed previously, the SDDP-IND policy depends heavily on natural gas. In the MAX scenario, the optimal SDDP-IND policy uses all available capacity from the most economical technologies and then supplements generation by importing electricity or using existing conventional technologies, regardless of resource costs. On the other hand, the electricity generation resulting from the SDDP- α policies shows a lower dependence on imported electricity. Imports even completely disappear with $\alpha = 0.5$ but then reappear with $\alpha = 0.25$ to satisfy a more important electrification of the system due certainly to a consideration of more extreme scenarios in the optimization.

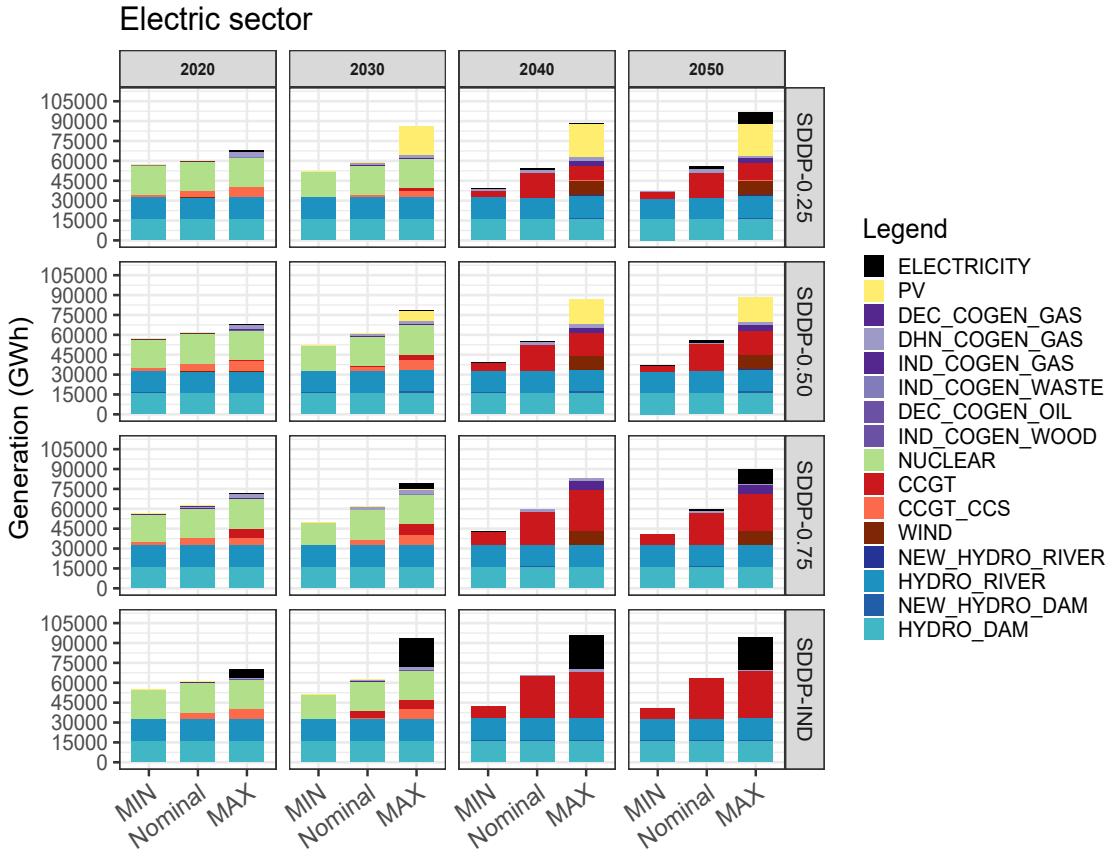


Figure 2.8: Power generation for the SDDP-IND and SDDP- α policies for the maximum (MAX), minimum (MIN), nominal (Nominal) scenarios.

Finally, Figure 2.9 shows the evolution of the heating sector. The electrification of the heating sector is confirmed in the MAX scenario as the α decreases. We thus observe a higher penetration of the Electrical Heat Pumps, Direct Elec. and Solar Thermal technologies.



Figure 2.9: Installed capacity for the heating sector for the SDDP-IND and SDDP- α policies for the maximum (MAX), minimum (MIN), nominal (Nominal) scenarios.

2.6 Conclusions and Future Work

Strategic energy planning models are undoubtedly important to indicate the investments that must be made in technologies in the current context where there is considerable uncertainty about the evolution of costs and demand for energy. Multi-stage stochastic programming provides a very useful framework in that regard, as it allows for investment decisions to be revised periodically as a function of the uncertainty observed up to that point. The use of multi-stage models in this setting, however, faces two challenges: first, defining the input stochastic processes for the model requires modeling the evolution of costs and demand over time; second, the total number of scenarios grows quickly since strategic energy planning models are characterized by the presence of many sources of uncertainty—both in the objective function and in the constraints—which requires the use of specialized algorithms to solve the problem.

In this paper, we have presented an approach that tackles the two challenges presented above. The key to our methodology is to start from ranges of values for the uncertain components—information that typically can be obtained from external studies in the literature—and then to develop auto-regressive processes that are restricted to those ranges. Such development requires special care, as it differs from standard auto-regressive models with normally distributed noise; indeed, processes with normal noise have unbounded support and hence scenarios with unrealistic values might be generated. In contrast, our approach not only stays within the pre-specified ranges in each period, but also introduces a modeling parameter α that allows the user to

control the variability of the process from one period to the next. We also show that the resulting model can be solved with the well-known Stochastic Dual Dynamic Programming (SDDP) algorithm by using a combination of extra state-variables and Markov chains to represent the AR processes.

We have applied our models to the Swiss energy system. The results show the benefits of using a multi-stage model as opposed to two-stage or robust optimization approaches proposed for this problem in the literature, as the latter models require making all the investment decisions at the beginning of the horizon, whereas our multi-stage model allows for making the decisions over time, thereby adapting them to the observed uncertainty. The results also demonstrate the importance of modeling the dependence of the input processes across stages; in fact, the case where the uncertainty is assumed to be independent across stages does not allow for adaptability—although the investment decisions can change from period to period, they do not adapt to the particular scenarios observed in each period. Finally, the results show the crucial role played by the variability-controlling parameter α . As discussed in the paper, lower values of α allow for protection against extreme scenarios.

The limitations that exist when using time dependence within SDDP are that processes that are only on the right-hand side of the constraints can be represented as affine functions of the errors, and with the Markov chain (which is not limited to considering only the right-hand side as uncertain), it is not possible to consider a large number of states Markov chain. According to the papers presented in the literature review, most of them are formulated as a mixed-integer programming model, since the conversion units (technologies) are integer and not continuous. Therefore for future work, it is to be able to consider the integrality of the technologies and compare how these solutions impact this kind of relaxation. Also, as the Swiss government proposes as a policy to establish zero emissions by 2050, it is necessary to include restrictions in the model to reduce emissions during the planning horizon.

Chapter 3

General Conclusion

This thesis focuses on the development of dynamic models, also on the analysis of the assumptions of the distribution and temporal correlation of uncertain parameters in long-term energy planning problems. Three optimization methods were used, such as two-stage stochastic programming (TSSP), two-stage distributionally robust optimization (TS-DRO) and multistage stochastic programming (MSSP). Machine learning techniques and a discretization method of continuous processes to a finite-state Markov chain were used for the formulation of the TS-DRO and MSSP, respectively. This application was divided into two parts.

In Chapter 1, we present the EnergyScope strategic planning model. It is a simplified but complete version of a long-term planning model applied to a real Swiss energy system, considering multiple energy sectors such as transportation, heating and electricity, with a 12-month multiperiod formulation that takes into account seasonality and energy storage for a target year. The level of detail provided by EnergyScope is relatively low compared to others using typical or finer days, but simplification makes it possible to include hundreds of uncertain parameters in a robust optimization modeling framework in an efficient manner, where the uncertainty is represented by variation ranges. Using the probabilistic approach, we analyze how the distribution assumptions of uncertain parameters affect investment decisions. To do so, we started by assigning different probability distributions to the uncertain parameters (considering the range of variation as the support of the distribution), and solved several TSSPs associated with the assumed probability distributions. First results showed that the distribution assumptions of the uncertain parameters impact the investment decisions, and to have solutions less dependent on the chosen probability distribution, an approach combining machine learning with distributionally robust optimization was proposed. Machine learning is used to reduce the size of the ambiguity set to a set of more important uncertain parameters, so that TS-DRO could be a computationally tractable problem. Out-of-sample results show that TS-DRO provides more stable solutions and reduces the impact on investment solutions related to the benchmark distribution. It is known that one of the weaknesses of stochastic programming is the knowledge of the true probability distribution of the uncertain parameter, so we believe that this whole methodology presented in chapter one can be a useful tool to protect against the possible errors that can be made when assuming or assigning one or several probability distributions.

In Chapter 2, we maintain the same application but with an approach more suitable for long-term planning problems. In general, multistage stochastic optimization allows sequential decisions to be made as uncertainty is revealed over the planning horizon. But the major difficulty of these models is how to incorporate dynamic models to account for temporal correlations of uncertainty. In particular, the SDDP algorithm can incorporate dynamic models with linear properties to preserve the convexity of the optimization problem to be solved, limiting it only to the uncertainties on the right-hand side of the constraints. Advanced SDDP algorithms can incorporate Markov chains in addition to the dynamic models, and have the advantage that uncertainty can be considered anywhere in the model, such as in the objective function and in the matrix, but great care must be taken, since the consideration of a few uncertain parameters using Markov chains can lead to a computationally intractable problem. On the other hand, the main contribution of Chapter 2 is the

method that allows us to construct first-order autoregressive processes when historical data are not explicitly available but rather ranges of variation. The AR model resulting from our method has the property that the random variables corresponding to each period are uniformly distributed within those pre-specified ranges, and also allows balancing the occurrence of zigzag and extreme scenarios using an α control. Our numerical experiments show that the assumptions used on temporal correlation are a key factor in the construction of optimal policies, since different installed capacity configurations can be obtained given the degree of correlation assumed. Therefore, this method can be used to verify which parameters can be considered stage-wise independent and which must be stage-wise dependent through a sensitivity analysis.

Through the different optimization approaches presented in the previous chapters of this thesis, we can conclude that when uncertainty is taken into account when performing the energy plan, the solutions obtained in stochastic and other robust approaches reported in the literature show better performance in out-of-sample simulations than the solutions obtained by the classical deterministic approach. Therefore, in order to have robust and stable strategic decisions it is crucial to incorporate the uncertainty of the key parameters in the optimization.

Although TS-DRO and MSSP are very different approaches, the similarities in terms of installed capacity are due to the fact that both share information on the uncertainties associated with the most impacting parameters of the model, such as electricity, natural gas and coal prices. For example, the TS-DRO with $\epsilon = 1$ projects a similar mix of technologies as the SDDP-0.25 in the maximum scenario, for the year 2040, such as PV, wind and cogeneration. The MSSP also shows that during the first decade there were no significant investments in the electricity system due to the participation of nuclear power plants, but thanks to decisions made by the Swiss government to close nuclear power plants by 2030, optimal SDDP- α policies begin to anticipate investment in new power generation technologies based on the degree of correlation and the scenario. Fossil fuel consumption for the years 2040 and 2050 are highly dependent on the time correlation assumption, i.e., assuming stage-wise independent, the SDDP-IND policy encourages investment in natural gas-based technologies as a substitute for nuclear power, and when stage-wise dependent is assumed, natural gas consumption gradually decreases as the level of time correlation (α) increases. However, it is important to remember that these primary energy sources (coal and oil) are not unlimited resources. The authors in [Hannah Ritchie and Rosado \[2020\]](#) provide an overview of global fossil fuel consumption and describe that fossil fuel consumption over the last half century has increased eightfold since 1950 and has roughly doubled since 1980. One of the reasons for the continued use of these fuels is that they tend to be cheaper than other currently available alternatives [\[Roser, 2021\]](#). In energy markets, when energy prices are very low, this drives economic activity, but also leads to higher consumption, and when energy prices are high (either because of extraction costs or scarcity) this will curb consumption and encourage research to find more efficient ways to use resources [\[O'Brien, 2012\]](#). In our results we can see this effect in the MAX scenario for the electricity sector, thus motivating the energy transition to use alternative technological sources, such as renewable technologies, which increases the uncertainty of these models.

In summary, the main contributions of this thesis include:

- The methodology used to restrict the DRO to a reduced set of the most important uncertain parameters of the model, where the training data are constructed by solving multiple problems with a single scenario.
- The methodology for constructing a first-order autoregressive process for time series, using as input data ranges of variation reported in the literature.

Future research in the application of strategic energy planning could include modeling the relationships between uncertain variables over time. Since the AR model does not capture the relationship between multiple quantities as they change over time, we propose to extend the method to vector autoregression (VAR). For example, resource prices and demands, as these act as a market equilibrium mechanism. Another application would be to consider the carbon tax as an environmental regulation and/or to restrict CO_2 emissions over the planning horizon, establishing for the last period zero emissions. Furthermore, the application of these two approaches (TSSP and TS-DRO) could be extended to a new version of the model called EnergyScope

TD, which can consider different time scales for a target year, however, it retains some limitations of the first version: it does not consider the current existing system (in the multistage stochastic programming model presented in Chapter 2 the existing system at the time of planning is considered), it is formulated in a fully deterministic approach, and the annual demand is exogenous to the problem, so it does not result from a supply-demand equilibrium. In addition, it is important to be able to consider the integrality of the technologies in the SDDP and compare how these solutions affect this type of relaxation.

Bibliography

- 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 and A. Nemirovski. Robust convex optimization. *Mathematics of Operations Research*, 23(4):769–805, 1998. doi: 10.1287/moor.23.4.769.
- 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.
- A. Ben-Tal, D. den Hertog, and J.-P. Vial. Deriving robust counterparts of nonlinear uncertain inequalities. *Mathematical Programming*, 149(1):265–299, Feb 2015. doi: 10.1007/s10107-014-0750-8.
- 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. Decomposition and partitioning methods for multistage stochastic linear programs. *Oper Res*, 33(5):989–1007, 1985.
- J. R. Birge and F. Louveaux. *Introduction to Stochastic Programming*. Springer Publishing Company, Incorporated, 2nd edition, 2011. ISBN 1461402360, 9781461402367.
- J. Blanchet, Y. Kang, and K. Murthy. Robust Wasserstein profile inference and applications to machine learning. *Journal of Applied Probability*, 56(3):830–857, 2019.
- G. Box, G. Jenkins, G. Reinsel, and G. Ljung. *Time Series Analysis: Forecasting and Control*. Wiley Series in Probability and Statistics. Wiley, 2015. ISBN 9781118674925. URL <https://books.google.cl/books?id=rNt5CgAAQBAJ>.
- N. Campodónico, S. Binato, R. Kelman, M. Pereira, M. Tinoco, F. Montoya, M. Zhang, and F. Mayaki. Expansion planning of generation and interconnections under uncertainty. *Proc. Ann. 3rd Balkans Power Conf.*, 2003., 2003.
- 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>.
- V. de Matos, D. Morton, and E. Finardi. Assessing policy quality in a multistage stochastic program for long-term hydrothermal scheduling. *Annals of Operations Research*, 253(2):713–731, June 2017. ISSN 0254-5330. doi: 10.1007/s10479-016-2107-6.
- L. Ding, S. Ahmed, and A. Shapiro. A python package for multi-stage stochastic programming. In *Optimization online*, pages 1–41, 2019.
- O. Dowson and L. Kapelevich. SDDP.jl: a Julia package for stochastic dual dynamic programming. *INFORMS Journal on Computing*, 2020. doi: <https://doi.org/10.1287/ijoc.2020.0987>. Articles in Advance.
- 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.
- R. M. Dudley. The speed of mean glivenko-cantelli convergence. *The Annals of Mathematical Statistics*, 40(1):40–50, 1969.
- G. Fella, G. Gallipoli, and J. Pan. Markov-chain approximations for life-cycle models. *Review of Economic Dynamics*, 34:183 – 201, 2019. ISSN 1094-2025. doi: <https://doi.org/10.1016/j.red.2019.03.013>. URL <http://www.sciencedirect.com/science/article/pii/S1094202519301565>.
- 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>.
- P. Gabrielli, F. Fürer, G. Mavromatis, and M. Mazzotti. Robust and optimal design of multi-energy systems with seasonal storage through uncertainty analysis. *Applied Energy*, 238:1192–1210, 2019. ISSN 0306-2619. doi: <https://doi.org/10.1016/j.apenergy.2019.01.064>. URL <http://www.sciencedirect.com/science/article/pii/S0306261919300649>.
- 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 Pt 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.
- E. Guevara, F. Babonneau, T. H. de Mello, and S. Moret. A machine learning and distributionally robust optimization framework for strategic energy planning under uncertainty. *Applied Energy*, 271:115005, 2020. ISSN 0306-2619. doi: <https://doi.org/10.1016/j.apenergy.2020.115005>. URL <http://www.sciencedirect.com/science/article/pii/S0306261920305171>.
- E. Guevara, F. Babonneau, and T. H. de Mello. Modeling uncertainty processes for multi-stage optimization of strategic energy planning : An auto-regressive and markov chain formulation. 2022. sent to EJOR.

- Y. Guo, K. Baker, E. Dall'Anese, Z. Hu, and T. H. Summers. Data-based distributionally robust stochastic optimal power flow—part 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.
- M. R. Hannah Ritchie and P. Rosado. Energy. *Our World in Data*, 2020. <https://ourworldindata.org/energy>.
- H. Heitsch and W. Roemisch. Scenario tree modeling for multistage stochastic programs. *Math. Program.*, 118:371–406, 05 2009. doi: 10.1007/s10107-007-0197-2.
- 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>.
- T. Homem-de Mello, V. de Matos, and E. Finardi. Sampling strategies and stopping criteria for stochastic dual dynamic programming: A case study in long-term hydrothermal scheduling. *Energy Systems*, 2:1–31, 03 2011. doi: 10.1007/s12667-011-0024-y.
- 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>.
- A. Ioannou, G. Fuzuli, F. Brennan, S. W. Yudha, and A. Angus. Multi-stage stochastic optimization framework for power generation system planning integrating hybrid uncertainty modelling. *Energy Economics*, 80:760 – 776, 2019. ISSN 0140-9883. doi: <https://doi.org/10.1016/j.eneco.2019.02.013>. URL <http://www.sciencedirect.com/science/article/pii/S0140988319300702>.
- S. Jin, S. Ryan, J.-P. Watson, and D. Woodruff. Modeling and solving a large-scale generation expansion planning problem under uncertainty. *Energy Systems*, 2:209–242, 11 2011. doi: 10.1007/s12667-011-0042-9.
- K. A. Kopecky and R. M. Suen. Finite state markov-chain approximations to highly persistent processes. *Review of Economic Dynamics*, 13(3):701 – 714, 2010. ISSN 1094-2025. doi: <https://doi.org/10.1016/j.red.2010.02.002>. URL <http://www.sciencedirect.com/science/article/pii/S1094202510000128>.
- J. Krzemień. 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>.
- C. L. Lara, D. S. Mallapragada, D. J. Papageorgiou, A. Venkatesh, and I. E. Grossmann. Deterministic electric power infrastructure planning: Mixed-integer programming model and nested decomposition algorithm. *European Journal of Operational Research*, 271(3):1037 – 1054, 2018. ISSN 0377-2217. doi: <https://doi.org/10.1016/j.ejor.2018.05.039>. URL <http://www.sciencedirect.com/science/article/pii/S0377221718304466>.

- C. L. Lara, J. D. Siirola, and I. Grossmann. Electric power infrastructure planning under uncertainty: stochastic dual dynamic integer programming (sddip) and parallelization scheme. *Optimization and Engineering*, pages 1–39, 2019.
- 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>.
- Y. Liu, R. Sioshansi, and A. J. Conejo. Multistage stochastic investment planning with multiscale representation of uncertainties and decisions. *IEEE Transactions on Power Systems*, 33(1):781–791, 2018.
- N. Löhndorf and A. Shapiro. Modeling time-dependent randomness in stochastic dual dynamic programming. *European Journal of Operational Research*, 273(2):650 – 661, 2019. ISSN 0377-2217. doi: <https://doi.org/10.1016/j.ejor.2018.08.001>. URL <http://www.sciencedirect.com/science/article/pii/S0377221718306787>.
- 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](https://doi.org/10.1016/S0167-6377(98)00054-6). URL <http://www.sciencedirect.com/science/article/pii/S0167637798000546>.
- R. Marathe and S. Ryan. On the validity of the geometric brownian motion assumption. *The Engineering Economist*, 50, 04 2005. doi: 10.1080/00137910590949904.
- M. D. McKay, R. J. Beckman, and W. J. Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239–245, 1979. ISSN 00401706. URL <http://www.jstor.org/stable/1268522>.
- N. Memon, S. B. Patel, and D. P. Patel. Comparative analysis of artificial neural network and xgboost algorithm for polsar image classification. In B. Deka, P. Maji, S. Mitra, D. K. Bhattacharyya, P. K. Bora, and S. K. Pal, editors, *Pattern Recognition and Machine Intelligence*, pages 452–460, Cham, 2019. Springer International Publishing. ISBN 978-3-030-34869-4.
- 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, École 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. Gironès, 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. Moret, F. Babonneau, M. Bierlaire, and F. Maréchal. Overcapacity in european power systems: Analysis and robust optimization approach. *Applied Energy*, 259:113970, 2020c. ISSN 0306-2619. doi: <https://doi.org/10.1016/j.apenergy.2019.113970>. URL <http://www.sciencedirect.com/science/article/pii/S0306261919316575>.
- 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, 2020d. ISSN 0377-2217. doi: <https://doi.org/10.1016/j.ejor.2019.06.015>. URL <https://www.sciencedirect.com/science/article/pii/S0377221719304904>.
- C. Nicolas. *Robust Energy and Climate Modeling for Policy Assessment*. PhD thesis, Paris 10, France, 2016.
- D. O'Brien. Scarcity of fossil fuels banished for the time being. <https://www.irishtimes.com/business/scarcity-of-fossil-fuels-banished-for-the-time-being-1.543570>, 2012. Accessed: 03 may 2022.
- H. Park and R. Baldick. Multi-year stochastic generation capacity expansion planning under environmental energy policy. *Applied Energy*, 183:737 – 745, 2016. ISSN 0306-2619. doi: <https://doi.org/10.1016/j.apenergy.2016.08.164>. URL <http://www.sciencedirect.com/science/article/pii/S0306261916312739>.
- M. V. Pereira and L. M. Pinto. Multi-stage stochastic optimization applied to energy planning. *Math. Program.*, 52(1–3):359–375, May 1991. ISSN 0025-5610.
- S. Pfenniger and B. Pickering. Calliope: a multi-scale energy systems modelling framework. *Journal of Open Source Software*, 3:825, 2018. doi: 10.21105/joss.00825. URL <https://github.com/calliope-project/calliope>.
- A. Philpott and V. de Matos. Dynamic sampling algorithms for multi-stage stochastic programs with risk aversion. *European Journal of Operational Research*, 218(2):470 – 483, 2012. ISSN 0377-2217. doi: <https://doi.org/10.1016/j.ejor.2011.10.056>. URL <http://www.sciencedirect.com/science/article/pii/S0377221711010332>.
- W. B. Powell, A. George, H. Simão, 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.
- A. Prékopa. *Stochastic Programming*. Springer Netherlands, Dordrecht, 1995. ISBN 978-94-017-3087-7.
- 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, 2019a. ISSN 1436-4646. doi: 10.1007/s10107-017-1224-6. URL <https://doi.org/10.1007/s10107-017-1224-6>.

- H. Rahimian, G. Bayraksan, and T. Homem-de Mello. Controlling risk and demand ambiguity in newsvendor models. *European Journal of Operational Research*, 279(3):854–868, 2019b.
- S. Rebennack. Generation expansion planning under uncertainty with emissions quotas. *Electric Power Systems Research*, 114:78–85, 2014.
- H.-K. Ringkjøb, P. M. Haugan, and I. M. Solbrekke. A review of modelling tools for energy and electricity systems with large shares of variable renewables. *Renewable and Sustainable Energy Reviews*, 96:440 – 459, 2018. doi: <https://doi.org/10.1016/j.rser.2018.08.002>.
- 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.
- J. C. Rojas-Zerpa and J. M. Yusta. Methodologies, technologies and applications for electric supply planning in rural remote areas. *Energy for Sustainable Development*, 20:66 – 76, 2014. ISSN 0973-0826. doi: <https://doi.org/10.1016/j.esd.2014.03.003>. URL <http://www.sciencedirect.com/science/article/pii/S0973082614000271>.
- M. Roser. The argument for a carbon price. <https://ourworldindata.org/carbon-price>, 2021. Accessed: 03 may 2022.
- K. Rouwenhorst. Asset pricing implications of equilibrium business cycle models. In T. F. Cooley, editor, *Frontiers of Business Cycle Research*, page 294–330, 1995. Princeton University Press.
- C. Senkpiel, S. Shammugam, W. Biener, N. S. Hussein, C. Kost, N. Kreifels, and W. Hauser. Concept of evaluating chances and risks of grid autarky. In *2016 13th International Conference on the European Energy Market (EEM)*, pages 1–5, June 2016. doi: 10.1109/EEM.2016.7521177.
- A. Shapiro and A. Nemirovski. *On Complexity of Stochastic Programming Problems*, pages 111–146. Springer US, Boston, MA, 2005. ISBN 978-0-387-26771-5. doi: 10.1007/0-387-26771-9_4. URL https://doi.org/10.1007/0-387-26771-9_4.
- A. Shapiro, W. Tekaya, J. P. da Costa, and M. P. Soares. Risk neutral and risk averse stochastic dual dynamic programming method. *European Journal of Operational Research*, 224(2):375 – 391, 2013. ISSN 0377-2217. doi: <https://doi.org/10.1016/j.ejor.2012.08.022>. URL <http://www.sciencedirect.com/science/article/pii/S0377221712006455>.
- A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on stochastic programming : modeling and theory*. SIAM, 2nd edition, 2014.
- J. Singh, Kavinesh, B. Philpott, Andy, and K. Wood. Dantzig-wolfe decomposition for solving multistage stochastic capacity-planning problems, 2009. URL <https://calhoun.nps.edu/handle/10945/38417>.
- A. L. Soyster. Technical note—convex 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>.
- G. Tauchen. Finite state markov-chain approximations to univariate and vector autoregressions. *Economics Letters*, 20(2):177 – 181, 1986. ISSN 0165-1765. doi: [https://doi.org/10.1016/0165-1765\(86\)90168-0](https://doi.org/10.1016/0165-1765(86)90168-0). URL <http://www.sciencedirect.com/science/article/pii/0165176586901680>.
- G. Tauchen and R. Hussey. Quadrature-based methods for obtaining approximate solutions to nonlinear asset pricing models. *Econometrica*, 59(2):371–396, 1991. ISSN 00129682, 14680262. URL <http://www.jstor.org/stable/2938261>.

- F. Thomé, R. C. Perez, L. Okamura, A. Soares, and S. Binato. Stochastic multistage co-optimization of generation and transmission expansion planning, 2019.
- W. Usher and N. Strachan. Critical mid-term uncertainties in long-term decarbonisation pathways. *Energy Policy*, 41:433 – 444, 2012. doi: <https://doi.org/10.1016/j.enpol.2011.11.004>. Modeling Transport (Energy) Demand and Policies.
- A. Velloso, D. Pozo, and A. Street. Distributionally Robust Transmission Expansion Planning: a Multi-scale Uncertainty Approach. *arXiv e-prints*, art. arXiv:1810.05212, 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.
- A. Zerrahn and W.-P. Schill. A dispatch and investment evaluation tool with endogenous renewables dieter, 2015.
- J. Zou, S. Ahmed, and X. A. Sun. Stochastic dual dynamic integer programming. *Math. Program.*, 175(1–2): 461–502, May 2019. ISSN 0025-5610. doi: 10.1007/s10107-018-1249-5. URL <https://doi.org/10.1007/s10107-018-1249-5>.

Appendices

Appendix A

Two-stage stochastic programming

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 $\mathbf{1}_A : X \rightarrow \{0, 1\}$ defined as:

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

(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
$T-EUT\{eut\}$: Set of technologies $\forall eut \in EUT$	$T-EUC\{euc\}$: Set of technologies $\forall euc \in EUC$

(II) Definition of variables

Name	Description	Units
G_{Public}	: Ratio [0; 1] public mobility over total passenger mobility	
G_{Rail}	: Ratio [0; 1] rail transport over total freight transport	
G_{Dhn}	: Ratio [0; 1] centralized over total low-temperature heat	
F_i	: Installed capacity with respect to main output i , $\forall i \in T$	[GW]
Y_i^{solar}	: If 1, technologies i is backup technology for decentralized solar else 0, $\forall i \in T$	
N_i	: Number integer of installed units i of size f_i^{ref} , $\forall i \in T$	
GWP_i^{constr}	: Technology construction GHG emissions, $\forall i \in T$	[ktCO ₂ -eq.]

Table A.1: Variables for the first-stage problem

Name	Description	Units
$\mathbf{Ft}_{i,t}$: Operation the i in each period t , $\forall i \in T \cup R$, $\forall t \in P$	[GW]
$\mathbf{Sto}_{j,l,t}^+$: Input to storage units $j \in Sto$ the $l \in L$ in period $t \in P$	[GW]
$\mathbf{Sto}_{j,l,t}^-$: Output from storage units $j \in Sto$ the $l \in L$ in period $t \in P$	[GW]
$\mathbf{D}_{l,t}$: End-uses demand. Set to 0 if $l \notin EUT$, $\forall l \in L$, $\forall t \in P$	[GW]
\mathbf{GWP}^{tot}	: Total yearly GHG emissions of the energy system	[ktCO ₂ -eq./y]
\mathbf{GWP}_r^{op}	: Total GHG emissions of resources, $\forall r \in R$	[ktCO ₂ -eq./y]
$\mathbf{Loss}_{eut,t}$: Losses in the networks (grid and DHN), $\forall eut \in EUT$, $\forall t \in P$	[GW]

Table A.2: Variables for the second-stage problem

(III) Definition of parameters

Name	Description	Units
$eUYear_{eui,s}$: Annual end-uses in energy services per sector s , $\forall s \in S$, $\forall eui \in EUI$	[GWh/y]
eUI_{eui}	: short name of $endUses_{year}$ Total annual end-uses in energy services eui , $\forall eui \in EUI$ $eUI_{eui} = \sum_{s \in S} eUYear_{eui,s}$	[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	
$\underline{g}_k, \bar{g}_k$: Upper and lower limit to G_k , $\forall k \in \{\%Public, \%DHN, \%Rail\}$	
h_t	: Time periods t duration, $\forall t \in P$	[h]
$\%lighting_t$: 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^{constr}	: Technology construction specific GHG emissions, $\forall i \in T$	[ktCO ₂ -eq./GW]
n_i	: Technology i lifetime, $\forall i \in T$	[y]
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)	
\hat{k}_i	: Yearly capacity i factor, $\forall i \in T$	
$c_{r,t}^{op}$: Specific cost of resources r in periods t , $\forall r \in R$, $\forall t \in P$	[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), $\forall eut \in EUT$	
$\%PeakDHN$: Ratio peak/max. average DHN heat demand	

(IV) Model formulation

Objective for the first-stage problem

$$\min \sum_{i \in T} \tau_i \cdot c_i^{Inv} \cdot \mathbf{F}_i + c_i^{Maint} \cdot \mathbf{F}_i \quad (\text{A.1})$$

The total investment cost of each technology results from the multiplication of its specific investment cost (c^{Inv}) and installed size (\mathbf{F}), which is then annualized with the factor (τ), calculated based on the

interest rate (i_{rate}) and the technology lifetime (n). The total O&M cost is calculated by means of the product of maintenance cost (c^{Maint}) and installed size (\mathbf{F}).

Constraints for the first-stage problem

Constraints (A.2)-(A.12) define the constraints for the first-stage problem, where Constraints (A.7)-(A.12) are added to simplify the use of the model and adapt it to the specific case study of Switzerland. Constraint (A.2) represents the total emissions related to the construction of technologies and is equal to the product of the specific emissions (gwp^{constr}) and the installed size (\mathbf{F}).

$$\mathbf{GWP}_i^{constr} = gwp_i^{constr} \cdot \mathbf{F}_i \quad \forall i \in T \quad (\text{A.2})$$

Constraint (A.3) set the upper and lower limits to the installed capacity of each technology are set by f^{max} and f^{min} , respectively. The latter allows accounting for old technologies still existing in the target year. Constraint (A.4) forces the number of installed units of a technology to be an integer multiple (N) of the reference size f^{ref} .

$$f_i^{min} \leq \mathbf{F}_i \leq f_i^{max} \quad \forall i \in T \quad (\text{A.3})$$

$$\mathbf{N}_i \cdot f_i^{ref} = \mathbf{F}_i \quad \forall i \in T \setminus I \quad (\text{A.4})$$

Constraint (A.5) set the upper and lower limits to the share public vs private mobility, train vs truck in freight transportation and DHN vs decentralized for low-Temperature heating demand.

$$\underline{g}_k \leq \mathbf{G}_k \leq \bar{g}_k \quad \forall k \in \{\%Public, \%Rail, \%DHN\} \quad (\text{A.5})$$

Constraint (A.6) is used to select only one technology as backup for solar in winter months, if decentralized solar thermal (*Decsolar*) panels are installed.

$$\sum_{i \in T} Y_i^{solar} \leq 1 \quad (\text{A.6})$$

Constraint (A.7) links linearly the storage capacity to the new installed power. Constraint (A.8) associates the cost of investment the *PowerToGas* unit to the maximum size of two conversion units.

$$\mathbf{F}_{StoHydro} \leq f_{StoHydro}^{max} \frac{\mathbf{F}_{NewHydroDam} - f_{NewHydroDam}^{min}}{f_{NewHydroDam}^{max} - f_{NewHydroDam}^{min}} \quad (\text{A.7})$$

$$\mathbf{F}_{PowerToGas} = \max \{\mathbf{F}_{PowerToGas}; \mathbf{F}_{GasToPower}\} \quad (\text{A.8})$$

Constraint (A.9) is used to calculate the energy efficiency as a fixed cost. Constraint (A.10) represent an additional investment cost of 9.4 billion CHF₂₀₁₅ is linked proportionally to the deployment of stochastic renewables. Constraint (A.11) links the DHN size to the total size of the installed centralized energy conversion technologies.

$$\mathbf{F}_{EFFICIENCY} = \frac{1}{1 + i_{rate}} \quad (\text{A.9})$$

$$\mathbf{F}_{Grid} \geq 1 + \frac{9400}{c_{Grid}^{Inv}} \frac{\mathbf{F}_{Wind} + \mathbf{F}_{PV}}{f_{Wind}^{max} + f_{PV}^{max}} \quad (\text{A.10})$$

$$\mathbf{F}_{DHN} \geq \sum_{i \in T-EUT\{HeatDHN\}} \mathbf{F}_i \quad (\text{A.11})$$

Constraint (A.12) complies with the decision of the Swiss government of eliminate nuclear power plants at the end of their useful life.

$$\mathbf{F}_{NUCLEAR} = 0 \quad (\text{A.12})$$

Objective for the second-stage problem

$$\min \sum_{r \in R} \sum_{t \in P} h_t \cdot c_{r,t}^{op} \cdot \mathbf{Ft}_{r,t} \quad (\text{A.13})$$

The total operational cost is calculated as the sum of the use over different periods and resources multiplied by the period duration (h_t) and the specific cost of the resource (c^{op}).

Constraints for the second-stage problem

In the second-stage problem, all first-stage decision variables are considered as fixed parameters. Constraints (A.14)-(A.15) link the installed size of a technology to its actual use in each period (\mathbf{Ft}_t) via the two capacity factors, respectively. Constraint (A.16) is used to limited the total use of resources for its yearly availability ($avail_r$).

$$\mathbf{Ft}_{i,t} \leq \mathbf{F}_i \cdot k_{i,t} \quad \forall i \in T, \forall t \in P \quad (\text{A.14})$$

$$\sum_{t \in P} \mathbf{Ft}_{i,t} \cdot h_t \leq \mathbf{F}_i \cdot \hat{k}_i \sum_{t \in P} h_t \quad \forall i \in T \quad (\text{A.15})$$

$$\sum_{t \in P} \mathbf{Ft}_{r,t} \cdot h_t \leq avail_r \quad \forall r \in R \quad (\text{A.16})$$

Constraint (A.17) expresses the balance for each layer: all outputs from resources and technologies (including storage) are used to satisfy the End-Uses-Demand or as inputs to other resources and technologies.

$$\sum_{i \in R \cup T \setminus Sto} f_{i,l} \mathbf{Ft}_{i,t} + \sum_{j \in Sto} (\mathbf{Sto}_{j,l,t}^- - \mathbf{Sto}_{j,l,t}^+) - \mathbf{D}_{l,t} \cdot \mathbf{1}_A(l) \cdot \mathbf{Loss}_{l,t} = 0 \quad \forall l \in L, \forall t \in P, A = \{HeatDHN\} \quad (\text{A.17})$$

Constraint (A.18) regulates the operation (\mathbf{Ft}_t) of decentralized technologies, enforcing that the relative share of heat produced by a given technology j be the same for all t . If decentralized solar thermal (*Decsolar*) panels are not installed, i.e. $\mathbf{F}(Decsolar) = 0$, then the percentage of heat produced by all other technologies j is constant in all periods t . This constraint is linearized as in Moret et al. [2020a].

$$\mathbf{Ft}_{i,t} + \mathbf{Ft}_{Decsolar,t} \cdot \mathbf{Y}_i^{solar} \geq \frac{\mathbf{D}_{HeatDHN,t} + \mathbf{D}_{HeatDec,t}}{eUI_{heatSH} + eUI_{heatHW}} \sum_{t \in P} \mathbf{Ft}_{i,t} \cdot h_t \quad \forall i \in T \setminus EUT\{HeatDec\} \setminus \{Decsolar\}, \forall t \in P \quad (\text{A.18})$$

Constraint (A.19) corresponds to the loss of electricity in the grid and DHN, which are calculated as a percentage (%loss) of the total production and import in the corresponding layers.

$$\mathbf{Loss}_{eut,t} = \sum_{i \in R \cup T \setminus Sto, f_{i,eut} > 0} f_{i,eut} \cdot \mathbf{Ft}_{i,t} \cdot \%loss_{eut} \quad \forall eut \in EUT, \forall t \in P \quad (\text{A.19})$$

Constraint (A.21) is used to calculate the total yearly emissions of the system (\mathbf{GWP}^{tot}), by sum of the emissions related to the construction and end-of-life of the energy conversion technologies (\mathbf{GWP}^{constr}), allocated to one year based on the technology lifetime, and the emissions related to resources (\mathbf{GWP}^{op}) calculated in constraint (A.20).

$$\mathbf{GWP}_r^{op} = \sum_{t \in P} gwp_{r,t}^{op} \mathbf{Ft}_{r,t} \cdot h_t \quad \forall r \in R \quad (\text{A.20})$$

$$\mathbf{GWP}^{tot} = \sum_{i \in T} \frac{\mathbf{GWP}_i^{constr}}{n_i} + \sum_{r \in R} \mathbf{GWP}_r^{op} \quad (\text{A.21})$$

Constraint (A.22) is used to avoid underestimating the cost of centralized heat production, a multiplication factor is introduced to account for peak demand, defined as a $\%Peak_{DHN}$ times the maximum monthly average heat demand. This constraint is linearized as in Moret et al. [2020a].

$$\sum_{i \in T-EUT\{HeatDHN\}} \mathbf{F}_i \geq \%peak_{DHN} \max_{t \in P} \{\mathbf{D}_{HeatDHN,t} + \mathbf{Loss}_{HeatDHN,t}\} \quad (\text{A.22})$$

Constraint (A.23) is complementary to constraint (A.3), as it expresses the minimum ($f^{min,\%}$) and maximum ($f^{max,\%}$) yearly output shares of each technology for each type of EUD.

$$\sum_{i' \in T-EUT(eut)} f_i^{min,\%} \sum_{t \in P} \mathbf{Ft}_{i',t} h_t \leq \sum_{t \in P} \mathbf{Ft}_{i,t} h_t \leq \sum_{i' \in T-EUT(eut)} f_i^{max,\%} \sum_{t \in P} \mathbf{Ft}_{i',t} h_t \quad \forall eut \in EUT, \forall i \in T-EUT\{eut\} \quad (\text{A.23})$$

Constraint (A.24) imposes that the share of the different technologies for mobility be the same in each period.

$$\mathbf{Ft}_{i,t} \sum_{t' \in P} h_{t'} \geq \sum_{t' \in P} \mathbf{Ft}_{i,t'} h_{t'} \quad \forall t \in P, \forall i \in T-EUC\{MobPass\} \cup T-EUC\{MobFreight\} \quad (\text{A.24})$$

Constraint (A.25) is a complement of constraint (A.7) and it is used for ensures that the shifted production in a given time period does not exceed the electricity production by the dams in that period.

$$\mathbf{Sto}_{StoHydro,Elec,t}^+ \leq \mathbf{Ft}_{HydroDam,t} + \mathbf{Ft}_{NewHydroDam,t} \quad \forall t \in P \quad (\text{A.25})$$

Constraint (A.26) is used for modeled the storage as a “tank” whose level (\mathbf{F}_t) in period t is equal to the level at the end of the previous period plus input to the storage (\mathbf{Sto}^+) minus output (\mathbf{Sto}^-) in t .

$$\mathbf{Ft}_{j,t} = \mathbf{Ft}_{j,t-1} + h_t \sum_{\substack{l \in L \\ \eta_{j,l}^+ > 0}} \mathbf{Sto}_{j,l,t}^+ \eta_{j,l}^+ - h_t \sum_{\substack{l \in L \\ \eta_{j,l}^- > 0}} \mathbf{Sto}_{j,l,t}^- / \eta_{j,l}^- \quad \forall j \in Sto, \forall t \in P \quad (\text{A.26})$$

Constraint (A.27)-(A.32) shows the constraints relative to the calculation of the EUD in each period t , starting from the projected total yearly demand \mathbf{D} (eUI, input from demand-side model) summed across the different energy sectors (households, services, industry, transport).

$$\mathbf{D}_{Elec,t} = \frac{eUI_{Elec}}{\sum_{t' \in P} h_{t'}} + eUI_{lighting} \frac{\%lighting_t}{h_t} + \mathbf{Loss}_{Elec,t} \quad \forall t \in P \quad (\text{A.27})$$

$$\mathbf{D}_{q,t} = \left(\frac{eUI_{heatHW}}{\sum_{t' \in P} h_{t'}} + eUI_{heatSH} \frac{\%sh_t}{h_t} \right) (\mathbf{1}_B(q) + (-1)^{\mathbf{1}_B(q)} \mathbf{G}_{\%Dhn}) \quad \begin{array}{l} \forall t \in P; \\ q \in \{HeatDHN, HeatDec\}; \\ B = \{HeatDec\} \end{array} \quad (\text{A.28})$$

$$\mathbf{D}_{q,t} = \frac{eUI_{passenger}}{\sum_{t' \in P} h_{t'}} (\mathbf{1}_{\{Pub, Pri\}}(q) + (-1)^{\mathbf{1}_{\{Pub\}}(q)} \mathbf{G}_{\%Public}) \quad \forall t \in P, q \in \{Pub, Pri\} \quad (\text{A.29})$$

$$\mathbf{D}_{q,t} = \frac{eUI_{freight}}{\sum_{t_1 \in P} h_{t_1}} (\mathbf{1}_{\{Road\}}(q) + (-1)^{\mathbf{1}_{\{Road\}}(q)} \mathbf{G}_{\%Rail}) \quad \forall t \in P, q \in \{Rail, Road\} \quad (\text{A.30})$$

$$\mathbf{D}_{HeatT,t} = \frac{eUI_{HeatT}}{\sum_{t' \in P} h_{t'}} \quad \forall t \in P \quad (\text{A.31})$$

$$\mathbf{D}_{r,t} = 0 \quad \forall t \in P, r \in R \setminus \{BioFuels \cup Export\} \quad (\text{A.32})$$

Appendix B

Sample generation and optimality gap

B.1 The sample average approximation method

Consider the stochastic programming problem

$$v^* = \min_{\mathbf{x} \in X} \{g(\mathbf{x}) := c^T \mathbf{x} + \mathbb{E}[Q(\mathbf{x}, \xi)]\} \quad (\text{B.1})$$

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 optimality 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:

$$\hat{v}_N = \min_{\mathbf{x} \in X} \left\{ g(\mathbf{x}) := c^T \mathbf{x} + \frac{1}{N} \sum_{i=1}^N Q(\mathbf{x}, \xi_i) \right\} \quad (\text{B.2})$$

To reduce the computational effort in solving the problem (B.2), 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 (B.2). Let \hat{v}_N^k and \hat{x}_N^k be the corresponding optimal objective and optimal solutions, respectively, with $k = 1, \dots, K$.

Then we can estimate v^* by

$$\bar{v}_N^K = \frac{1}{K} \sum_{k=1}^K \hat{v}_N^k \quad (\text{B.3})$$

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:

$$\hat{g}_{N'}(\hat{x}) = c^T \hat{x} + \frac{1}{N'} \sum_{i=1}^{N'} Q(\hat{x}, \xi_i) \quad (\text{B.4})$$

where $\xi_1, \dots, \xi_{N'}$ 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 $N = 1,500$.

Appendix C

Multi-stage stochastic programming

C.1 Mathematical model formulation

In this subsection, we define the sets, the parameters and the variables that are used in the multi-stage strategic energy planning model. Figure C.1 summarizes the sets with their relative indices used throughout the paper.

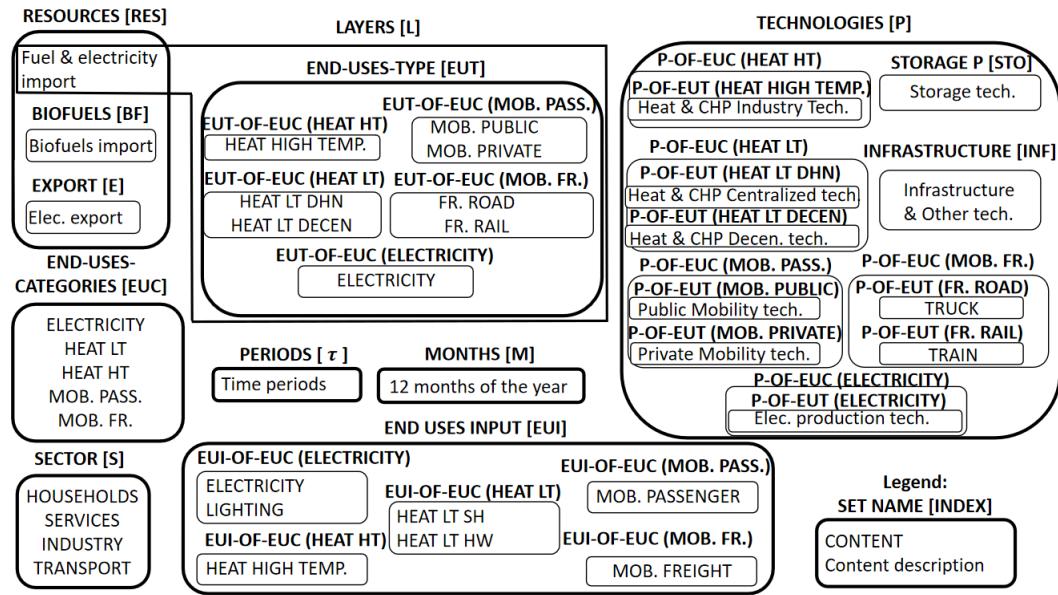


Figure C.1: Diagram of sets and indices used in the MSLP model (source: [Moret et al. \[2017\]](#)). Abbreviations: space heating (SH), hot water (HW), temperature (Temp.), mobility (MOB), Low Temperature (LT), High Temperature (HT), Freight (FR.), Passenger (PASS.).

In Figure C.1, we represent the elements defining the set by a square with rounded edges. For example, the set **EUC** is one of the main sets of the model and the elements of this set will be used as indexes for new subsets. The set **EUT** is the union of the subsets **EUT-OF-EUC** indexed by elements of the set **EUC**. In addition, the end-use demands are calculated based on these elements. The set **EUI** is the union of the subsets **EUI-OF-EUC** indexed by the elements of the set **EUC**. The set **P** (or processes) also contains subsets indexed by elements of the **EUC** set, but in turn, the subsets formed by the elements **HEAT LT**,

MOB. PASS. and MOB. FR. contain two subsets indexed by their corresponding $eut \in \mathbf{EUT}$ element.

The set **L** is a special set, since on these elements the model will have to balance what is generated, produced, consumed or stored in each month. Also each $l \in \mathbf{L}$ element defines a type of end-use demand that must be satisfied. As an example, in the case of decentralized heat production with a NG boiler, the amount of NG consumed by the boiler is an input layer, and the amount of heat produced by the boiler is an output layer. In the case of electricity, the model can import or produce electricity within the system to meet the electricity end use demand; in addition, electricity can be stored in hydroelectric dams or used as an input to other energy conversion technologies (such as electric vehicles) to meet other types of end-use demand.

(I) Sets (and Indices).

\mathcal{T}	time periods, $t \in \mathcal{T} = \{1, \dots, T\}$
\mathbf{M}	twelve-month period, $m \in \mathbf{M} = \{1, \dots, 12\}$
\mathbf{RES}	imported or local resources, $r \in \mathbf{RES} = \{\text{coal, gas natural (NG), electricity, wood, ...}\}$
\mathbf{INF}	infrastructure: DHN, grid, and intermediate energy conversion technologies (i.e. not directly supplying end-use demand), $inf \in \mathbf{INF} = \{\text{electrolysis, pyrolysis, ...}\}$
\mathbf{E}	exported electricity ($\subset \mathbf{RES}$), $ex \in \mathbf{E} = \{\text{electricity export}\}$
\mathbf{S}	sectors of the energy system, $s \in \mathbf{S} = \{\text{households, services, industry, transport}\}$
\mathbf{STO}	storage units, $u \in \mathbf{STO} = \{\text{Power-to-Gas, Pumped Hydro}\}$
\mathbf{EUC}	categories of demand (end-uses), $euc \in \mathbf{EUC} = \{\text{electricity, heat low temperature, ...}\}$
\mathbf{EUI}	input to the model, $eui \in \mathbf{EUI} = \{\text{electricity, lighting, heat lt sh, heat lt hw, ...}\}$
	$\mathbf{EUI} = \bigcup_{euc \in \mathbf{EUC}} \mathbf{EUI-OF-EUC}_{euc}$
\mathbf{EUT}	types of demand (end-uses), $eut \in \mathbf{EUT} = \{\text{electricity, heat high temperature, ...}\}$
	$\mathbf{EUT} = \bigcup_{euc \in \mathbf{EUC}} \mathbf{EUT-OF-EUC}_{euc}$
\mathbf{L}	layers are used to balance resources/processes in the system, $l \in \mathbf{L} = \mathbf{RES} \setminus \{\mathbf{BF} \cup \mathbf{E}\} \cup \mathbf{EUT}$
$\mathbf{P-OF-EUT}_{ELECTRICITY}$	energy conversion technologies for electricity generation with $p \in \{\text{nuclear, ccgt, pv, wind, hydro dam, hydro river, coal, ...}\}$
$\mathbf{P-OF-EUT}_{HEAT\ HIGH\ TEMP.}$	energy conversion technologies for heat production in industrial processes at high temperatures with $p \in \{\text{boiler, CHP, elec. direct heating}\}$
$\mathbf{P-OF-EUT}_{HEAT\ LT\ DHN}$	energy conversion technologies for centralized heat production (DHN) with $p \in \{\text{boiler, CHP, heat pump, deep geothermal}\}$
$\mathbf{P-OF-EUT}_{HEAT\ LT\ DECEN}$	energy conversion technologies for decentralized heat production with $p \in \{\text{boiler, CHP, fuel cell, solar thermal, heat pump, elec. direct heating}\}$
$\mathbf{P-OF-EUT}_{MOB.\ PUBLIC}$	energy conversion technologies for passenger mobility in the public sector with $p \in \{\text{gasolina car, diesel car, NG car, electric car, ...}\}$
$\mathbf{P-OF-EUT}_{MOB.\ PRIVATE}$	energy conversion technologies for passenger mobility in the private sector with $p \in \{\text{diesel bus, hybrid bus, trolley bus, train/metro, ...}\}$
$\mathbf{P-OF-EUT}_{FR.\ ROAD}$	energy conversion technologies for road freight transport with $p \in \{\text{truck}\}$
$\mathbf{P-OF-EUT}_{FR.\ RAIL}$	energy conversion technologies for rail freight transport with $p \in \{\text{train freight}\}$
$\mathbf{P-OF-EUC}_{ELECTRICITY}$	energy conversion technologies in the electricity category with $p \in \{k : k \in \mathbf{P-OF-EUT}_{ELECTRICITY}\}$
$\mathbf{P-OF-EUC}_{HEAT\ HT}$	energy conversion technologies in the heat high temp. category with $p \in \{k : k \in \mathbf{P-OF-EUT}_{HEAT\ HIGH\ TEMP.}\}$
$\mathbf{P-OF-EUC}_{HEAT\ LT}$	energy conversion technologies in the heat low temp. category with $p \in \{k : k \in \mathbf{P-OF-EUT}_{HEAT\ LT\ DHN} \cup \mathbf{P-OF-EUT}_{HEAT\ LT\ DECEN}\}$
$\mathbf{P-OF-EUC}_{MOB.\ PASS.}$	energy conversion technologies in the mobility passenger category with $p \in \{k : k \in \mathbf{P-OF-EUT}_{MOB.\ PUBLIC} \cup \mathbf{P-OF-EUT}_{MOB.\ PRIVATE}\}$
$\mathbf{P-OF-EUC}_{MOB.\ FR.}$	energy conversion technologies in the mobility freight category with $p \in \{k : k \in \mathbf{P-OF-EUT}_{MOB.\ ROAD} \cup \mathbf{P-OF-EUT}_{MOB.\ RAIL}\}$
\mathbf{P}	processes with $p \in \left\{ k : k \in \left(\bigcup_{euc \in \mathbf{EUC}} \mathbf{P-OF-EUC}_{euc} \right) \cup \mathbf{STO} \cup \mathbf{INF} \right\}$
\mathbf{MOB}	energy conversion technologies for passenger and freight transportation ($\subset \mathbf{P}$), $p \in \mathbf{MOB} = \mathbf{P-OF-EUC}_{MOB.PASS.} \cup \mathbf{P-OF-EUC}_{MOB.FR.}$

(II) Parameters

¹[Mpkm] (millions of passenger-km) for passenger, [Mtkm] (millions of ton-km) for freight mobility end-uses

²[Mpkm/h] for passenger, [Mtkm/h] for freight mobility end-uses

Name	Description	Units
T	Parameters used in investments	
a_l	number of planning periods $a_l = 1$ if $l = \text{HEAT LT DHN}$, $a_l = 0$ otherwise. $l \in \mathbf{L}$	
Δ_T	number of years in each period	
i_{rate}	discount rate	
$\bar{g}_{\%public}, g_{\%public}$	Upper and lower limit to $\%_{public}$ (mobility public)	
$\bar{g}_{\%rail}, g_{\%rail}$	Upper and lower limit to $\%_{rail}$ (mobility rail)	
$\bar{g}_{\%dhn}, g_{\%dhn}$	Upper and lower limit to $\%_{dhn}$ (heat low temp. DHN)	
$c_{t,p}^{Inv}$	investment cost of technology $p \in \mathbf{P}$ in the time period $t \in \mathcal{T}$	[MCHF/GW] ^{2,3}
lt_p	lifetime of technology $p \in \mathbf{P}$	
ς_t	annualization factor in time period $t \in \mathcal{T}$	
$rest_{t,p}$	residual capacity of technology $p \in \mathbf{P}$ in time period $t \in \mathcal{T}$	[GW]
$\bar{f}_p^{min}, \bar{f}_p^{max}$	minimum/maximum available capacity of the technology installed $p \in \mathbf{P}$	[GW]
Parameters used for system operations		
$f_{t,p}^{min}, f_{t,p}^{max}$	minimum/maximum capacity of installed technology $p \in \mathbf{P}$ in time period $t \in \mathcal{T}$	[GW] ^{2,3}
$f_p^{min,\%}, f_p^{max,\%}$	it expresses the minimum and maximum yearly output shares of each technology $p \in P$ for each type of end uses demand	
$avail_r$	available amount of the resource $r \in \mathbf{RES}$	[GWh]
$k_{p,m}$	capacity factor of technology $p \in \mathbf{P}$ in month $m \in \mathbf{M}$	
k_p	yearly capacity factor of technology $p \in \mathbf{P}$	
$c_{t,r,m}^p$	cost of the resource $r \in \mathbf{RES}$ in month $m \in \mathbf{M}$ and time period $t \in \mathcal{T}$	[MCHF/GWh]
gwp_r^{Op}	emissions associated to fuels (from cradle to combustion) and imports of electricity. $r \in \mathbf{RES}$	[ktCO ₂ -eq./GWh]
$\eta_{u,l}^+, \eta_{u,l}^-$	efficiency [0:1] of storage $u \in \mathbf{STO}$ input from/output to layer $l \in \mathbf{L}$	
$\%loss_{eut}$	electrical and thermal energy loss factor [0,1] (GRID and DHN). $eut \in \{\text{ELECTRICITY, HEAT LT DHN}\}$	
$\%Peak_{DHN}$	factor used to avoid underestimating the cost of centralized heat production	
$eUYear_{t,eui,s}$	annual end-uses in energy services $eui \in \mathbf{EUI}$ per sector $s \in \mathbf{S}$ in the time period $t \in \mathcal{T}$	[GWh] ¹
$eUI_{t,eui}$	total annual end-uses in energy services $eui \in \mathbf{EUI}$ in the time period $t \in \mathcal{T}$	[GWh] ¹
h_m	$eUI_{t,eui} = \sum_{s \in S} eUYear_{t,eui,s} \quad \forall eui \in \mathbf{EUI}, t \in \mathcal{T}$ duration of the month $m \in \mathbf{M}$ in hours	[h]
$\%lighting_m$	monthly distribution factor for lighting end use input (adding up to 1, $\sum_{m \in M} \%lighting_m = 1$)	
$\%sh_m$	monthly distribution factor for space heating end use input (adding up to 1, $\sum_{m \in M} \%sh_m = 1$)	
$f_{i,l}^{Maint}$	input/output resources/technologies to layers. $i \in \mathbf{RES} \cup \mathbf{P} \setminus \mathbf{STO}, l \in \mathbf{L}$	[GW] ²
$c_{t,p}^{Maint}$	technology maintenance cost $p \in \mathbf{P}$ in time period $t \in \mathcal{T}$	[MCHF/GW] ^{2,3}

(III) Variables

Name	Description	Units
variables used for investments		
$x_{t,p}$	total available capacity of technology $p \in \mathbf{P}$ in time period $t \in \mathcal{T}$	[GW] ³
$y_{t,p}$	new installed capacity of technology $p \in \mathbf{P}$ in time period $t \in \mathcal{T}$	[GW] ³
$G_{t,\%public}$	ratio [0; 1] public mobility over total passenger mobility in time period $t \in \mathcal{T}$	
$G_{t,\%rail}$	ratio [0; 1] rail transport over total freight transport in time period $t \in \mathcal{T}$	
$G_{t,\%dhn}$	ratio [0; 1] centralized over total low-temperature heat in time period $t \in \mathcal{T}$	
C_t^{Inv}	discounted total investment cost in time period $t \in \mathcal{T}$	[MCHF]
C_t^{SV}	discounted salvage value in time period $t \in \mathcal{T}$	[MCHF]
Variables used for system operations		
$z_{t,p,m}$	the operation of resources and technologies in each month $m \in \mathbf{M}$ and time period $t \in \mathcal{T}$	[GW] ^{2,3}
$Sto_{t,u,l,m}^+$	input to storage units $u \in \mathbf{STO}, l \in \mathbf{L}, m \in \mathbf{M}, t \in \mathcal{T}$	[GW]
$Sto_{t,u,l,m}^-$	output from storage units $u \in \mathbf{STO}, l \in \mathbf{L}, m \in \mathbf{M}, t \in \mathcal{T}$	[GW]
$D_{t,l,m}$	end-use demand $l \in \mathbf{L}, m \in \mathbf{M}, t \in \mathcal{T}$. Set to 0 if $l \notin \mathbf{EUT}$	[GW]
$GWP_{t,r}^{op}$	GHG emissions of resources $r \in \mathbf{RES}$ in time period $t \in \mathcal{T}$	[ktCO ₂ -eq.]
$Loss_{t,eut,m}$	losses in the networks (grid and DHN), $t \in \mathcal{T}, eut \in \mathbf{EUT}, m \in \mathbf{M}$	[GW]
$C_t^{O\&M}$	discounted O&M cost in time period $t \in \mathcal{T}$	[MCHF]

³[GWh] if $p \in \mathbf{STO}$

C.2 Objective function

The objective function of the energy problem minimizes the total discounted system cost (i.e., investment and operation costs) minus the salvage value of the residual life of installed technologies at the end of the planning horizon.

$$\min \sum_{t=1}^T \frac{\mathbf{C}_t^{Inv} - \mathbf{C}_t^{SV} + \mathbf{C}_t^{O\&M}}{(1 + i_{rate})^{(t-1) \cdot \Delta_T}} \quad (\text{C.1})$$

The investment cost \mathbf{C}_t^{Inv} on period t is given by

$$\mathbf{C}_t^{Inv} = \sum_{p \in \mathbf{P}} c_{t,p}^{Inv} \cdot \mathbf{y}_{t,p}. \quad (\text{C.2})$$

The O&M cost $\mathbf{C}_t^{O\&M}$ on period t is defined as follows

$$\mathbf{C}_t^{O\&M} = \varsigma_t \cdot \sum_{p \in \mathbf{P}} c_{t,p}^{Maint} \cdot \mathbf{x}_{t,p} + \varsigma_t \cdot \sum_{r \in \mathbf{RES}} \sum_{m \in \mathbf{M}} c_{t,r,m}^{Op} \cdot \mathbf{z}_{t,r,m} \cdot h_m^{Op}, \quad (\text{C.3})$$

where parameter ς_t is an annualization factor. Maintenance and operation costs are linked to the available capacities and resource activities, respectively. Finally the salvage value \mathbf{C}_t^{SV} is defined as

$$\mathbf{C}_t^{SV} = \sum_{p \in \mathbf{P}} sv_{t,p} \cdot c_{t,p}^{Inv} \cdot \mathbf{y}_{t,p} \cdot \mathbf{1}_{\{t+lt_p > T+1\}}(t), \quad (\text{C.4})$$

with

$$sv_{t,p} = \frac{1 - (1 + i_{rate})^{-\Delta_T \cdot (t+lt_p - T - 1)}}{\frac{(1+i_{rate})^{\Delta_T \cdot (T+1-t)}}{(1-(1+i_{rate})^{-\Delta_T \cdot lt_p})}}.$$

C.3 Constraints

The first set of restrictions refers to the initial design of the capacity and its subsequent expansion, which considers the new installed capacity and the decommissioning plan. The available capacity $\mathbf{x}_{t,p}$ of technology p in time period t , as shown in (C.6), is continuously updated by balancing the accumulated capacity $\mathbf{x}_{t-1,p}$ in time period $t-1$, the new installed capacity $\mathbf{y}_{t,p}$, the elimination of unused capacity $rest_{t,p} - rest_{t-1,p}$ at the beginning of the time period t and the retirement capacity of technologies that were installed for optimization but have reached their useful life.

$$\mathbf{x}_{0,p} = res_{0,p}, \quad \forall p \in \mathbf{P} \quad (\text{C.5})$$

$$\mathbf{x}_{t,p} = \begin{cases} \mathbf{x}_{t-1,p} + \mathbf{y}_{t,p} - rest_{t-1,p} + rest_{t,p} - \mathbf{y}_{t-lt_p,p} & \text{if } t > lt_p \quad \forall p \in \mathbf{P}, t \in \mathcal{T} \\ \mathbf{x}_{t-1,p} + \mathbf{y}_{t,p} - rest_{t-1,p} + rest_{t,p} & \text{if } 0 < t \leq lt_p \quad \forall p \in \mathbf{P}, t \in \mathcal{T} \end{cases} \quad (\text{C.6})$$

At $t = 0$, the initial capacity is the existing infrastructure at that moment as show in (C.5). Once a technology is installed, it can be operated until its end-of-life. For the available capacity we use the upper and lower limits for the entire planning horizon (C.7), as follows

$$\bar{f}_p^{min} \leq \mathbf{x}_{t,p} \leq \bar{f}_p^{max} \quad \forall p \in \mathbf{P}, t \in \mathcal{T} \quad (\text{C.7})$$

C.3.0.1 Long-term investment equations

Constraint (C.8) set the upper and lower bounds to the new installed capacity for each time period $t \in \mathcal{T}$ and technology $p \in \mathbf{P}$.

$$f_{t,p}^{min} \leq \mathbf{y}_{t,p} \leq f_{t,p}^{max} \quad \forall p \in \mathbf{P}, t \in \mathcal{T} \quad (\text{C.8})$$

Constraint (C.9) forces the new installed capacity at each time period t to equal total annual end-use for passenger and freight mobility. The addition of this constraint is motivated by the fact that the investment cost of passenger and freight transport technologies is not accounted for in the model. This helps to avoid the installation of more capacity than necessary in the transportation sector.

$$\sum_{mob \in \mathbf{P}\text{-OF-EUC}\{euc\}} \mathbf{y}_{t,mob} = \frac{eUI_{t,euc}}{\sum_{m \in \mathbf{M}} h_m} \quad \forall euc \in \{\mathbf{MOB. PASS.}, \mathbf{MOB. FR.}\}, t \in \mathcal{T} \quad (\text{C.9})$$

Constraints (C.10)-(C.12) set the upper and lower limits to the share public vs private mobility, train vs truck in freight transportation and DHN vs decentralized for low-temperature heating demand. Note that, although these variables are not investment decisions, they are related to them, since these variables divide mobility (passenger and freight) and heat low demands, which can affect investment decisions.

$$\underline{g}_{\%_{public}} \leq \mathbf{G}_{t,\%_{public}} \leq \bar{g}_{\%_{public}} \quad \forall t \in \mathcal{T} \quad (\text{C.10})$$

$$\underline{g}_{\%_{rail}} \leq \mathbf{G}_{t,\%_{rail}} \leq \bar{g}_{\%_{rail}} \quad \forall t \in \mathcal{T} \quad (\text{C.11})$$

$$\underline{g}_{\%_{dhn}} \leq \mathbf{G}_{t,\%_{dhn}} \leq \bar{g}_{\%_{dhn}} \quad \forall t \in \mathcal{T} \quad (\text{C.12})$$

There are a number of other constraints related to investment in new capacity, but we present them later because they are constraints specific to the Swiss energy problem.

C.3.1 Short-term operating equations

The operation of resources and technologies are determined by the operating variable \mathbf{z} . Constraints (C.13)-(C.14) link the total capacity available of a technology to its actual use in each month and time period via two capacity factors: a capacity factor for each month ($k_{p,m}$) depending on resource availability (e.g. renewable) and a yearly capacity factor (\hat{k}_p) accounting for technology downtime and maintenance.

$$\mathbf{z}_{t,p,m} \leq \mathbf{x}_{t,p} \cdot k_{p,m} \quad \forall p \in \mathbf{P}, \forall m \in \mathbf{M}, t \in \mathcal{T} \quad (\text{C.13})$$

$$\sum_{m \in \mathbf{M}} \mathbf{z}_{t,p,m} \cdot h_m \leq \mathbf{x}_{t,p} \cdot \hat{k}_p \sum_{m \in \mathbf{M}} h_m \quad \forall p \in \mathbf{P}, t \in \mathcal{T} \quad (\text{C.14})$$

In the constraint (C.15), the total use of resources is limited by the yearly availability ($avail_r$).

$$\sum_{m \in \mathbf{M}} \mathbf{z}_{t,r,m} \cdot h_m \leq avail_r \quad \forall r \in \mathbf{RES}, t \in \mathcal{T} \quad (\text{C.15})$$

Constraint (C.16) is the balancing equation for storage units. In this constraint when $m = 1$ then $\mathbf{z}_{t,u,0} = \mathbf{z}_{t,u,12}$.

$$\mathbf{z}_{t,u,m} = \mathbf{z}_{t,u,m-1} + h_m \sum_{\substack{l \in \mathbf{L} \\ \eta_{u,l}^+ > 0}} \mathbf{Sto}_{t,u,l,m}^+ \eta_{u,l}^+ - h_m \sum_{\substack{l \in \mathbf{L} \\ \eta_{u,l}^- > 0}} \mathbf{Sto}_{t,u,l,m}^- / \eta_{u,l}^- \quad \forall u \in \mathbf{STO}, \forall m \in \mathbf{M}, t \in \mathcal{T} \quad (\text{C.16})$$

Constraint (C.17)-(C.25) shows the calculation of the end-uses demand \mathbf{D} in each month m and layer l , as a function of end-use input (eUI).

$$\mathbf{D}_{t,ELEC.,m} = \frac{eUI_{t,ELECTRICITY}}{\sum_{m' \in \mathbf{M}} h_{m'}} + eUI_{t,lighting} \frac{\%lighting_m}{h_m} + \mathbf{Loss}_{t,ELEC.,m} \quad \forall m \in \mathbf{M}, t \in \mathcal{T} \quad (\text{C.17})$$

$$\mathbf{D}_{t,\text{HEAT LT DHN},m} = \left(\frac{eUI_{t,\text{HEAT LT HW}}}{\sum_{m' \in \mathbf{M}} h_{m'}} + eUI_{t,\text{HEAT LT SH}} \frac{\%sh_m}{h_m} \right) \cdot \mathbf{G}_{t,\%dh_n} \quad \forall m \in \mathbf{M}, t \in \mathcal{T} \quad (\text{C.18})$$

$$\mathbf{D}_{t,\text{HEAT LT DEC},m} = \left(\frac{eUI_{t,\text{HEAT LT HW}}}{\sum_{m' \in \mathbf{M}} h_{m'}} + eUI_{t,\text{HEAT LT SH}} \frac{\%sh_m}{h_m} \right) \cdot (1 - \mathbf{G}_{t,\%dh_n}) \quad \forall m \in \mathbf{M}, t \in \mathcal{T} \quad (\text{C.19})$$

$$\mathbf{D}_{t,\text{MOB. PUBLIC},m} = \frac{eUI_{t,\text{MOB. PASSENGER}}}{\sum_{m' \in \mathbf{M}} h_{m'}} \cdot \mathbf{G}_{t,\%public} \quad \forall m \in \mathbf{M}, t \in \mathcal{T} \quad (\text{C.20})$$

$$\mathbf{D}_{t,\text{MOB. PRIVATE},m} = \frac{eUI_{t,\text{MOB. PASSENGER}}}{\sum_{m' \in \mathbf{M}} h_{m'}} \cdot (1 - \mathbf{G}_{t,\%public}) \quad \forall m \in \mathbf{M}, t \in \mathcal{T} \quad (\text{C.21})$$

$$\mathbf{D}_{t,\text{FR. RAIL},m} = \frac{eUI_{t,\text{MOB. FREIGHT}}}{\sum_{m' \in \mathbf{M}} h_{m'}} \cdot \mathbf{G}_{t,\%rail} \quad \forall m \in \mathbf{M}, t \in \mathcal{T} \quad (\text{C.22})$$

$$\mathbf{D}_{t,\text{FR. ROAD},m} = \frac{eUI_{t,\text{MOB. FREIGHT}}}{\sum_{m' \in \mathbf{M}} h_{m'}} \cdot (1 - \mathbf{G}_{t,\%rail}) \quad \forall m \in \mathbf{M}, t \in \mathcal{T} \quad (\text{C.23})$$

$$\mathbf{D}_{t,\text{HEAT HIGH TEMP.},m} = \frac{eUI_{t,\text{HEAT HIGH TEMP.}}}{\sum_{m' \in \mathbf{M}} h_{m'}} \quad \forall m \in \mathbf{M}, t \in \mathcal{T} \quad (\text{C.24})$$

$$\mathbf{D}_{t,r,m} = 0 \quad \forall m \in \mathbf{M}, \forall r \in \mathbf{RES} \setminus \{\mathbf{BF} \cup \mathbf{E}\}, t \in \mathcal{T} \quad (\text{C.25})$$

As we can see, the variable $\mathbf{D}_{t,ELEC.,m}$ calculates the final electricity use which results from the sum of the electricity demand (constant in all months), the lighting demand, distributed over the months according to the % of lighting and the electricity loss. The low temperature heat demand is the sum of the annual demand for hot water (HW), evenly distributed over the year, and space heating (SH), distributed over the months according to the %sh. The percentage that multiplies the sum of the annual demand for hot water (HW) and space heating (SH) divides low temperature heat demand into centralized (district heating network (DHN)) and decentralized and is defined by the variable $\mathbf{G}_{t,\%dh_n}$. The high temperature heat and transport demand is evenly distributed among the months. Passenger mobility and freight demand are expressed in passenger-kilometers (pkms) and ton-kilometers (tkms), respectively. The variables $\mathbf{G}_{t,\%public}$ defines the penetration of public transport in passenger mobility and $\mathbf{G}_{t,\%rail}$ defines the penetration of rail in freight.

Layers are defined as all the elements in the system that need to be balanced in each month, such as resources and end-uses demand. For example, the electricity imported or produced in the system are layers that can be stored in hydroelectric dams or used as inputs to other energy conversion technologies (such as heat pumps) to meet the end-use demand for electricity which also represents a layer. Constraint (C.26) expresses the balance for each layer: all outputs from resources and technologies (including storage) are used to satisfy the end-uses-demand or as inputs to other resources and technologies. The matrix f defines for all technologies and resources outputs to (positive) and inputs from (negative) layers.

$$\sum_{i \in \mathbf{P} \cup \mathbf{RES} \setminus \mathbf{STO}} f_{i,l} \cdot \mathbf{z}_{t,i,m} + \sum_{u \in \mathbf{STO}} (\mathbf{Sto}_{t,u,l,m}^- - \mathbf{Sto}_{t,u,l,m}^+) - \mathbf{D}_{t,l,m} - a_l \cdot \mathbf{Loss}_{t,l,m} = 0 \quad \forall l \in \mathbf{L}, \forall m \in \mathbf{M}, t \in \mathcal{T} \quad (\text{C.26})$$

Losses (**Loss**) are considered for the electricity grid and for the DHN, constraint (C.27) calculates the amount of electricity that is lost from both produced and imported electricity in the corresponding layers.

$$\mathbf{Loss}_{t,eut,m} = \sum_{i \in \mathbf{RES} \cup \mathbf{P} \setminus \mathbf{STO}, f_{i,eut} > 0} f_{i,eut} \cdot \mathbf{z}_{t,i,m} \cdot \%loss_{eut} \quad \forall eut \in \mathbf{EUT}, \forall m \in \mathbf{M}, t \in \mathcal{T} \quad (\text{C.27})$$

The constraint (C.28) shows the calculation of GHG emissions. It is calculated as the sum of the products of the amount of resource used (local or imported), the duration of each month (hour) and the emission related

to the resource (gwp^{Op}).

$$\mathbf{GWP}_{t,r}^{Op} = \sum_{m \in \mathbf{M}} gwp_{r,m}^{Op} \cdot \mathbf{z}_{t,r,m} \cdot h_m \quad \forall r \in \mathbf{RES}, t \in \mathcal{T} \quad (\text{C.28})$$

In our model there is no emissions target, so only the equation allows us to estimate the amount of emissions produced by the system in each time period.

C.4 Additional constraints

Equations (C.1)-(C.28) define the main constraints of the energy model, which can be adapted or forced by adding constraints that limit the degrees of freedom of the model. Additional restrictions necessary for the specific case of the Swiss energy system are introduced below.

C.4.1 Long-term investment equations

Constraint (C.29) links linearly the storage capacity to the new installed power. Constraint (C.30) associates the cost of investment the *PowerToGas* unit to the maximum size of two conversion units. This constraint is displayed in a compact nonlinear formulation.

$$\mathbf{y}_{t,StoHydro} \leq f_{StoHydro}^{max} \frac{\mathbf{y}_{t,NewHydroDam} - f_{NewHydroDam}^{min}}{f_{NewHydroDam}^{max} - f_{NewHydroDam}^{min}}, \quad \forall t \in \mathcal{T} \quad (\text{C.29})$$

$$\mathbf{y}_{t,PowerToGas} = \max \{ \mathbf{y}_{t,PowerToGas}; \mathbf{y}_{t,GasToPower} \}, \quad \forall t \in \mathcal{T} \quad (\text{C.30})$$

Constraint (C.31) is used to calculate the energy efficiency as a fixed cost. Constraint (C.32) represent an additional investment cost of 9.4 billion CHF₂₀₁₅ is linked proportionally to the deployment of stochastic renewables. Constraint (C.33) links the DHN size to the total size of the installed centralized energy conversion technologies.

$$\mathbf{y}_{t,EFFICIENCY} = \frac{1}{1 + i_{rate}}, \quad \forall t \in \mathcal{T} \quad (\text{C.31})$$

$$\mathbf{y}_{t,Grid} \geq 1 + \frac{9400}{c_{t,Grid}^{Inv}} \frac{\mathbf{y}_{t,Wind} + \mathbf{y}_{t,PV}}{f_{Wind}^{max} + f_{PV}^{max}}, \quad \forall t \in \mathcal{T} \quad (\text{C.32})$$

$$\mathbf{y}_{t,DHN} \geq \sum_{p \in \mathbf{P-OEUT}_{\text{HEAT LT DHN}}} \mathbf{y}_{t,p}, \quad \forall t \in \mathcal{T} \quad (\text{C.33})$$

Constraint (C.34) complies with the decision of the Swiss government to phase out the nuclear power plants at the end of their useful life. Therefore there is no investment in nuclear plants over the entire planning horizon

$$\mathbf{y}_{t,NUCLEAR} = 0 \quad \forall t \in \mathcal{T} \quad (\text{C.34})$$

C.4.2 Short-term operating equations

Constraint (C.35) is used to avoid underestimating the cost of centralized heat production, a multiplication factor is introduced to account for peak demand, defined as a %*PeakDHN* times the maximum monthly average heat demand (also expressed in a compact non-linear formulation⁴).

$$\sum_{i \in \mathbf{P-OEUT}_{\text{HEAT LT DHN}}} \mathbf{x}_{t,i} \geq \%peak_{DHN} \max_{m \in \mathbf{M}} \{ \mathbf{D}_{t,\text{HEAT LT DHN},m} + \mathbf{Loss}_{t,\text{HEAT LT DHN},m} \}, \quad t \in \mathcal{T} \quad (\text{C.35})$$

⁴All equations expressed in a compact non-linear form in this section (Eqs. C.37 y C.42) can be easily linearized

Constraints (C.36)-(C.37) are complementary to constraint (C.8), as it expresses the minimum ($f^{min,\%}$) and maximum ($f^{max,\%}$) yearly output shares of each technology for each type of end-use demand.

$$\sum_{m \in \mathbf{M}} \mathbf{z}_{t,p,m} h_m \geq f_p^{min,\%} \sum_{p' \in \mathbf{P}-\mathbf{OF}-\mathbf{EUT}_{eut}} \sum_{m \in \mathbf{M}} \mathbf{z}_{t,p',m} h_m \quad \forall eut \in \mathbf{EUT}, \forall p \in \mathbf{P}-\mathbf{OF}-\mathbf{EUT}_{eut}, t \in \mathcal{T} \quad (\text{C.36})$$

$$\sum_{m \in \mathbf{M}} \mathbf{z}_{t,p,m} h_m \leq f_p^{max,\%} \sum_{p' \in \mathbf{P}-\mathbf{OF}-\mathbf{EUT}_{eut}} \sum_{m \in \mathbf{M}} \mathbf{z}_{t,p',m} h_m \quad \forall eut \in \mathbf{EUT}, \forall p \in \mathbf{P}-\mathbf{OF}-\mathbf{EUT}_{eut}, t \in \mathcal{T} \quad (\text{C.37})$$

Constraint (C.38) imposes that the share of the different technologies for mobility be the same in each month.

$$\mathbf{z}_{t,mob,m} \sum_{m' \in \mathbf{M}} h_{m'} \geq \sum_{m' \in \mathbf{M}} \mathbf{z}_{t,mob,m'} h_{m'} \quad \forall m \in \mathbf{M}, \forall mob \in \mathbf{MOB}, t \in \mathcal{T} \quad (\text{C.38})$$

Constraint (C.39) is a complement of constraint (C.29) and it is used for ensures that the shifted production in a given month does not exceed the electricity production by the dams in that month.

$$\mathbf{Sto}_{t,StoHydro,Elec,m}^+ \leq \mathbf{z}_{t,HydroDam,m} + \mathbf{z}_{t,NewHydroDam,m} \quad \forall m \in \mathbf{M}, t \in \mathcal{T} \quad (\text{C.39})$$

Appendix D

Assumptions for uncertain parameters

D.1 Uncertain parameters

Table D.1: Nominal, min and max values for end-uses demands in heating, electricity and mobility sectors. Abbreviations: Temperature (T), Hot Water (HW), Space Heating (SH). Parameter $eUI_{year,EUIdemands} [GWh]$

EUI demands	Year	Nominal	Min	Max
HEAT_HIGH_T	2020	21,535	13,283	29,786
	2030	19,144	11,219	27,070
	2040	17,413	8,475	26,350
	2050	15,925	7,174	24,676
	2060	13,864	4,770	22,958
HEAT_LOW_T_SH	2020	61,432	42,379	80,486
	2030	52,416	31,806	73,026
	2040	44,451	20,280	68,622
	2050	37,013	14,235	59,791
	2060	33,352	9,646	57,059
HEAT_LOW_T_HW	2020	11,116	8,860	13,372
	2030	11,575	8,797	14,354
	2040	11,891	7,848	15,935
	2050	12,179	7,386	16,973
	2060	12,567	6,879	18,255
ELECTRICITY	2020	34,829	26,441	43,217
	2030	34,754	24,561	44,948
	2040	35,469	21,207	49,731

Continued on next page

Table D.1: Nominal, min and max values for end-uses demands in heating, electricity and mobility sectors. Abbreviations: Temperature (T), Hot Water (HW), Space Heating (SH). Parameter $eUI_{year,EUIdemands} [GWh]$ (Continued)

	2050	37,001	19,929	54,074
	2060	35,322	17,312	53,332
LIGHTING	2020	6,149	4,635	7,663
	2030	5,495	4,037	6,953
	2040	5,202	3,264	7,140
	2050	5,000	2,977	7,023
	2060	4,961	2,910	7,013
	2020	131,144	115,239	147,049
MOBILITY_PASSENGER	2030	141,147	118,635	163,659
	2040	150,250	119,851	180,649
	2050	151,351	111,305	191,396
	2060	160,904	113,611	208,197
MOBILITY_FREIGHT	2020	33,917	29,227	38,607
	2030	38,776	33,001	44,551
	2040	41,255	32,516	49,995
	2050	41,950	29,500	54,399
	2060	45,619	31,145	60,093

Table D.2: Nominal, min and max values for resource costs. Abbreviations: Natural Gas (NG), Synthetic Natural Gas (SNG), Carbon Capture and Storage (CCS). Parameter $C_{year,resources}^{Op} [MCHF/GWh]$

Resources	Year	Nominal	Min	Max
ELECTRICITY	2010	0.0662	0.0564	0.1058
	2020	0.0757	0.0595	0.1287
	2030	0.0833	0.0585	0.1541
	2040	0.0909	0.0578	0.1736
	2050	0.0927	0.0554	0.1780
	2010	0.0709	0.0604	0.1134
GASOLINE/BIOETHANOL	2020	0.0775	0.0662	0.1243
	2030	0.0849	0.0618	0.1527
	2040	0.0920	0.0604	0.1757

Continued on next page

Table D.2: Nominal, min and max values for resource costs. Abbreviations: Natural Gas (NG), Synthetic Natural Gas (SNG), Carbon Capture and Storage (CCS). Parameter $C_{year,resources,.}^{Op}$ [MCHF/GWh] (Continued)

	2050	0.1002	0.0565	0.1964
	2010	0.0686	0.0399	0.1029
	2020	0.0752	0.0591	0.1279
DIESEL/BIODIESEL	2030	0.0822	0.0577	0.1520
	2040	0.0892	0.0567	0.1703
	2050	0.0972	0.0548	0.1904
	2010	0.0488	0.0455	0.0781
	2020	0.0535	0.0420	0.0909
LFO	2030	0.0584	0.0411	0.1081
	2040	0.0634	0.0403	0.1210
	2050	0.0690	0.0389	0.1332
	2010	0.0248	0.0212	0.0422
	2020	0.0288	0.0202	0.0518
NG/NG_CCS/SNG	2030	0.0329	0.0196	0.0608
	2040	0.0370	0.0187	0.0706
	2050	0.0414	0.0182	0.0811
	2010	0.0475	0.0421	0.0537
	2020	0.0658	0.0460	0.0710
WOOD	2030	0.0797	0.0436	0.0900
	2040	0.0937	0.0420	0.1246
	2050	0.0949	0.0407	0.1451
	2010	0.0206	0.0201	0.0371
	2020	0.0244	0.0192	0.0452
COAL/COAL_CCS	2030	0.0274	0.0189	0.0515
	2040	0.0303	0.0187	0.0573
	2050	0.0306	0.0183	0.0579
	2010	0.0038	0.0032	0.0046
	2020	0.0039	0.0031	0.0051
URANIUM	2030	0.0040	0.0028	0.0064
	2040	0.0041	0.0026	0.0081
	2050	0.0041	0.0024	0.0081

Appendix E

Modeling the price of resources and demands

E.1 First-order autoregressive process

Figure E.1 below displays the distribution of 30,000 scenarios of demands of electricity, heating and lighting for different values of α using the AR processes. The figure also highlights seven random trajectories of the processes (in blue color). Figure E.2 is similar but results are displayed for energy prices.

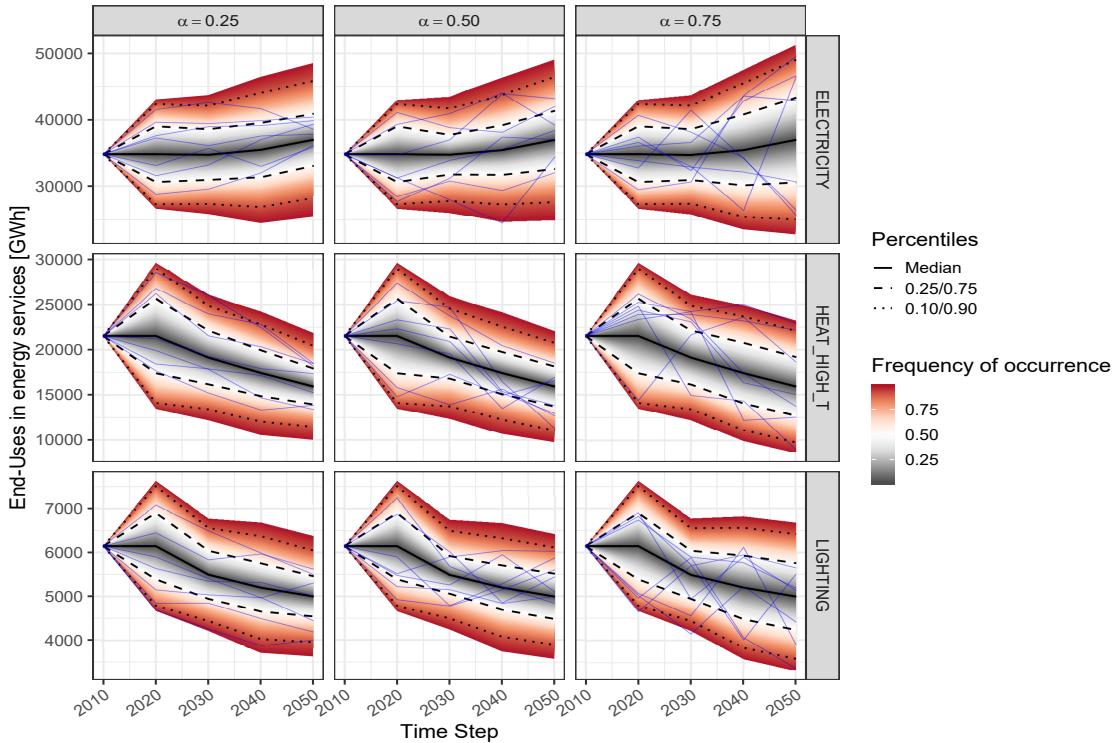


Figure E.1: 30,000 scenarios generated from the original AR processes with $\alpha \in 0.75, 0.50, 0.25$ for demand of ELECTRICITY, HEAT_HIGH_T. and LIGHTING.

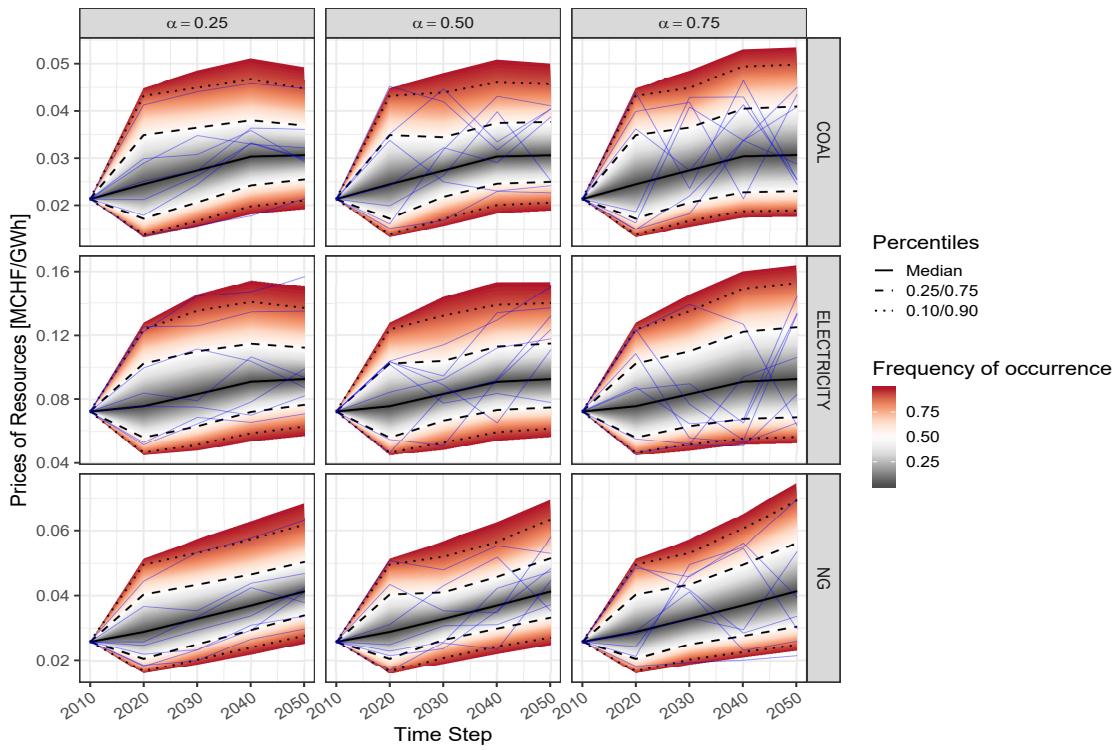


Figure E.2: 30,000 scenarios generated from the original AR processes with $\alpha \in 0.75, 0.50, 0.25$ for price of COAL, ELECTRICITY and NG.

Appendix F

Markov chains for electricity and natural gas prices

F.1 Transition probability matrix of a three-state Markov chain

Figures F.1 and F.2 display the Markov chains obtained from discretizing the electricity and gas prices, for different values of the parameter α .

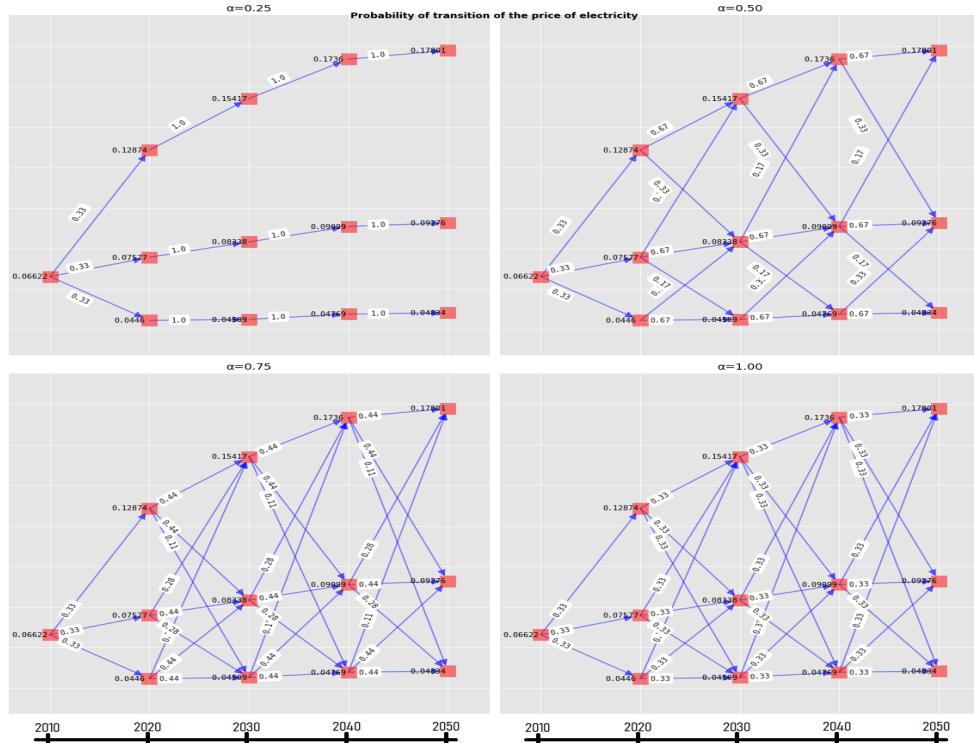
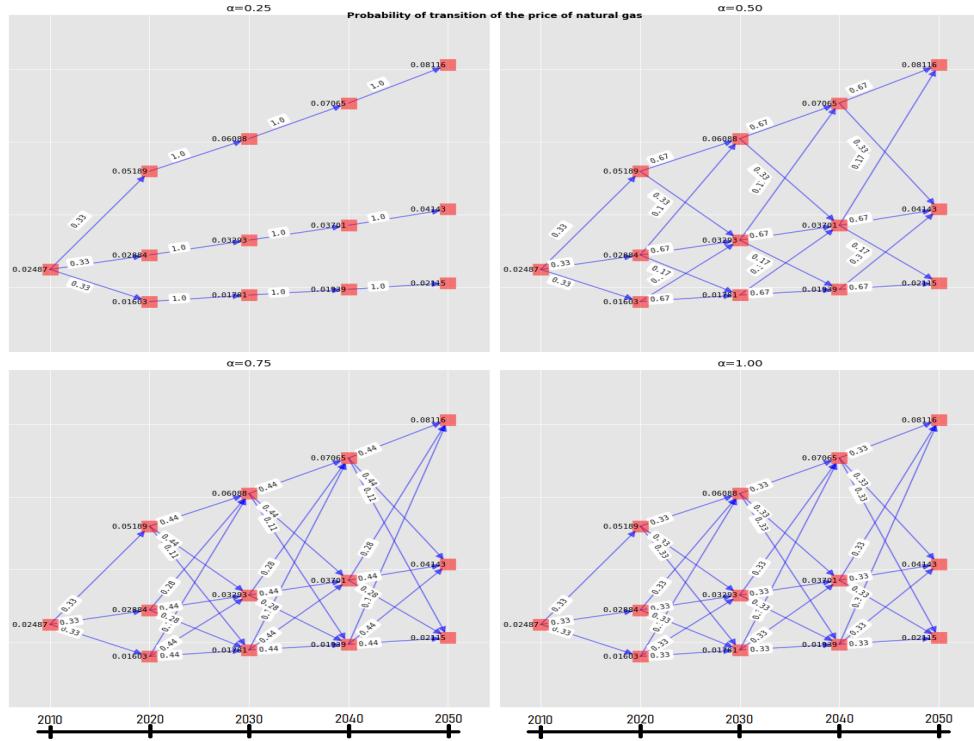


Figure F.1: Markov Chains for electricity prices for different values of α .

Figure F.2: Markov Chains for Natural Gas prices for different values of α .