UNIVERSITÀ DEGLI STUDI DI BERGAMO
SCHOOL OF ENGINEERING
DEPARTMENT OF ENGINEERING AND APPLIED SCIENCES
DOCTORAL PROGRAMME IN ENGINEERING AND APPLIED SCIENCES

MATHEMATICAL MODELS AND SOLUTION ALGORITHMS FOR GENERATION AND TRANSMISSION EXPANSION PLANNING WITH HIGH SHARES OF RENEWABLES

Doctoral Dissertation of
Giovanni Micheli

Supervisor
Prof.ssa Maria Teresa Vespucci

The Chair of the Doctoral Programme
Prof. Valerio Re

A.A. 2019/2020 – XXXIII Cycle

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This chapter provides an introduction to the thesis work reported in this dissertation. First, we present an overview of the generation and transmission expansion planning problem and we motivate our research work in this field. Second, a literature review of the topic pertaining to this thesis is provided. Finally, the objectives of the dissertation as well as the thesis structure are illustrated.

1.1 Thesis motivation

The generation and transmission expansion planning (GTEP) problem aims at determining the evolution of power systems over a long-term planning horizon, by defining technology, capacity and location of new generating units, as well as new interconnections to be built. The definition of joint expansion plans is one of the most relevant problems in the field of power systems. Indeed, this kind of analysis provides a lot of useful information, allowing for instance to study the impact of some policy decisions and the possibility to achieve targets such as decarbonisation, integration of large shares of renewables or reduction of CO_2 emissions.

Many different agents are interested in the joint generation and transmission expansion planning. Indeed, in all countries where the unbundling of the energy sector is still ongoing, this problem is addressed by vertically integrated monopoly utilities to establish a strategic master plan, to secure long-term supply, promoting affordable and reliable electricity and reducing power outages. Instead, in an unbundled energy environment, expansion planning analysis is performed by transmission planners (e.g., ISO) for what is called *anticipative planning*, in which those planners determine the expansion plan that is optimal for the energy system as a whole to identify the best network reinforcement and to set incentives that could induce generation companies to invest in a socially efficient manner [1, 2, 3].

The GTEP problem also has great relevance in operations research, being generally addressed through the formulation of large-scale MILP decision models, which require the development of dedicated solving algorithms.

This dissertation is mainly concerned with the GTEP for the transition to low emission power systems. Specifically, the achievement of such challenging sustainability goals requires installing large shares of renewable power capacity in power systems. In the presence of high levels of renewable penetration, power systems face great challenges to meet the demand, because of the unpredictable daily and seasonal nature of renewable generation. Such a variability has to be managed in the short-term to ensure that power systems are operated in an efficient and reliable way.

The need to install large shares of renewables in power systems motivates the development of a mathematical tool that could find the optimal capacity expansion strategy for a power system, while evaluating the achievement of the challenging decarbonisation targets. Such a tool must be capable of capturing the short-term variability and addressing all the issues related to integrating large shares of intermittent renewable power sources.

Moreover, addressing the GTEP problem for modern power systems requires facing several challenges. First, investments in power facilities are capital-intensive decisions, which are usually characterized by expected lifetimes greater than 30-50 years. Thus, when planning the expansion of a power system, planning horizons of several decades are typically applied. Second, in the presence of high levels of renewable penetration, the GTEP analysis requires considering a high level of temporal detail, evaluating power system operation with an hourly resolution, in order to catch the fluctuation of wind and solar power production. Third, the GTEP problem is characterized by a high level of uncertainty: indeed, since expansion plans are usually provided for a long-term planning horizon, the future system conditions are generally uncertain at the time the expansion plans are decided. Fourth, in addressing the expansion planning, many interconnections between power systems and other sectors, such as the fossil fuel sectors, should be considered in order to define more efficient solutions for the whole energy system.

Because of the long-term horizon, the hourly resolution, the high uncertainty and the need to model interactions with other sectors, GTEP models are large-scale problems, whose solution is computationally challenging. The high dimensionality of GTEP models motivates the research in the field of advanced algorithms that could solve these large-scale models to the desired accuracy.

1.2 Literature review

GTEP problems have received considerable attention in the last three decades and many mathematical programming models have been proposed to address different research questions. Works concerning the investment planning in the power sector can be classified in several ways. This section reviews the existing literature for GTEP analysis, by introducing the basic choices associated with the design of a GTEP model, which include three main features: (i) sectors to be included in the analysis; (ii) formulation structure; and (iii) representation of the transmission network.

1.2.1 Sectors coverage

When designing a model for GTEP, the first decision to be taken concerns the choice of which sectors should be included in the analysis and which other sectors should be exogenously modeled. Indeed, there exist many interconnections between power systems and several sectors, such as the transportation sector, the heating sector or the fossil fuels sectors. In terms of sectors coverage, two different approaches are possible:

1. Perform a single sectoral analysis, focusing only on power sector and considering the interactions with other sectors as boundary conditions.
2. Consider multiple sectors to study the high-level interactions within the whole energy system.

Considering multiple sectors allows evaluating the interdependencies between sectors, usually finding more efficient solutions for the whole energy system. However, the drawback of this approach is an increasing computational cost. To maintain the problem computationally tractable, multi-sector models usually introduce strong approximations in the way single sectors are modeled. For instance, reference [4] proposes an equilibrium model to plan infrastructure investments in the whole energy system to satisfy the future energy requirements of three different demand sectors, namely the industry sector, the residential sector and the transportation sector, while considering climate policies and governmental regulation. A very wide geographical scope is considered in this analysis, modeling the whole world through 30 nodes, of which 15 are European countries and the remaining 15 model the rest of the world, while the planning horizon of 40 years, from 2010 up to 2050, is discretized in 10-year steps. The energy requirements of demand sectors are satisfied by using various energy carriers (fuels) depending on relative costs, efficiencies, as well as regulatory and technical

constraints. The fuels included in the analysis are: crude oil, oil products, natural gas, coal, lignite, electricity, biofuels, renewables, and (as an input for power generation only) nuclear and hydro. In this work, a very simplified representation of the power sector is considered: indeed, the power sector is modeled as a transformation process, converting input fuels into electricity, considering only capacity constraints stated in terms of output quantities.

When a high level of technical detail is required, the multi-sector models can be applied only by limiting the geographical scope. For instance, authors in [5] perform a long-term investment planning analysis on the Danish heat and electricity sector, considering the uncertainty related to the wind power production, which is modeled as a stochastic parameter. In this paper, a more detailed representation of the power sector is considered, including some technical constraints, such as peaking reserve constraints, and considering different scenarios for the wind power availability. However, the application of this analysis is limited to a single country, i.e., Denmark.

From the works above, it can be noticed the existence of a trade-off between sectors coverage, geographical scope and technical detail: a multi-sector analysis can be performed only by limiting the geographical scope or by considering a low level of technical detail. Instead, by exogenously modeling all sectors but the power sector it is possible to consider technical and geographical characteristics in greater detail. The choice of the approach to adopt depends on the objective of the analysis: the multi-sector approach is to be applied when the research focus is to evaluate long-term trends, such as decarbonisation pathways, taking into account the interactions between different economic sectors and the energy system; the single sectoral analysis is suited to plan investments in the power sector, since it allows a more detailed representation of the power system, as well as of the short-term operational dispatch, and therefore provides more reliable decisions to the actors involved in the GTEP problem.

In this chapter, we will focus on single sectoral analysis, by introducing several approaches designed to perform a GTEP analysis exogenously modeling all sectors but the power sector. In such models, the feedback effects between sectors are ignored and parameters like fuel prices and fuels availability are exogenous parameters rather than decision variables. However, this choice allows providing a more detailed representation of power systems, which is required when modeling both long-term investment decisions and short-term operational dispatch.

1.2.2 Formulation structure

Works that address the GTEP problem focusing only on power sector modeling can be further divided in two categories according to the problem formulation structure, which determines the decision variables to be introduced in the analysis. The first class includes decoupled models, i.e., researches that only address either generation expansion or transmission expansion. Specifically, many studies deal with only transmission expansion planning (TEP), a problem that is typically addressed by Transmission System Operators to identify the optimal transmission reinforcements to be carried out with the aim of facilitating energy exchange among producers and consumers. In such models, generation expansion decisions are exogenous parameters rather than model variables. Pioneering work in this area is due to Garver [6], who in 1970 proposed a linear programming problem determining the transmission expansion plans based on the location of overloads. Since then, many relevant contributions based on mathematical programming have been produced. For instance, in [7] authors deal with the TEP problem considering the integration of large-scale wind power production, while in [8] a bi-level TEP model using conic AC power flow formulation is presented.

Many other studies only address generation expansion planning (GEP), evaluating the adequacy of generating facilities used to supply load and analysing whether it is necessary to build new power plants. The GEP problem is typically motivated by the growth of the demand for electricity and by the aging of existing generating facilities. For a very detailed review of GEP models we refer the reader to [9], which provides a comprehensive description of the state-of-the-art of recently developed approaches dealing with the GEP problem, organizing them into seven key categories including the interaction of generation expansion planning with: (i) the transmission expansion planning; (ii) natural gas system; (iii) short-term operation of power markets; (iv) electric vehicles; (v) demand-side management and storage; (vi) risk-based decision-making; and (vii) applied energy policy including security of supply. As regards to the transmission network, in some GEP models transmission constraints are totally neglected. For instance, reference [10] proposes a mixed integer non-linear programming model to perform a generation expansion planning analysis minimizing the planning cost and environmental pollution at the same time, while considering energy storage systems. In such a paper, the test system consists of an isolated system, which does not consider any representation of the transmission network. Instead, other works that address GEP include also the transmission network in the analysis by considering the network configuration as an exogenous parameter. Examples of these

works are represented by references [11, 12]. Specifically, in [11] a multi-objective generation expansion problem is considered taking into account costs, environmental impacts and portfolio investment risk. The proposed model decides the location of the planned generation units in a multi-period planning horizon by minimizing simultaneously costs, environmental impact, imported fuel and fuel price risks. In [12] the generation expansion problem is instead addressed considering a market framework. In this work, the strategic behaviour of the producer is represented through a bi-level model: the upper-level considers both investment decisions and strategic production actions and the lower-level corresponds to market clearing.

Studies in the second class optimize both the generation and the transmission expansion plan. Indeed, if the demand for electricity in a zone can be supplied with both local generation and transportation of power from other zones, generation and transmission decisions are substitutes and have to be simultaneously considered in the optimization process [3]. By endogenously modeling the interactions between generation and transmission expansion decisions, the joint GTEP analysis provides solutions that are less expensive than expansion plans provided by the decoupled models.

Two main approaches have been proposed to address the joint GTEP: centralized and decentralized models. Specifically, decentralized models consider that power systems consist of multiple decision makers with different objectives. Decentralized models are usually formulated as equilibrium problems, with several participants maximizing their own objective. For instance, in [13] the GTEP problem is formulated as a tri-level problem, consisting of the pool-based market, the generation system and the transmission system. In [14], authors propose mathematical models for sequential coordination of transmission expansion planning with strategic generation investments. The interaction between transmission company and strategic generation companies is modeled using the sequential-move game, while the interaction between the strategic generation companies is modeled as a simultaneous-move game.

Centralized models instead approach the GTEP problem as just one problem, with a unique decision maker, such as the authority or the ministry of energy, taking all the relevant decisions. Although this approach does not reflect the real structure of modern power systems, consisting of several actors involved, centralized models are widely used to perform the so-called anticipative planning, whose objective is to identify policies and incentives that could induce generation companies to invest in a socially efficient manner. Several contributions have been developed in this area. For example, reference [15] proposes a mixed integer linear programming formulation for generation and

transmission planning considering the value of lost load. A centralized approach is also adopted in [16] to study the effect of wind speed's spatial distribution on the simultaneous generation and transmission expansion planning of power systems including wind farms. Specifically, in such a work investment decisions are defined by means of a mixed integer linear programming model that minimizes the sum of total investment cost of thermal generation units, transmission lines and wind farms, operation cost of thermal generators, and loss cost of transmission lines. The wind generation is also accurately considered in [17], where a mixed integer linear programming model is proposed to study the GTEP problem with a high wind power penetration rate in large-scale power grids. Also in [18, 19, 20], the co-optimization problem is formulated as a mixed integer programming model, with the addition of reliability constraints enforced iteratively. Finally, authors in [21] propose a multi-objective, multi-area and multi-stage model to long-term expansion planning of integrated generation and transmission facilities. The proposed model considers three objectives: (i) minimization of investments and operation costs of power generation and transmission facilities; (ii) minimization of Life-Cycle Greenhouse Gas Emissions; and (iii) maximization of the diversification of electricity generation mix.

Both the centralized and the decentralized approaches are useful to analyse optimal policies in the power sector. The two approaches are complementary to each other and are taken into account in separate phases of the decision process, using different models. Specifically, the centralized modeling is the approach adopted by regulators to search for optimal policies, modeling the power system and the short-term operations in great details, but considering a very simplified representation of market aspects. By contrast, decentralized models allow to validate policies, by reducing the technical detail of the analysis to focus on the interactions between different agents involved in the liberalized power sector. Because of computational restrictions, models with a detailed representation of both power system technical operations and market aspects cannot be applied to plan investment decisions in real-scale power systems. Due to our interest in finding optimal policies for the evolution of power systems, rather than in validating the outcomes of already specified policies, our research is focused on centralized models. For a detailed description of market aspects, we refer the reader to [22] and the references contained in it.

Beside the choice between centralized and decentralized models, when addressing the joint GTEP problem decision variables representing infrastructure investments in generation and transmission capacity for the power system have to be introduced. Different types of variables can be used according to the technical detail of the analysis.

Specifically, some models introduce binary variables to represent the selection of discrete facilities within a set of candidate generating plants and candidate transmission lines. This approach is typically adopted in studies that require a very high level of technical detail, modeling the operations of every power plant. Instead, most of the works in the literature employ linear decision variables, representing the total aggregated capacity per technology at each node, meaning that traditional unit commitment constraints are neglected. According to the objective of the analysis, ignoring unit commitment constraints could be a very restrictive approximation for expansion planning models. Specifically, when studying power systems with large shares of renewables it is necessary to plan also the investments in new flexible resources that could respond to the variability and uncertainty of stochastic generation. Ignoring unit commitment constraints leads to the impossibility of properly evaluating such a flexibility and, consequently, to underestimate the required new generation capacity as well as the system costs, as shown in [23].

1.2.3 Transmission network

In a GTEP analysis transmission network can be modeled in different ways. In the order of increasing realism, the available approaches include the transportation model, the Direct Current (DC) power flow model and the Alternating Current (AC) power flow model. As discussed in [3], in the transportation model transmission lines are characterized by an efficiency parameter describing transportation losses. Energy flows on transmission lines are subject to capacity limits and nodal power balances are imposed at every node, while Kirchhoff's voltage law is ignored. A more realistic modeling approach is the DC power flow model, which, in addition to transmission flow limits and nodal power balances, consists of a linear relation between power flows and voltage angle differences. Although the equations that describe the DC power flow model are linear, when integrating this formulation in the GTEP problem the resulting model is a non-linear and non-convex problem. Indeed, in such a formulation binary variables are introduced to model investment decisions in candidate transmission lines and the voltage angles are divided by line reactances, which are both decision variables. Thus, the DC power flow model provides a greater fidelity but at the cost of an increasing computational complexity. Finally, the most realistic representation of the transmission network is the AC power flow model, which is strongly non-linear since it models both active and reactive power flows in the transmission grid through non-linear constraints that involve nodal voltages and angles and line impedances. Although the AC model provides the best representation of power flows, the complexity in this approach is

further increased: the trade-off between modeling detail and computational complexity holds also for the choice of the transmission network modeling approach. Usually, when a detailed geographical scope is considered, modeling every bus of the transmission network, AC and DC models are more appropriate choices, while the transportation approach is generally preferred with higher geographical scopes, such as when nodes of the transmission network represent zones or countries.

In the literature there exist also some hybrid approaches. For instance, reference [24] uses an iterative approach where investment decisions are optimized using a transportation model, followed by a separate load flow model for grid expansion to check the feasibility and the robustness of the solution provided.

1.3 Thesis objectives

Due to computational restrictions, most of the existing planning tools for the GTEP employ a low level of temporal and technical detail, evaluating system operations on a daily or weekly basis and ignoring unit commitment constraints. However, by providing an approximate representation of power system operations, such tools cannot capture the short-term volatility, overestimating the renewable capacity and underestimating the need for flexible resources and the expected costs.

The main objective of this thesis is to develop a novel computational tool to perform the GTEP analysis in scenarios with large shares of renewables. Such a tool is specifically designed to plan the joint capacity expansion of energy systems evaluating the achievement of decarbonisation targets set by the European Commission. Indeed, the distinct feature of the proposed analysis is the very detailed representation of the short-term operations, which is needed to accurately address all the challenges related to integrating high shares of intermittent renewable energy sources.

This dissertation summarizes the research activities performed to develop the novel computational tool. Specifically, this document is organized as follows.

Chapter 2 introduces the first activity performed in this research project, aimed at providing a detailed representation of power system operations, while considering the computational burden of the problem. As described in the chapter, accurate estimates of the short-term operations for computationally tractable problems are obtained by including a tight formulation for the thermal unit commitment problem and by selecting a small set of representative days, rather than considering every day of the planning horizon.

Chapter 3 integrates the short-term model formulated in Chapter 2 into the expansion planning framework, by proposing a deterministic MILP model designed to support government authorities in generation and transmission expansion planning. Such a model optimizes strategic decisions including retirement of existing capacity and investments in new generation, transmission and storage facilities, as well as operational decisions.

Chapter 4 deals with the inclusion of the long-term uncertainties in the decision-making framework, formulating the expansion planning problem as a two-stage stochastic MILP model, being investment decisions first-stage variables and operational decisions second-stage variables. Due to the high dimensionality, the inclusion of the stochasticity makes the GTEP models computationally intractable, requiring the application of advanced algorithms to solve these large-scale problems. Thus, our solution algorithm based on the multi-cut Benders decomposition approach is introduced in the chapter.

Chapter 5 addresses the expansion planning of integrated electricity and gas systems. The stochastic model and the solution algorithm introduced in the previous chapter are here modified to include equations describing the natural gas system operations, modeling a bi-directional energy conversion between electricity and gas. Since the inclusion of the gas system in the decision making framework further increases the computational burden of the expansion planning model, in this chapter also a clustered unit commitment formulation is presented, by grouping similar thermal power plants into clusters.

Chapter 6 deals with the modeling of demand elasticity, by including demand response programs in the expansion planning analysis. Two different reactions of customers to electricity prices are modeled, namely the load shifting and the peak shaving. Then, an optimization model to plan investments also in demand response devices is introduced.

Chapter 7 concludes this dissertation providing a summary, several relevant conclusions and suggestions for future research.

Appendix provides the list of dissemination activities connected to the thesis work, which include publications in international journals and conference proceedings, as well as presentations at international conferences.

The research activities described in this thesis were carried out in the context of a three-year collaboration with CESI S.p.A. (Centro Elettrotecnico Sperimentale Italiano, Italian Electrical and Technical Experimental Centre), a major industrial research centre on power systems in Italy. Specifically, the development of a computational model for

the expansion planning analysis has a high strategic importance for CESI S.p.A. and more in general for all the companies providing worldwide consultancy services in the power sector, in order to support governments and regulatory authorities in defining strategic investment decisions. For instance, the solution of the GTEP model has a crucial importance in:

- Establishing a strategy and a master plan for developing countries, in order to secure long-term supply, to promote affordable and reliable electricity and reducing power outages;
- Determining the best choice between the trade-offs of selling electricity or natural gas;
- Supporting investors developing regional trade of electrical energy and gas.

The research activities described in this dissertation are also the result of the collaboration with Professor Andrés Ramos, from the IIT Department at the Comillas University, mainly regarding the inclusion of the stochasticity in the decision making framework.

Chapter 2

Evaluating power systems short-term operations

2.1 Research motivation

Generation and transmission expansion planning to achieve decarbonisation targets in the power sector requires installing relevant shares of production from renewable energy sources. In the presence of high shares of non-dispatchable renewable sources, power system operations must be evaluated with an hourly resolution to take into account the different operating conditions due to the intermittency of renewable power generation. However, investment decisions in the power sector are long-term decisions: when addressing the generation and transmission expansion problem, planning horizon of several decades are typically applied.

Due to the long-term horizon, providing an hourly resolution to the expansion planning problem is computationally infeasible. In order to limit computational costs, most of the existing planning tools employ a low level of temporal detail, evaluating system operations on a daily or weekly basis. However, as shown in [25], the low level of temporal detail can significantly affect the results, especially in a context of large penetration of renewables. Indeed, planning tools that employ a low level of temporal detail cannot capture the short-term volatility, overestimating the renewable capacity and, thus, resulting not suited to accurately study all the challenges related to integrating high shares of intermittent energy sources.

Moreover, most of the existing models for power generation and transmission expansion planning consider also a low level of technical detail, generally modeling thermal capacity expansion through continuous variables and, thus, without considering unit commitment constraints on a plant-by-plant level. However, as discussed in [23], this approach, although very common in the literature, is no more an appropriate approximation for expansion planning models. Indeed, due to the increasing penetration of intermittent renewable energy sources in power systems, the need is growing for flexible resources that could respond to the variability and uncertainty of stochastic generation. Ignoring unit commitment constraints leads to the impossibility of properly

evaluating such a flexibility and, consequently, to underestimate the required new generation capacity as well as the system costs.

To provide reliable expansion plans for large-scale energy systems with high shares of renewables, a high level of both technical and temporal detail has to be considered at operational level. This chapter describes the analysis performed to obtain accurate estimates of the operational decisions in the expansion planning framework while keeping the problem computationally tractable. Specifically, a high level of technical detail is included in the analysis by considering unit commitment constraints, while a high level of temporal detail is obtained by working with representative days, discretized in hours. However, the use of representative days raises the crucial issues regarding: (i) how to consider the seasonality of hydroelectric dispatch; and (ii) how to set the initial ON/OFF status of thermal power plants in representative days. Both these crucial issues are addressed in this chapter.

Specifically, the structure of the chapter is as follows. Section 2.2 introduces the thermal unit commitment problem. The selection of representative days is addressed in Section 2.3. Section 2.4 proposes two different approaches to capture the seasonality of the hydroelectric dispatch when working with representative days. Section 2.5 describes the proposed method to determine the initial ON/OFF status of thermal power plants in representative days. Section 2.6 summarizes contributions provided in this chapter by formulating a Mixed-Integer Linear Programming model to evaluate short-term operations through representative days. The results of different case studies are presented in Section 2.7. Finally, Section 2.8 concludes the chapter.

2.2 The thermal unit commitment problem

This section summarizes the comparison of different approaches to model the thermal unit commitment problem within the expansion planning framework. Indeed, for this problem several formulations have been proposed in the literature, that mainly differ in the number of binary variables used and in the way the minimum up/down time constraints are expressed. Since different model formulations may present different solution times, an analysis has been conducted to identify the computationally most efficient formulation for our specific application. In particular, we have considered three formulations compatible with our approach to the thermal unit commitment problem and analysed their solution times on a case study of suitable dimension.

The structure of this section is as follows. Section 2.2.1 describes the thermal unit commitment problem and the assumptions introduced in our analysis. The notation needed to formulate the thermal unit commitment problem is introduced in Section 2.2.2. Section 2.2.3 introduces the formulation that proved to be the most efficient one in our specific application. Alternative formulations are introduced in Section 2.2.4. Finally, results from testing the different formulations are presented in Section 0.

2.2.1 Problem description

The thermal unit commitment problem consists in scheduling the activation of a set of thermal power plants to supply at minimum cost the demand for electricity while respecting some technical constraints. In its general form, the thermal unit commitment problem is formulated as a Mixed-Integer Non-Linear Programming (MINLP) problem, since it includes binary variables to describe the ON/OFF status of thermal power plants and non-linear constraints, mainly introduced to model cost functions or flows in the transmission network. For the thermal unit commitment problem there exists a huge literature, whose review is outside the scope of this dissertation. We refer the reader to [26] and [27] for detailed reviews of the problem.

The thermal unit commitment problem we consider in this work is formulated as a large-scale Mixed-Integer Linear Programming (MILP) problem, based on the following assumptions:

- A zonal representation of the power system is adopted, i.e., power system is partitioned into different price zones that exchange energy through a capacitated power network.
- Time is represented with hourly resolution.
- Thermal power plants flexibilities are considered in the analysis by including *minimum up/down time constraints*, which force thermal power plants to stay ON/OFF for a given amount of hours after each switching manoeuvre, and *start-up costs*, which are additional costs incurred to ignite and heat up thermal power plants. Instead, *ramping constraints*, which bound the change of power production in consecutive hours, are neglected in this work. Indeed, CESI S.p.A. considers minimum up/down time constraints and start-up costs to be sufficient to model the flexibility of thermal power plants. However, given the linear formulation of ramping constraints, these restrictions could be easily integrated in the current formulation of the unit commitment problem.

- Thermal production costs are supposed to be linear functions of the power output, being CM_k the slopes of these linear relationships. Indeed, piece-wise linear or quadratic formulations for the cost-functions would significantly increase the complexity of the problem while only marginally improving the accuracy.

2.2.2 Notation

To formulate the thermal unit commitment problem, the following notation is introduced.

Sets

$\mathcal{Z}$	Set of zones, indexed by z
$\mathcal{K}$	Set of thermal power plants, indexed by k
$\Omega_z^k \subset \mathcal{K}$	Set of thermal power plants located in zone z
$\mathcal{T}$	Set of hours, indexed by t and τ
$\mathcal{L}$	Set of transmission lines, indexed by l
$rz(l)$	Receiving-end zone of transmission line l
$sz(l)$	Sending-end zone of transmission line l

Parameters

c_{ENP}	[€/MWh]	Penalty for energy not provided
c_{OG}	[€/MWh]	Penalty for over-generation
CM_k	[€/MWh]	Marginal production cost of thermal power plant k
$\underline{P}_k$	[MW]	Minimum power output of thermal power plant k
$\overline{P}_k$	[MW]	Maximum power produced by thermal plant k
SUC_k	[€]	Start-up cost of thermal power plant k
γ_{k_0}	[−]	Initial ON/OFF status of thermal power plant k
MUT_k	[h]	Minimum up time of thermal power plant k
MDT_k	[h]	Minimum down time of thermal power plant k
$\underline{F}_l$	[MW]	Minimum capacity of transmission line l
$\overline{F}_l$	[MW]	Maximum capacity of transmission line l
$NL_{z,t}$	[MW]	Net load in zone z in hour t
$R_{z,t}$	[MW]	Reserve requirement for zone z in hour t

Variables

$\gamma_{k,t}$	[−]	Commitment status of unit k at time t: $\gamma_{k,t} = 1$, if unit k is online at time t and $\gamma_{k,t} = 0$, otherwise
$\alpha_{k,t}$	[−]	Start-up status of unit k at time t: $\alpha_{k,t} = 1$, if unit k is started up at time t and $\alpha_{k,t} = 0$, otherwise
$\beta_{k,t}$	[−]	Shut-down status unit k at time t: $\beta_{k,t} = 1$, if unit k is shut down at time t and $\beta_{k,t} = 0$, otherwise
$p_{k,t}$	[MW]	Power production of thermal power plant k at time t above its minimum output $\underline{P}_k$
$x_{l,t}$	[MW]	Power flow on transmission line l in hour t
$ENP_{z,t}$	[MWh]	Energy not provided in zone z in hour t
$OG_{z,t}$	[MWh]	Over-generation in zone z in hour t

2.2.3 Mathematical formulation

The thermal unit commitment problem is formulated as the following MILP model

$$\min z = \sum_{k\in K} CM_k \sum_{t\in T}\left(\gamma_{k,t}\,\underline{P}_k + p_{k,t}\right) + \sum_{k\in K} SUC_k \sum_{t\in T} \alpha_{k,t} + c_{ENP}\sum_{z\in Z}\sum_{t\in T} ENP_{z,t} + c_{OG}\sum_{z\in Z}\sum_{t\in T} OG_{z,t} \tag{2.1}$$

subject to

$$p_{k,t} \le \gamma_{k,t}\left(\overline{P}_k - \underline{P}_k\right) \qquad k \in \mathcal{K}, t \in \mathcal{T} \tag{2.2}$$

$$\gamma_{k,t} - \gamma_{k_0} = \alpha_{k,t} - \beta_{k,t} \qquad k \in \mathcal{K}, t = 1 \tag{2.3}$$

$$\gamma_{k,t} - \gamma_{k,t-1} = \alpha_{k,t} - \beta_{k,t} \qquad k \in \mathcal{K}, t > 1 \tag{2.4}$$

$$\sum_{\tau=t-MUT_k+1}^{t} \alpha_{k,\tau} \le \gamma_{k,t} \qquad k \in \mathcal{K}, t > MUT_k \tag{2.5}$$

$$\sum_{\tau=t-MDT_k+1}^{t} \beta_{k,\tau} \le 1 - \gamma_{k,t} \qquad k \in \mathcal{K}, t > MDT_k \tag{2.6}$$

$$\underline{F}_l \le x_{l,t} \le \overline{F}_l \qquad l \in \mathcal{L}, t \in \mathcal{T} \tag{2.7}$$

$$\sum_{k\epsilon\Omega_z^k}\left(\underline{P}_k\,\gamma_{k,t}+p_{k,t}\right)+\sum_{l|rz(l)=z}x_{l,t}+ENP_{z,t}=NL_{z,t}+\sum_{l|sz(l)=z}x_{l,t}+OG_{z,t,w}$$
$$z\in\mathcal{Z},t\in\mathcal{T} \qquad (2.8)$$

$$\sum_{k\epsilon\Omega_z}\left[\left(\overline{P}_k-\underline{P}_k\right)\gamma_{k,t}-p_{k,t}\right]\geq R_{z,t} \qquad z\in\mathcal{Z},t\in\mathcal{T} \qquad (2.9)$$

$$\gamma_{k,t}\,,\alpha_{k,t}\,,\beta_{k,t}\in\{0,1\} \qquad k\in\mathcal{K},t\in\mathcal{T} \qquad (2.10)$$

$$p_{k,t}\geq 0 \qquad k\in\mathcal{K},t\in\mathcal{T} \qquad (2.11)$$

$$x_{l,t}\ \text{free variable} \qquad l\in\mathcal{L},t\in\mathcal{T} \qquad (2.12)$$

$$ENP_{z,t},OG_{z,t}\geq 0 \qquad z\in\mathcal{Z},t\in\mathcal{T}. \qquad (2.13)$$

In particular, the objective function (2.1) is the sum of four terms: (i) production costs, considered as linear functions of the power produced by thermal plants; (ii) start-up costs; (iii) penalties for energy not provided; and (iv) penalties for over-generation.

Constraints (2.2) state that the maximum power above the minimum is either bounded above by $\overline{P}_k-\underline{P}_k$, if unit k is online ($\gamma_{k,t}=1$), or zero if unit k is offline ($\gamma_{k,t}=0$). Constraints (2.3) and (2.4) enforce consistency between binary variables that represent start-up, shut down and statuses in adjacent hours. Specifically, while equations (2.3) hold only for the first hour of the planning horizon, constraints (2.4) are imposed in all the hours of the planning horizon, except the first one. Inequalities (2.5) are the minimum up time constraints and they impose that in an interval of MUT_k consecutive time periods a unit can be started-up at most once. Inequalities (2.6) work similarly for the shut-down case and they are referred to as minimum down time constraints. Inequalities (2.7) impose upper and lower bounds for the power flows on transmission lines. Equations (2.8) ensure the generation-demand balance at each zone of the system: the left-hand side represents the hourly energy sources of zone z (given by thermal generation and incoming energy flows) and the right-hand side describes the energy uses (represented by the net load and outgoing energy flows). The variables $ENP_{z,t}$ and $OG_{z,t}$ allow detecting and evaluating problems in the simulated system that can cause a mismatch between supply and demand. The spinning reserve is the amount of unused capacity in online power plants which can compensate for power shortages or frequency drops within a given period of time. Inequalities (2.9) guarantee the fulfilment of zonal

upward reserve requirements provided by thermal power plants. Finally, constraints (2.10)–(2.13) define the optimization variables.

2.2.4 Alternative formulations

Different formulations for the thermal unit commitment problem have been proposed in the literature. We refer the reader to [28] and [29] for a detailed comparison between them. Specifically, other formulations in the literature differ from the model proposed in Section 2.2.3 for two aspects: (i) the way the minimum up/down time constraints are imposed; and (ii) the number of binary variables introduced.

Indeed, minimum up time constraints (2.5) can be equivalently expressed with the following inequalities:

$$\sum_{\tau=t}^{t+MUT_k-1} \gamma_{k,\tau} \geq MUT_k\, \alpha_{k,t} \qquad k \in \mathcal{K}, t \leq |\mathcal{T}| - MUT_k + 1 \qquad (2.14)$$

$$\sum_{\tau=t}^{|\mathcal{T}|} \gamma_{k,\tau} \geq (|\mathcal{T}| - t + 1)\alpha_{k,t} \qquad k \in \mathcal{K}, |\mathcal{T}| - MUT_k \leq t \leq |\mathcal{T}|. \qquad (2.15)$$

Specifically, constraints (2.14) state that if unit k is started-up at time t, it has to stay ON for at least MUT_k consecutive hours in the planning horizon. Constraints (2.15) model the final periods of the planning horizon in which if unit k is started-up, it remains ON until the end of the planning horizon.

Minimum down time constraints (2.6) can be formulated as:

$$\sum_{\tau=t}^{t+MDT_k-1} \left(1 - \gamma_{k,\tau}\right) \geq MDT_k\, \beta_{k,t} \qquad k \in \mathcal{K}, t \leq |\mathcal{T}| - MDT_k + 1 \qquad (2.16)$$

$$\sum_{\tau=t}^{|\mathcal{T}|} \left(1 - \gamma_{k,\tau}\right) \geq (|\mathcal{T}| - t + 1)\beta_{k,t} \qquad k \in \mathcal{K}, |\mathcal{T}| - MDT_k \leq t \leq |\mathcal{T}|. \qquad (2.17)$$

Similarly to constraints (2.14) and (2.15), while inequalities (2.16) hold in the central hours of the planning horizon, constraints (2.17) model the final periods in which if unit k is shut-down, it remains OFF until the end of the planning horizon.

Since binary variables describing start-up, shut-down and commitment statuses are related by equations (2.3) and (2.4), the thermal unit commitment problem can be formulated also by introducing a single set of binary variables (per unit and per period) to represent the commitment status and expressing start-up and shut-down decisions as a function of the commitment decision variables. In such a formulation, additional

continuous variables $C_{k,t}^{SU}$ are introduced to model start-up costs by imposing the following constraints:

$$C_{k,t}^{SU} \geq SUC_k \left(\gamma_{k,t} - \gamma_{k,t-1}\right) \qquad k \in \mathcal{K}, t \in \mathcal{T} \tag{2.18}$$

$$C_{k,t}^{SU} \geq 0 \qquad k \in \mathcal{K}, t \in \mathcal{T}. \tag{2.19}$$

The objective function (2.1) is modified by including the new formulation of the start-up costs, therefore obtaining

$$\begin{aligned} \min z \quad = \quad & \sum_{k \in K} CM_k \sum_{t \in T} \left(\gamma_{k,t} \underline{P}_k + p_{k,t}\right) + \sum_{k \in K} \sum_{t \in T} C_{k,t}^{SU} \\ + \quad & c_{ENP} \sum_{z \in Z} \sum_{t \in T} ENP_{z,t} + c_{OG} \sum_{z \in Z} \sum_{t \in T} OG_{z,t} \end{aligned} \tag{2.20}$$

Finally, the minimum up/down time constraints are formulated by replacing $\alpha_{k,t}$ with $\left(\gamma_{k,t} - \gamma_{k,t-1}\right)$ in equations (2.14) and (2.15) and $\beta_{k,t}$ with $\left(\gamma_{k,t-1} - \gamma_{k,t}\right)$ in constraints (2.16) and (2.17), therefore obtaining the new set of constraints

$$\sum_{\tau=t}^{t+MUT_k-1} \gamma_{k,\tau} \geq MUT_k \left(\gamma_{k,t} - \gamma_{k,t-1}\right) \qquad k \in \mathcal{K}, t \leq |\mathcal{T}| - MUT_k + 1 \tag{2.21}$$

$$\sum_{\tau=t}^{|\mathcal{T}|} \gamma_{k,\tau} \geq (|\mathcal{T}| - t + 1)\left(\gamma_{k,t} - \gamma_{k,t-1}\right) \qquad k \in \mathcal{K}, |\mathcal{T}| - MUT_k \leq t \leq |\mathcal{T}| \tag{2.22}$$

$$\sum_{\tau=t}^{t+MDT_k-1} \left(1 - \gamma_{k,\tau}\right) \geq MDT_k \left(\gamma_{k,t-1} - \gamma_{k,t}\right) \qquad k \in \mathcal{K}, t \leq |\mathcal{T}| - MDT_k + 1 \tag{2.23}$$

$$\sum_{\tau=t}^{|\mathcal{T}|} \left(1 - \gamma_{k,\tau}\right) \geq (|\mathcal{T}| - t + 1)\left(\gamma_{k,t-1} - \gamma_{k,t}\right) \qquad k \in \mathcal{K}, |\mathcal{T}| - MDT_k \leq t \leq |\mathcal{T}|. \tag{2.24}$$

2.2.5 Tests on different formulations

The different formulations of the thermal unit commitment problem are equivalent in the sense that they impose the same set of constraints on thermal power plants, but different formulations may present different solution times. Several researches have been conducted in the literature to identify the fastest formulation [29, 30, 31]. However, the obtained results strongly depend on the empirical data and the assumptions introduced in the formulation of the thermal unit commitment problem.

To identify the fastest formulation for our specific application, several tests have been performed. Specifically, the following four formulations have been implemented.

- *1bin*: This formulation introduces a single set of binary variables to represent commitment decisions, expressing the start-up and the shut-down manoeuvres as a function of the commitment decision variables. Therefore, *1bin* is formulated by considering equations (2.2)–(2.4), (2.7)–(2.13), (2.18)–(2.24).
- *3binB*: This formulation includes constraints (2.1)–(2.4), (2.7)–(2.17). Thus, formulation *3binB* differs from the model presented in Section 2.2.3 in the way minimum up/down time constraints are formulated.
- *3binA*: This is the formulation presented in Section 2.2.3, including equations (2.1)–(2.13).
- *3binR*: This is the continuous relaxation of model *3binA*, obtained by relaxing integrality constraints (2.10). Formulation *3binR* is implemented as a benchmark for the computational times.

The four formulations have been tested on a scenario elaborated by CESI S.p.A. consisting in three zones connected in a tree network and 48 thermal power plants under a planning horizon of 8760 hours. All the numerical tests have been performed on an ASUS laptop with a 3 GHz Intel Core i7-5500U Processor and 4 GB of RAM using solver Gurobi under GAMS 24.7.4. Since none of the considered formulations could provide a solution to the yearly problem, the planning horizon of one year has been divided into different time blocks: semesters (two instances with 4380 hours each), four-month periods (three instances with 2920 hours), quarters (four instances with 2190 hours each), two-month periods (six instances with 1460 hours each), and single months (twelve instances with 730 hours each). For MILP models an optimality gap of 0.1% has been fixed.

Table 2.1 Solution time [min] required by the four formulations for different lengths of the planning horizon

	1 month	**2 months**	**3 months**	**4 months**	**6 months**
1bin	36.60	126.21	–	–	–
3binB	36.95	42.39	56.54	74.56	–
3binA	36.23	39.20	44.57	45.75	55.57
3binR	34.42	37.01	42.50	44.13	52.72

Table 2.1 reports the solution time required by the different formulations to solve the thermal unit commitment problem for different lengths of the planning horizon. As can be noticed, formulation *3binA* is the most efficient formulation for three reasons. First, this is the only formulation providing a solution to the thermal unit commitment problem when the year is divided into two semesters. Indeed, formulation *3binB* computes a solution to the problem only for optimization horizons shorter than or equal to 4 months, while formulation 1bin provides a solution only for monthly or two-month time blocks: larger instances cause out-of-memory errors.

Second, for every length of the optimization horizon, formulation *3binA* is the fastest MILP model. Third, solution times of formulation *3binA* are very close to the computational time of the relaxed model *3binR*. As can be noticed, most of the computational time for formulation *3binA* is needed to solve the relaxation of the problem, while the optimal integer solution is determined in a small amount of time. For instance, the total time needed to solve the year broken down into four-month blocks in formulation *3binA* is only 3.7% greater than the solution time of model *3binR*. Instead, when semesters are considered, formulation *3binA* requires roughly 5.4% more time than model *3binR*. Indeed, as discussed in [29], the equations introduced to model the thermal unit commitment problem in *3binA* tighten the feasible region by reducing the distance between relaxed and integer solutions. As a consequence of the reduction of the search space explored by MILP solvers, computational times in formulation *3binA* are dramatically reduced.

2.3 The selection of representative days

2.3.1 Literature review

To provide a better representation of the short-term operation while maintaining the problem computationally tractable, some energy planning models use a small number of representative periods (i.e., days or weeks) instead of modelling every hour of the planning horizon. Different approaches have been proposed to identify representative periods. Some authors select representative days by using simple heuristics. For instance, to capture the fluctuations in demand during the year, in [32] three time periods are selected as the day that contains the minimum demand level of the year, the day that contains the maximum demand level and the day that contains the largest demand spread in 24 hours. Each of these representative days is then weighted in such

way that the weighted sum of the hourly demand of the three selected days equals the overall original demand of the entire year. Instead, reference [33] proposes to represent the operation of a power system by means of four representative weeks, one for every season. The load profile of each of these representative weeks is the average of the load profiles of all the weeks of a season. In the objective function, the operational cost of each representative week is weighted by the number of weeks in the season. To ensure that the power system has enough flexibility to manage extreme conditions, authors propose also to add to the previous four average weeks a fifth week representing extreme conditions. However, no explanations about the way the extreme week is selected are provided in the paper.

Other works combine heuristic approaches with the random selection of some additional days. For instance, in [34] unit commitment decisions within each investment period are optimized by considering a set of 12 days: two for each even-numbered month, with one day corresponding to the peak-load day and the second day randomly selected from days belonging to the same month. In [35], in order to significantly reduce the size of the planning problem, authors propose to consider 28 representative days in the whole planning horizon that are obtained by selecting 20 random days to characterize typical system behaviour and 8 specific days that contain hours with extreme meteorological and load events.

More advanced methods are based on clustering algorithms in order to group days with similar load, wind power production or solar power production into clusters: the cluster centroid or a specific historical day for each group is then taken as the representative day. Different clustering algorithms have been proposed in the literature. For example, reference [36] employs k-means algorithm, in [37] the Ward's hierarchical clustering algorithm is used, while reference [38] suggests to apply k-means algorithm using median representatives.

Finally, some works select representative days by considering the historical load duration curve and the one obtained from the load in the representative days (see e.g. [39] and [40]). Specifically, authors in [40] design a mixed integer linear programming model to select representative days and determine their weights so as to minimize the distance between the historical load duration curve and the one obtained by using the selected days.

From this brief review about the selection of representative days, four main observations can be drawn:

- Clustering algorithms provide better approximations than heuristic approaches [38, 40].
- Using as representative days historical days rather than clusters means provide better results, especially if power systems have large shares of renewables [37, 38].
- Representative days selected using clustering algorithms characterize typical system behaviour. However, the occurrence of extreme events should also be taken into account to properly design the expansion of a power system [33, 34, 35].
- The distance between the historical load duration curve and the one approximated by representative days can provide usefull information about the goodness of the representation [39, 40].

On the basis of the above observations, in this thesis a novel hybrid approach is introduced to identify representative days. Our method is based on the iterative application of the k-medoids algorithm, on the addition of some extreme days to the set of representative days identified by the clustering algorithm and on the selection of the most suited number of representative days to be used by considering load duration curves.

2.3.2 The proposed method to select representative days

In this section, we describe our procedure to select representative days from a set

$$D_{z,t}^{d},\ \mu_{z,t}^{d},\ \rho_{z,t}^{d} \qquad z \in \mathcal{Z}, 1 \le d \le 365, 1 \le t \le 24 \tag{2.25}$$

where the load data $D_{z,t}^{d}$ are either historical values (typically related to the last year before the planning horizon) or forecast values for the first year of the planning horizon, while $\mu_{z,t}^{d}$ and $\rho_{z,t}^{d}$ are technical production/capacity ratios for solar power production and wind power production, respectively. By performing the clustering analysis on this data set, correlations among production and load, as well as spatial correlations among zones, can be taken into account.

In the proposed procedure, before determining representative days, for every zone z the original hourly load data $D_{z,t}^{d}$, $1 \le d \le 365$ and $1 \le t \le 24$, are sorted in order of decreasing magnitude so as to determine the zonal load duration curves $LDC_{z,\tau}$, $1 \le \tau \le 8760$. These curves will be compared in the termination test with the zonal load duration

curves corresponding to the representative days and their associated weights to determine the number of representative days.

In order to take into account extreme conditions that cannot be captured by clustering analysis, representative days c_1 and c_2 are chosen as the days with minimum and maximum total load in the power system, i.e.,

$$D_{z,t}^{c_1},\ \mu_{z,t}^{c_1},\ \rho_{z,t}^{c_1},\quad c_1 = argmin_d\left(\sum_{z\in\mathcal{Z}}\sum_{t=1}^{24} D_{z,t}^{d}\right)\quad z\in\mathcal{Z}, 1\le t\le 24 \tag{2.26}$$

$$D_{z,t}^{c_2},\ \mu_{z,t}^{c_2},\ \rho_{z,t}^{c_2},\quad c_2 = argmax_d\left(\sum_{z\in\mathcal{Z}}\sum_{t=1}^{24} D_{z,t}^{d}\right)\quad z\in\mathcal{Z}, 1\le t\le 24 \tag{2.27}$$

Once the two extreme days c_1 and c_2 are identified, the original data set is modified by deleting days c_1 and c_2 and by normalizing the load values $D_{z,t}^{d}$ in the range of $[0; 1]$. Further representative days are selected by the following iterative procedure performed on the modified data set and based on the application of the k-medoids algorithm [41].

Like k-means, the objective of the k-medoids algorithm is to compute k clusters in order to minimize the deviation between observations and their representative (i.e., the centroids). However, the main difference between the two algorithms is that for k-means each cluster representative is computed as the mean of all the points within the cluster, while for k-medoids each cluster centroid is a vector belonging to that group. As discussed in the previous section, the use of specific historical days, rather than cluster means, as representative days usually provides better results. For this reason, in our approach we suggest to apply k-medoids rather than k-means. The steps of the iterative procedure are as follows:

1. Set $k = 2$;
2. The days of the modified data set are partitioned in k clusters by the k-medoids algorithm;
3. The representative day $c_{2+\xi}$, for ξ, $1 \le \xi \le k$, is selected from the original data set as the day corresponding to the centroid of cluster ξ; the weight associated to representative day $c_{2+\xi}$ is the number of days in cluster ξ;
4. For each zone, determine the zonal load duration curve corresponding to the $k + 2$ representative days and their associated weights (a unit weight is associated to c_1 and c_2) and compute the mean absolute percentage error between the original

zonal load duration curve and the one corresponding to the current set of representative days;

5. Compute the system average mean absolute percentage error by averaging the mean absolute percentage errors of zonal load duration curves;
6. If the system average mean absolute percentage error is below the given threshold, stop, otherwise increase k by 1 and go to step 2.

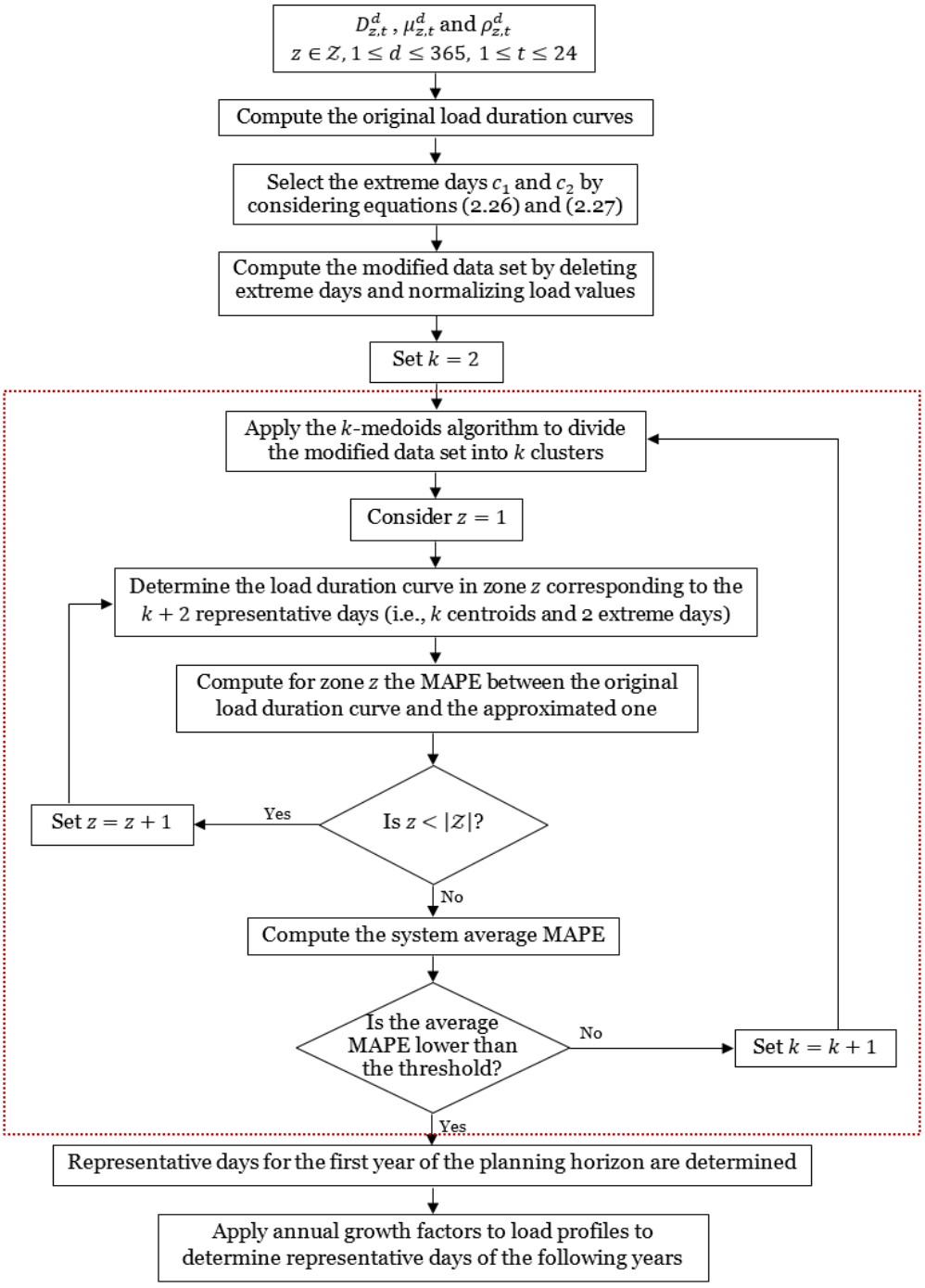


Fig. 2.1 Illustration of the representative days selection procedure

If the planning horizon includes several years, once the representative days for the first year of the planning horizon are determined, the representative days of the following years are derived by applying annual growth factors to load profiles. The drawback of this approach is to neglect the evolution of technical and socio-economic parameters, such as the impact of electric vehicles and heat pumps or an increase of capacity factors for wind and solar power technologies, which could lead to significant variations for the shape of both demand and RES profiles in the future. Since no reliable information about changes in the future shape of demand and RES profiles was available, in our application we decided to simply use annual growth factors. However, the proposed approach can be easily modified to include long-term variations. For instance, different shapes of load (or of RES production) profiles in the years of the planning horizon can be obtained by defining multiplicative coefficients varying on the set of hours of the representative days of every year: the hourly load (or RES production) forecast of every year in the planning horizon is then obtained by multiplying the hourly forecast of the first year by the corresponding multiplicative coefficient.

Fig. 2.1 summarizes the proposed strategy to select representative days from historical days.

2.4 Capturing seasonality of the hydroelectric dispatch with representative days

The use of representative days introduces a strong limitation in storage operation modeling. Indeed, representative days allow providing a good representation of storage operation within a day, but, since the chronology among representative days is not preserved, any energy storage system with a cycle longer than 24 hours cannot be modeled with great accuracy. Thus, the evaluation of operational conditions through representative days raises the crucial issue regarding how to consider the seasonality of hydropower plants, which are usually characterized by yearly cycles. To model the operation of hydropower facilities, two different approaches can be applied:

1. According to historical data or the results of a preliminary model, hydroelectric productions are assigned to each representative day.
2. Representative days are connected in order to catch the seasonality of hydropower plants operation. While some works consider representative days as temporal consecutive [42], some recent studies model the long-term storage by

considering the Cluster Index, which is a numeric column vector where each row indicates the cluster assignment of the corresponding day of the year [43].

The disadvantage of the first approach is the need to apply a statistical analysis or an optimization model to determine the medium-term or long-term energy content of storage facilities. Instead, by imposing constraints that connect representative days, the drawback of the second approach is an increasing computational complexity and the impossibility to decompose the problem by representative day, which is a strategy commonly applied to pursue scalability in large-scale models [23]. Therefore, the choice of the best approach to adopt depends on the specific application. When the disconnection between representative days allows obtaining a huge computational saving, the first approach should be implemented, while the second approach is preferred when working with already connected representative days or when historical data for the power system under study are not available.

In this section, both approaches are discussed in details. Specifically, while Section 2.4.1 illustrates the proposed analysis to determine the hydroelectric dispatch in each representative day by means of a preliminary optimization model, Section 2.4.2 introduces the constraints that connect representative days to model long-term storage operations.

2.4.1 A preliminary model for the hydroelectric dispatch

In this first approach, the hydroelectric dispatch is determined by applying a preliminary model based on the following assumptions:

- As for the thermal unit commitment problem, a zonal representation of the power system is adopted, i.e., the power system is partitioned into different price zones that exchange energy through a capacitated power network.
- Each hydraulic valley is represented as a single equivalent hydroelectric plant with reservoir.
- Time is represented with daily resolution. Each zone is characterized by a given daily net load $\widehat{NL}_{z,d}$, which is the difference between the daily load and the daily generation from non-programmable renewable power sources, while each reservoir presents given daily natural inflows $\hat{F}_{h,d}$.
- The objective of the model is to maximize the revenues related to the hydroelectric production, which can be computed as the product between the

zonal energy prices and the hydroelectric energy produced to supply the net load in a given market zone.

- Zonal prices are computed as a linear function of the zonal residual net load, i.e., the difference between the zonal net load $\widehat{NL}_{z,d}$ and the hydroelectric energy produced to supply the net load in a given market zone. Thus, zonal prices are endogenous parameters depending on the water resources dispatching decisions.

The solution of the hydroelectric model provides the daily energy production for each hydropower plant in each day of the planning horizon. Knowing the representative day assigned to each day of the year (an information provided by the clustering analysis), a post-processing activity is applied to determine for each representative day the hydroelectric energy production, which is used to limit the sum of the hourly hydroelectric generation in that representative day.

As previously mentioned, in the proposed analysis zonal prices are considered as a function of the net load. Thus, a preliminary analysis is needed to estimate the relation between energy prices and net load.

2.4.1.1 Price estimation

The relation between prices and net load can be estimated by applying a linear regression model to historical data. In this paragraph, the results related to the Italian power system are presented.

Specifically, we performed the regression analysis on a dataset provided by CESI S.p.A. including hourly energy prices and net load values for several European countries in years 2015 and 2016. This dataset has been modified by computing for each day of the planning horizon and for each system zone the average daily price and the total daily net load. On the modified dataset two linear regression models have been estimated. First, a simple linear regression model have been applied, being the daily zonal net load $\widehat{NL}_{z,d}$ the only covariate to explain the average zonal daily price $\pi_{z,d}$. Such a model can be formulated as:

$$\pi_{z,d} = \alpha_z + \beta_z \widehat{NL}_{z,d} \tag{2.28}$$

For instance, Table 2.2 reports the intercept and the slope of the linear relationship between prices and net load estimated on the Italian power system, graphically represented by the solid red line in Fig. 2.2. As can be noticed, the linear model provides a good approximation of data, being the coefficient of determination equal to 0.65.

Table 2.2 Estimates of model (2.28) regression coefficients for the Italian power system

Parameter	Value	Standard Deviation	p-value
α_z	8.71	0.35	$< 10^{-15}$
β_z	$6.26 \cdot 10^{-5}$	$4.83 \cdot 10^{-7}$	$< 10^{-15}$

Although the linear model seems to provide an accurate estimation of zonal prices, the results have been further improved by introducing in the regression model a monthly interaction effect

$$\pi_{z,d} = \alpha_z + \sum_{m=1}^{12} \beta_{z,m} \left(\widehat{NL}_{z,d} x_{m,d} \right) \tag{2.29}$$

being m the index for months and $x_{m,d}$ a dummy variable that is equal to 1 if day d belongs to month m. Specifically, in this model the effect of the daily net load on the average price changes in different months of the planning horizon. Table 2.3 reports the values of the regression coefficients estimated on the Italian power system. As can be noticed by analysing p-values, all the parameters contribute in a meaningful way to the explanation of the prices. The new value of the coefficient of determination is 0.75, which is consistently greater than the coefficient of determination of the model without the monthly interaction (i.e., 0.65). Fig. 2.3 illustrates the monthly linear relations between prices and net load estimated on the Italian power system.

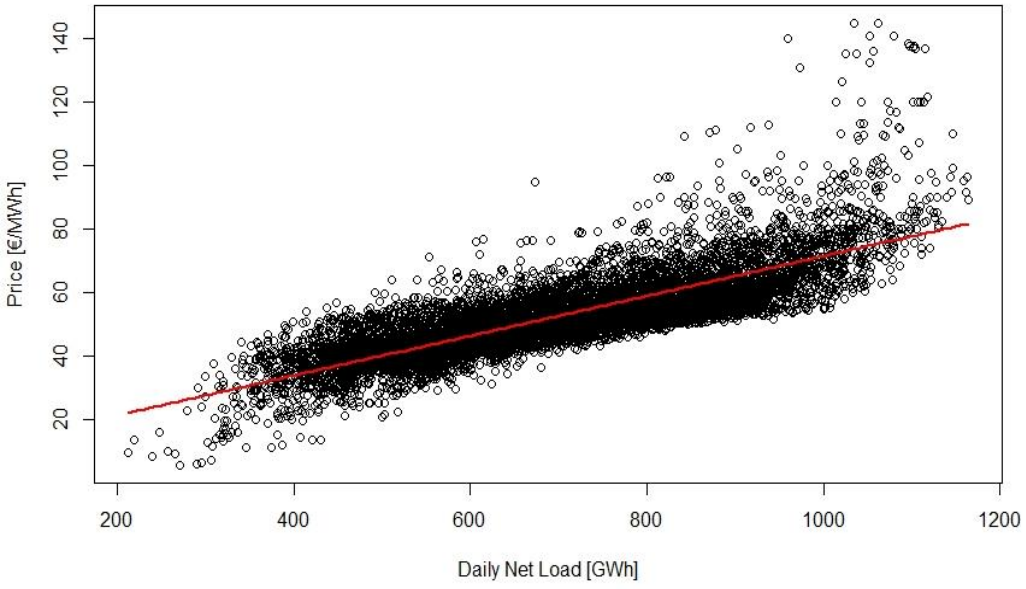


Fig. 2.2 Simple linear regression model estimated on the Italian power system

Table 2.3 Estimates of model (2.29) regression coefficients for the Italian power system

Parameter		Value	Standard Deviation	p-value
α_z		10.79	0.34	$< 10^{-15}$
β_z	Jan	$5.66 \cdot 10^{-5}$	$5.77 \cdot 10^{-7}$	$< 10^{-15}$
	Feb	$5.90 \cdot 10^{-5}$	$5.69 \cdot 10^{-7}$	$< 10^{-15}$
	Mar	$5.80 \cdot 10^{-5}$	$6.08 \cdot 10^{-7}$	$< 10^{-15}$
	Apr	$6.02 \cdot 10^{-5}$	$6.77 \cdot 10^{-7}$	$< 10^{-15}$
	May	$5.91 \cdot 10^{-5}$	$6.64 \cdot 10^{-7}$	$< 10^{-15}$
	Jun	$5.44 \cdot 10^{-5}$	$6.06 \cdot 10^{-7}$	$< 10^{-15}$
	Jul	$6.77 \cdot 10^{-5}$	$4.91 \cdot 10^{-7}$	$< 10^{-15}$
	Aug	$6.44 \cdot 10^{-5}$	$6.42 \cdot 10^{-7}$	$< 10^{-15}$
	Sep	$5.51 \cdot 10^{-5}$	$5.96 \cdot 10^{-7}$	$< 10^{-15}$
	Oct	$5.39 \cdot 10^{-5}$	$5.39 \cdot 10^{-7}$	$< 10^{-15}$
	Nov	$6.30 \cdot 10^{-5}$	$5.68 \cdot 10^{-7}$	$< 10^{-15}$
	Dec	$6.20 \cdot 10^{-5}$	$5.59 \cdot 10^{-7}$	$< 10^{-15}$

Described the procedure designed to estimate the prices for electricity, the following sections provide a detailed description of the preliminary hydroelectric model.

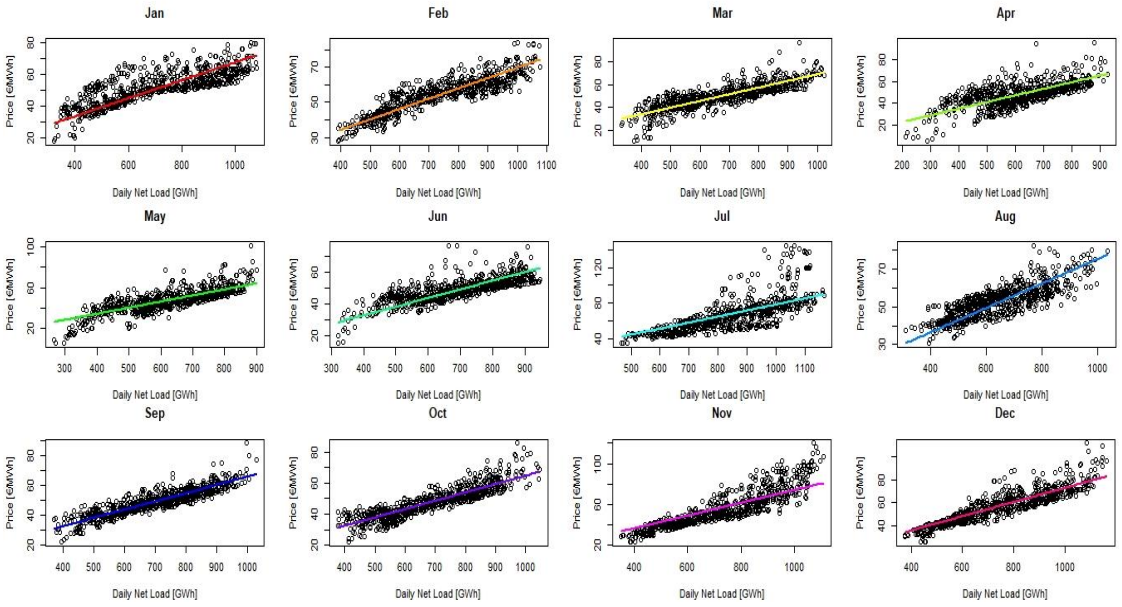


Fig. 2.3 Linear regression model with monthly interaction effect estimated on the Italian power system

2.4.1.2 Notation

To formulate the preliminary hydroelectric model, the following notation is introduced.

Sets

$\mathcal{Z}$	Set of zones, indexed by z
$\mathcal{H}$	Set of hydropower plants, indexed by h
$\Omega_z^h \subset \mathcal{H}$	Set of hydropower plants located in zone z
$\mathcal{M}$	Set of months, indexed by m
$\mathcal{D}$	Set of days, indexed by d
$\mathcal{D}^m$	Set of days belonging to month m
$\mathcal{L}$	Set of transmission lines, indexed by l
$rz(l)$	Receiving-end zone of transmission line l
$sz(l)$	Sending-end zone of transmission line l

Parameters

α_z	[€/MWh]	Intercept of the zonal relationship between price and net load in zone z
$\beta_{z,m}$	[€/MWh^2]	Slope of the zonal relationship between price and net load in zone z at month m
$\widehat{NL}_{z,d}$	[MWh]	Daily net demand for electricity in zone z at day d
$\bar{q}_h$	[MWh]	Maximum daily energy produced by hydropower plant h
$\overline{pp}_h$	[MWh]	Maximum daily pumping energy absorbed by hydropower plant h
$\bar{sl}_h$	[MWh]	Maximum daily energy spillage from reservoir h
$\bar{E}_h$	[MWh]	Maximum energy content of reservoir h
$E_{h\,0}$	[MWh]	Energy content of hydropower plant h at the beginning of the time horizon
λ_h	[−]	Loss coefficient for energy stored by hydropower plant h $(0 \leq \lambda_h \leq 1)$
λ_h^{IN}	[−]	Loss coefficient for hydro plant h pumping $\left(0 \leq \lambda_h^{\mathrm{IN}} \leq 1\right)$
λ_h^{OUT}	[−]	Loss coefficient for hydro plant h power generation $\left(\lambda_h^{\mathrm{OUT}} \geq 1\right)$
$\hat{F}_{h,d}$	[MWh]	Daily energy inflow for hydropower plant h at day d

$\underline{Fd_l}$	[MWh]	Lower bound on daily energy flows on transmission line l
$\overline{Fd_l}$	[MWh]	Upper bound on daily energy flows on transmission line l

Variables

$\varphi_{h,d}$	[−]	Binary status of hydropower plant h at day d, which is equal to 1 if plant h produces energy and 0 otherwise
$q_{h,d}$	[MWh]	Daily energy produced by hydropower plant h at day d
$pp_{h,d}$	[MWh]	Daily pumping energy absorbed by hydropower plant h at day d
$sl_{h,d}$	[MWh]	Daily energy spillage from reservoir h at day d
$E_{h,d}$	[MWh]	Energy content of reservoir h at the end of day d
$\hat{x}_{l,d}$	[MWh]	Daily energy flow on transmission line l at day d
$\pi_{z,d}$	[MW]	Daily market price in zone z at day d

2.4.1.3 Mathematical formulation

The hydroelectric preliminary model can be formulated as the following Quadratic Mixed-Integer Programming (QMIP) problem

$$\max \; z = \sum_{z\in\mathcal{Z}} \sum_{d\in\mathcal{D}} \pi_{z,d} \left(\sum_{h\in\Omega_z^h} \left(q_{h,d} - pp_{h,d}\right) + \sum_{l|rz(l)=z} \hat{x}_{l,d} - \sum_{l|sz(l)=z} \hat{x}_{l,d} \right) \tag{2.30}$$

subject to

$$\pi_{z,d} = \alpha_z + \beta_{z,m} \left[\widehat{NL}_{z,d} - \left(\sum_{h\in\Omega_z^h} \left(q_{h,d} - pp_{h,d}\right) + \sum_{l|rz(l)=z} \hat{x}_{l,d} - \sum_{l|sz(l)=z} \hat{x}_{l,d} \right) \right]$$

$$z \in \mathcal{Z}, m \in \mathcal{M}, d \in \mathcal{D}^m \tag{2.31}$$

$$q_{h,d} \leq \varphi_{h,d}\, \bar{q}_h \qquad h \in \mathcal{H}, d \in \mathcal{D} \tag{2.32}$$

$$pp_{h,d} \leq \left(1 - \varphi_{h,d}\right) \overline{pp}_h \qquad h \in \mathcal{H}, d \in \mathcal{D} \tag{2.33}$$

$$sl_{h,d} \leq \bar{sl}_h \qquad h \in \mathcal{H}, d \in \mathcal{D} \tag{2.34}$$

$$E_{h,d} \leq \bar{E}_h \qquad h \in \mathcal{H}, d \in \mathcal{D} \tag{2.35}$$

$$E_{h,d} = (1-\lambda_h)\, E_{h_0} + \hat{F}_{h,d} + \lambda_h^{\mathrm{IN}}\, pp_{h,d} - \lambda_h^{\mathrm{OUT}} q_{h,d} - sl_{h,d}$$

$$h \in \mathcal{H}, d = 1 \qquad (2.36)$$

$$E_{h,d} = (1-\lambda_h)\, E_{h,d-1} + \hat{F}_{h,d} + \lambda_h^{\mathrm{IN}}\, pp_{h,d} - \lambda_h^{\mathrm{OUT}} q_{h,d} - sl_{h,d}$$

$$h \in \mathcal{H}, d > 1 \qquad (2.37)$$

$$E_{h,|\mathcal{D}|} = E_{h_0} \qquad h \in \mathcal{H} \qquad (2.38)$$

$$\underline{Fd_l} \le \hat{x}_{l,d} \le \overline{Fd_l} \qquad l \in \mathcal{L}, d \in \mathcal{D} \qquad (2.39)$$

$$\varphi_{h,d} \in \{0,1\} \qquad h \in \mathcal{H}, d \in \mathcal{D} \qquad (2.40)$$

$$q_{h,d}, pp_{h,d}, sl_{h,d}, E_{h,d} \ge 0 \qquad h \in \mathcal{H}, d \in \mathcal{D} \qquad (2.41)$$

$\hat{x}_{l,d}$ free variable $\qquad l \in \mathcal{L}, d \in \mathcal{D} \qquad (2.42)$

$$\pi_{z,d} \ge 0 \qquad z \in \mathcal{Z}, d \in \mathcal{D}. \qquad (2.43)$$

The objective (2.30) of the proposed model is to maximize the total system revenues related to the hydroelectric production, which are computed by multiplying in each zone the market price by the energy produced to supply the zonal net load, i.e., the difference between produced and pumped energy plus the difference between incoming and outgoing energy flows. In particular, zonal prices are endogenous variables defined by constraints (2.31) as a linear function of the residual net load (i.e., the net load after dispatching the water resources). Indeed, prices are estimated considering that the hydropower used in a specific zone to supply the net load will reduce the net load itself, causing consequently a price reduction. Inequalities (2.32) both impose upper bounds to the energy produced by hydropower plants and enforce consistency between positive and binary variables related to the hydropower generation. Constraints (2.33) work similarly for the pumping case. Moreover, the joint application of constraints (2.32) and (2.33) ensures the exclusive mutuality of generation and pumping. Inequalities (2.34) and (2.35) impose upper bounds to the spillage and the energy level of reservoirs, respectively. Constraints (2.36) and (2.37) are flow conservation equations and they ensure that the energy stored by hydro plant h at the end of day d equals the energy

stored at the end of the previous day (reduced by the loss coefficient $\lambda_h \leq 1$), plus the natural inflows, plus the energy injected in h (reduced by the coefficient $\lambda_h^{\mathrm{IN}} \leq 1$), minus the energy released from the reservoir (reduced by the coefficient $\lambda_h^{\mathrm{OUT}} \geq 1$), minus the spillage. While equations (2.36) impose the energy balance for the first day of the planning horizon, constraints (2.37) apply to all days but the first one. Equations (2.38) ensure the equality between final and initial energy stored in each reservoir. Constraints (2.39) impose upper and lower bounds to the energy flows on transmission lines. Finally, constraints (2.40)–(2.43) define decision variables.

2.4.1.4 Post-processing

The model described in Section 2.4.1.3 determines the daily energy production and storage volumes of reservoirs for each hydropower plant in each day of the planning horizon. For instance, Fig. 2.4 shows inputs and outputs provided by the model for the equivalent hydropower plant *CH2a* located in Switzerland in the simulation of the European power system at year 2017. Further details about this analysis are provided in Section 2.7. Specifically, the two top graphs in Fig. 2.4 represent, respectively, the daily net load in Switzerland (blue line) and the daily natural inflows (red line) for the considered hydropower plant in the planning horizon. As can be observed, the net load in Switzerland shows a very particular behaviour, with high values in both the first and the last months of the year and a clear valley from day 120 to day 280 (i.e., in spring and summer months). The two bottom graphs in Fig. 2.4 depicts the model outputs, i.e., the daily energy production (yellow line) and the daily storage of the reservoir (green line).

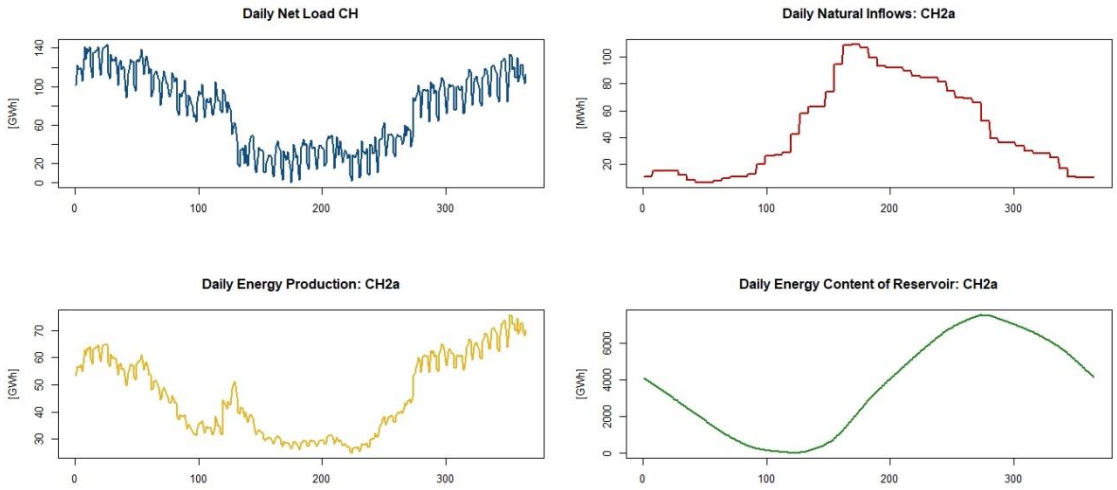


Fig. 2.4 Water resources dispatching for equivalent hydropower plant CH2a

As can be noticed, the energy production presents a profile very similar to the daily net load, with high productions during the first and the last months and low hydroelectric generation in the middle part of the year. Indeed, since the market price is computed as a linear function of the net load with positive coefficients, high values of net load imply high expected prices. Thus, in order to maximise revenues, the decision of the model is to increase the hydroelectric generation when the net load is higher. Given the low natural inflows during the first months of the year, to supply the high load the hydropower plant uses the water stored in the reservoir, causing a progressive reduction of the basin's level until the day 120, when the storage reaches its minimum value. From day 120 to day 280, both the lower generation and the high natural inflows determine an increase of energy content in the reservoir. In the last days of the year, the considerable increase in power generation causes a reduction of water volumes stored, until reaching the initial level at the end of the year, as imposed by constraints (2.38).

After determining the optimal daily dispatch of water resources for every plant, the model output is post-processed in order to coordinate the information obtained with the representative days provided by the clustering analysis. The first step for this coordination is the computation of the daily hydroelectric energy balance $Q_{h,d}$ for hydropower plant h at day d, which is defined as

$$Q_{h,d} = \lambda_h^{\mathrm{OUT}} q_{h,d} - \lambda_h^{\mathrm{IN}} pp_{h,d} \qquad h \in \mathcal{H}, d \in \mathcal{D} \qquad (2.44)$$

Let $\mathcal{C}$ denote the set of representative days and let $\mathrm{Map}_{d,c}$ denote the injective map of each day $d \in \mathcal{D}$ to a representative day $c \in \mathcal{C}$. In a first approach, the daily hydroelectric energy assigned to each representative day c can be computed as the average of daily hydroelectric energy balances for days associated to representative day c, i.e.,

$$Q_{h,c} = \frac{1}{wg_c} \sum_{d \in \mathrm{Map}_{d,c}} Q_{h,d} \qquad h \in \mathcal{H}, c \in \mathcal{C} \qquad (2.45)$$

where wg_c denotes the weight of representative day c, i.e., the number of days assigned to cluster c.

For instance, let us consider again the European scenario. In this case, the clustering analysis identifies 10 representative days with different weights. Let us consider in particular the first representative day (i.e., *Cluster 1*), whose weight is equal to 38, and the hydropower plant *CH2a* located in Switzerland.

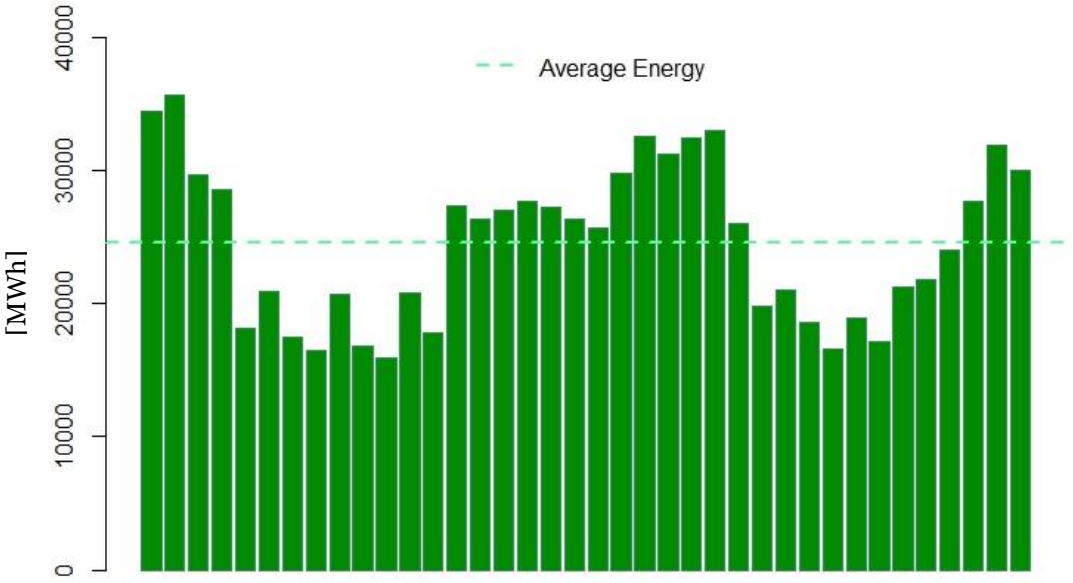


Fig. 2.5 Daily hydroelectric energy for hydropower plant CH2a in Cluster 1

Fig. 2.5 shows the results of the post-processing for *Cluster 1* and hydropower plant *CH2a*: each bar represents a specific day allocated to *Cluster 1* and it is characterized by a height equal to parameter $Q_{h,d}$ (i.e., the energy balance between hydroelectric generation and pumping in that day). The horizontal dotted line in the figure is the average of these productions and it represents the hydroelectric production assigned to representative day *1* by post-processing.

Alternatively, it is possible to associate to each representative day more than one value, as shown in Fig. 2.6.

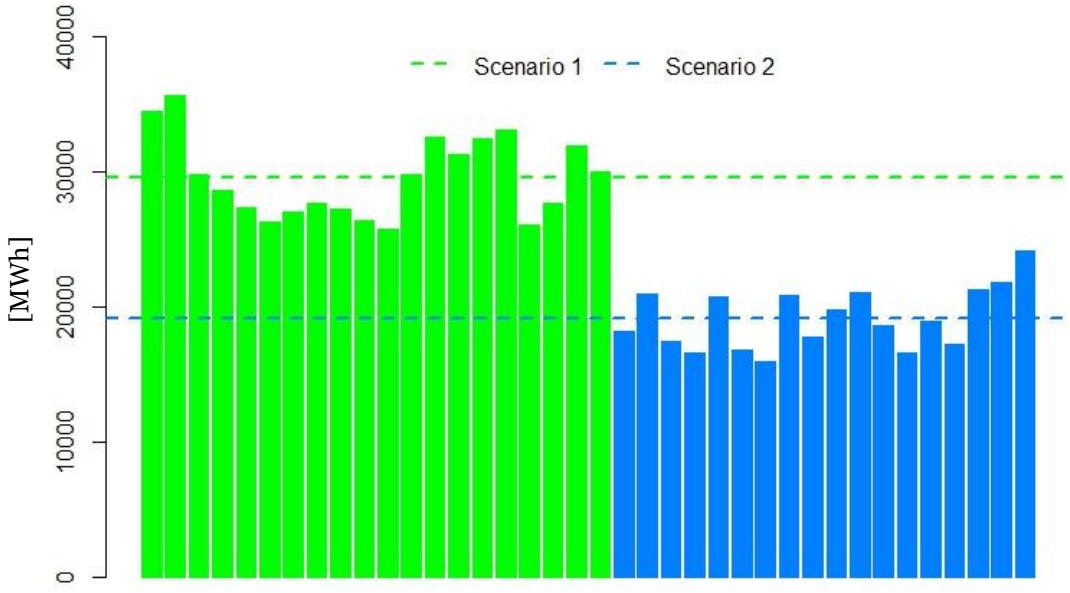


Fig. 2.6 Daily hydroelectric energies for hydropower plant CH2a in Cluster 1

In this second approach, the k-means algorithm is applied in order to divide data into two groups, identifying two different values of hydroelectric energy. In the considered example, these two values are 29 555 MWh (dotted green line) and 19 141 MWh (dotted blue line): while the higher value can be associated with 20 days of *Cluster 1*, the other 18 days show a hydroelectric production close to the lower value. Thus, in this second approach the original representative day *Cluster 1* is replaced by two representative days *Cluster 1A* and *Cluster 1B*, which present the same load, wind and solar profiles, but different hydroelectric balances.

Of course, this approach can be extended to consider a higher number of scenarios for hydroelectric generation at the cost of an increasing number of representative days.

The post-processing activity allows determining for each representative day c the total daily hydroelectric production, which is used to control the sum of hourly hydroelectric generation in representative day c. Specifically, to control the operations of hydropower plants in representative days, variables $E_{h,t}^{c,\mathrm{IN}}$ and $E_{h,t}^{c,\mathrm{OUT}}$ are introduced to represent the hourly energy charge and discharge for hydropower plant h in hour t ($1 \leq t \leq 24$) of representative day c.

The seasonality of the water resources dispatching is then considered in the short-term dispatch by imposing constraints

$$\sum_{t=1}^{24}\left(\lambda_h^{\mathrm{OUT}} E_{h,t}^{c,\mathrm{OUT}} - \lambda_h^{\mathrm{IN}} E_{h,t}^{c,\mathrm{IN}}\right) = Q_{h,c} \qquad h \in \mathcal{H}, c \in \mathcal{C} \qquad (2.46)$$

which ensure that the difference between the daily hydroelectric generation and the daily energy pumped equals the value assigned to the representative day after the preliminary dispatch.

2.4.2 Connecting representative days to model long-term storage

A first drawback of the approach described in the previous section is the need to increase the number of representative days introduced in the analysis to provide a better representation of the long-term hydroelectric dispatch, since days with similar load, wind and solar profiles could be characterized by very different values for hydroelectric production. Another disadvantage of the previous approach is the need to employ historical data to estimate parameters α_z and $\beta_{z,m}$ of the linear relations between prices and net load that drive long-term water resources dispatching decisions. However, in some applications this analysis could not be applied either because historical data are

not available for the power system under study or because past information is considered to be not reliable to predict the future behaviour of the power system.

In these cases, the long-term storage can be modeled by connecting representative days through some constraints that consider the assignment of every day to a representative day to create the continuity in storage across the entire time horizon [44].

Specifically, this approach is based on the concept of *cluster index*, which is the injective map $\text{Map}_{d,c}$ containing the relationship between days and representative days. For instance, let us consider a set of seven days $\mathcal{D} = \{d1, d2, \dots, d7\}$ and a set of two representative days $\mathcal{C} = \{c1, c2\}$. The set $\mathcal{D}$ could represent for example the days in a week, with the two representative days being working days and weekend days. In this case, the cluster index $\text{Map}_{d,c}$ would present the structure illustrated in Fig. 2.7.

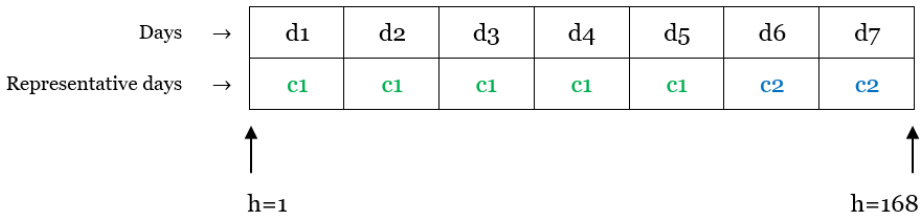


Fig. 2.7 Cluster index $\mathbf{Map}_{d,c}$ for the considered example

Integral constraints imposed to model the hydroelectric operation are divided into two types: *intra-day* and *inter-day* balance equations. Specifically, intra-day equations control the hydropower plants operation within each representative day. Denoting with $E_{h,t}^{c,\text{IN}}$, $E_{h,t}^{c,\text{OUT}}$, $E_{h,t}^{c}$ and $sl_{h,t}^{c}$ the decision variables representing the charge, the discharge, the energy level and the spillage for hydropower plant h in hour t of representative day c, intra-day balance equations can be expressed as

$$E_{h,t}^{c} = (1-\lambda_h)E_{h_0}^{c} + F_{h,t}^{c} + \lambda_h^{\text{IN}}\, E_{h,t}^{c,\text{IN}} - \lambda_h^{\text{OUT}}\, E_{h,t}^{c,\text{OUT}} - sl_{h,t}^{c} \qquad h \in \mathcal{H}, c \in \mathcal{C}, t = 1 \tag{2.47}$$

$$E_{h,t}^{c} = (1-\lambda_h)E_{h,t-1}^{c} + F_{h,t}^{c} + \lambda_h^{\text{IN}}\, E_{h,t}^{c,\text{IN}} - \lambda_h^{\text{OUT}}\, E_{h,t}^{c,\text{OUT}} - sl_{h,t}^{c} \qquad h \in \mathcal{H}, c \in \mathcal{C}, 2 \le t \le 24 \tag{2.48}$$

$$E_{h,t}^{c} \ge E_{h_{24}}^{c} \qquad h \in \mathcal{H}, c \in \mathcal{C}, t = 24, \tag{2.49}$$

where $E_{h_0}^{c}$ and $E_{h_{24}}^{c}$ are the energy levels of reservoir h at the beginning and the end of representative day c, $F_{h,t}^{c}$ denotes the hourly natural inflows, and parameters λ_h, λ_h^{IN} and λ_h^{OUT} are the loss coefficients for storage, charge and discharge activities, respectively. By

assigning specific values to parameters $E^c_{h_0}$ and $E^c_{h_{24}}$ the long-term storage can be controlled. However, the choice of the values to be assigned to these parameters is not trivial. Thus, when working with representative days, intra-day balance constraints are usually replaced by inter-day balance constraints.

In particular, inter-day balance equations check at regular intervals (every M days) the storage level. For instance, let us consider weekly checks (i.e., $M = 7$). In the example represented in Fig. 2.7, the storage level at the end of the week is the storage level at the beginning of the week, plus 5 times the energy balance in representative day $c1$, plus 2 times the energy balance in representative day $c2$. Specifically, the energy balance in representative day $c1$ is multiplied by 5 because in the considered week 5 days (i.e., days $d1$ to $d5$) are associated with the representative day $c1$, as shown by the cluster index. Similarly, the energy balance in day $c2$ is multiplied by 2 since days $d6$ and $d7$ are represented by $c2$.

More generally, denoting with $\mathcal{D}_M$ the subset of days in the planning horizon in which the long-term storage level is checked, with E_{h_0} the energy content of reservoir h at the beginning of the planning horizon and with $\hat{E}_h^{\mathrm{LT},d}$ the decision variable representing the energy level of reservoir h at the end of day d, inter-day constraints can be expressed as

$$\hat{E}_h^{\mathrm{LT},d} = E_{h_0} + \sum_{d'=d-M+1}^{d} \sum_{c\in \mathrm{Map}_{d',c}} \sum_{t=1}^{24} \left(F^c_{h,t} + \lambda^{\mathrm{IN}}_h E^{c,\mathrm{IN}}_{h,t} - \lambda^{\mathrm{OUT}}_h E^{c,\mathrm{OUT}}_{h,t} - sl^c_{h,t}\right) \qquad h \in \mathcal{H}, d = M \tag{2.50}$$

$$\hat{E}_h^{\mathrm{LT},d} = \hat{E}_h^{\mathrm{LT},d-M} + \sum_{d'=d-M+1}^{d} \sum_{c\in \mathrm{Map}_{d',c}} \sum_{t=1}^{24} \left(F^c_{h,t} + \lambda^{\mathrm{IN}}_h E^{c,\mathrm{IN}}_{h,t} - \lambda^{\mathrm{OUT}}_h E^{c,\mathrm{OUT}}_{h,t} - sl^c_{h,t}\right) \qquad h \in \mathcal{H}, d \in \mathcal{D}_M, d > M \tag{2.51}$$

$$0 \le \hat{E}_h^{\mathrm{LT},d} \le EPR_h \bar{E}_h^{\mathrm{IN}} \qquad h \in \mathcal{H}, d \in \mathcal{D}_M \tag{2.52}$$

In particular, equations (2.50) ensure that the storage level at the first checkpoint (i.e., at day M) is equal to the initial energy content of reservoir h, plus/minus the daily energy inflows/outflows in the first M days of the planning horizon. Constraints (2.51) hold for every subsequent checkpoint and they express the energy level at the checkpoint as the sum of the total energy at the previous checkpoint, plus/minus the energy inflows/outflows since the previous checkpoint. Finally, inequalities (2.52) ensure that the storage level at every checkpoint is non-negative and does not exceed the maximum

storage capacity $EPR_h \bar{E}_h^{\mathrm{IN}}$, being the parameter EPR_h the energy to power ratio for hydropower plant h.

Since constraints (2.50)–(2.52) control the storage level only at checkpoints, in such an approach there is no guarantee that the energy level in any hour different from the checkpoints is within bounds. To ensure the respect of storage bounds in every hour of the planning horizon, while avoiding the application of hourly constraints on the storage level that would dramatically increase the computational burden, we propose the following iterative procedure:

- Solve the optimization model by only controlling the storage level at checkpoints;
- According to the model outputs, reconstruct the hourly energy levels of storage facilities across the whole planning horizon;
- If the hourly energy levels are within their respective bounds, stop, otherwise go to step 4;
- Update equations (2.52) by reducing of a given factor (e.g., 5%) the upper bounds of constraints corresponding to all the couples of checkpoints that include a violation of the hourly storage level and go back to step 1.

Moreover, the application of constraints (2.50)–(2.52) allows creating the continuity in long-term storage modeling while using representative days, without performing any preliminary analysis. However, the drawback of this approach is an increasing computational complexity given by the interconnection between representative days. As previously explained, the choice of the approach to be used to capture the seasonality of hydroelectric generation should be made considering the specific objectives of the analysis performed.

2.5 Determining the initial ON/OFF status of thermal power plants in representative days

The use of representative days raises also the crucial issue regarding how these days should be linked in the expansion planning model to properly apply thermal unit commitment constraints without overestimating start-up costs. Most of the existing methods, such as [42], consider the representative days as temporally consecutive, linking these days according to an arbitrary order, from which, however, the model results may be affected. Other works, such as [33], assume that each representative day

is followed by similar days and, thus, state that the initial status of each representative day should be equal to its final status. Specifically, in this approach new decision variables are introduced to represent the initial status of each generating plants and constraints are imposed to enforce the equality between initial and final status of each representative period. However, this approach is not appropriate for extreme days, which represent unique conditions. Extreme days are usually preceded by average days, thus considering the initial status of an extreme day equal to its final status could bias the calculation of the optimal amount of flexibility. More sophisticated approaches, such as [43], connect representative days by computing the transition matrix, which gives the number of transitions between each pair of representative days. However, this approach implies the introduction of many constraints and the interconnection among days, which can consistently increase computational costs.

The simplest approach to deal with temporally disconnected representative days is to assume that all thermal plants are offline at the beginning of each representative day. In this approach many start-up manoeuvres need to take place in the first hour of each cluster in order to supply load, with the following consequences: (i) an over-estimation of start-up costs and (ii) a distortion of the system operation, since units with low start-up costs may result preferable with respect to plants supplying base-load, which usually have lower production costs but higher start-up costs. Thus, in order to provide an accurate solution to the unit commitment problem while maintaining representative days separate, it is necessary to apply a method that could accurately predict the online or offline status of every thermal power plant at the beginning of each representative day. In this section, different classification techniques to initialize representative days are introduced.

Specifically, the objective of a classifier is to assign the ON/OFF status at the beginning of a representative day of every thermal plant according to its technical characteristics, which affect the commitment decisions. Certainly, the commitment decisions depend on marginal costs: the lower the production cost, the higher the commitment. However, the marginal cost is not the only relevant attribute, since also unit flexibilities play an important role in determining commitment decisions, especially if power systems have large shares of renewables. Thus, to determine the initial ON/OFF status of each power plant, the following features are considered: (i) marginal cost ratio, i.e., the ratio between unit marginal cost and average marginal cost of available thermal plants; (ii) start-up cost; (iii) minimum up time; and (iv) minimum down time.

To classify the ON/OFF status of thermal plants according to the previous features, the classifier needs to learn a classification rule from a set of observed data. The greater

the number of instances in the training set, the better the ability of the classifier to identify patterns. In our approach, a large dataset is constructed by considering the commitment decisions of Italian thermal power plants in years 2015 and 2016. Specifically, for every thermal power plant k the considered features are marginal cost ratio, start-up cost, minimum up time and minimum down time; from data we also derive the vector of daily initial statuses $\gamma^d_{k\,0}$, which describes the ON/OFF status of thermal power plant k in the last hour of day $d-1$, $2 \leq d \leq 730$. The obtained data set is partitioned in two subsets: a training set containing 70% of total observations randomly selected and a test set including the remaining 30% observations. The training set is used to estimate a classification rule, which is then applied to the test set in order to assign to each thermal power plant a probability π_k^{ON} to be ON at the beginning of the day according to features values. By comparing for each thermal power plant the probability π_k^{ON} with the ON frequency on the test set (i.e., the ratio between the number of days in the test set in which the unit is ON and the total number of instances in the test set), the accuracy of a classification technique can be evaluated. Specifically, in the current application four methods are compared: logistic regression, artificial neural networks, decision trees and support vector machines.

2.5.1 Logistic Regression

Logistic regression is a regression model frequently used to determine the dependence of a dichotomous response variable, describing the presence or absence of an attribute of interest, from a vector of observed covariates [45]. Specifically, let y_i be a binary response variable: y_i can be seen as a realization of a random variable Y_i that may take the values 1 and 0 with probabilities π_i and $1-\pi_i$, respectively. The aim of a logistic model is to express probabilities π_i as a function of a vector of observed covariates x_i. The simplest approach would be to consider probabilities π_i being a linear function of the covariates

$$\pi_i = x_i'\beta \tag{2.53}$$

where β is a vector of regression coefficients. However, probabilities π_i have to be between 0 and 1, while the linear combination of covariates can take any real value, not guaranteeing that predicted probabilities are in the correct range.

The range restrictions are solved by the following transformation. First, we move from probabilities to the *odds*, i.e., the ratio of the probability to its complement

$$\text{odds}_i = \frac{\pi_i}{1-\pi_i} \tag{2.54}$$

Then, range restrictions are removed by computing the *logit* function

$$\text{logit}(\pi_i) = ln\frac{\pi_i}{1-\pi_i} \tag{2.55}$$

Indeed, as π_i goes to 0, the *odds* approaches 0 as well, while the *logit* approaches $-\infty$. Instead, as the probability goes to 1, both the *odds* and the *logit* approach $+\infty$.

Finally, logistic regression models assume that the *logit* of probability π_i is a linear function of the covariates

$$\text{logit}(\pi_i) = x_i'\beta \tag{2.56}$$

Solving equation (2.56) for the probability π_i gives the following equation

$$\pi_i = \frac{e^{x_i'\beta}}{1+e^{x_i'\beta}} \tag{2.57}$$

In the current application, observations in the training set are used to estimate regression coefficients β in equation (2.56). The estimations for regression coefficients are reported in Table 2.4 along with the correspondent standard deviations and the p-values. Specifically, for each term the p-value tests the null hypothesis that the regression coefficient is equal to zero, i.e., that predictor is not statistically significant. In this case, as can be noticed, all covariates present a very low p-value, indicating that all the predictors are meaningful in the model.

Table 2.4 Regression coefficients estimated on the training set

Covariate	Coefficient	Standard Deviation	p-value
Intercept	35.98	1.04	$< 10^{-15}$
MDT	0.37	0.07	$< 10^{-7}$
MUT	-0.18	0.01	$< 10^{-15}$
MC Ratio	-36.65	1.06	$< 10^{-15}$
SUC	$-1.18 \cdot 10^{-5}$	$3.98 \cdot 10^{-7}$	$< 10^{-15}$

Moreover, as expected, regression coefficient for the marginal cost ratio is negative: efficient thermal plants with low production costs have greater probabilities to be ON at the beginning of the day rather than power plants with high marginal costs. Also the coefficient for the start-up cost is negative, indicating that the probability to be ON at the beginning of the day increases as the start-up costs decrease.

Regression coefficients estimated on the training set are then used to compute probabilities of thermal plants in the test set having an initial ON status according to equation (2.57).

2.5.2 Artificial Neural Networks

An artificial neural network is a system based on biological neural system consisting of a pool of simple processing units that communicate by sending signals to each other over a large number of weighted connections [46].

Each unit in an artificial neural network is called an artificial neuron and it receives inputs from other units or external sources and elaborate this information to compute an output signal, which is propagated to other neurons. Fig. 2.8 illustrates the mathematical structure of an artificial neuron. The basic idea in this model is that each unit provides an additive contribution to the input of the neurons with which it is connected. As can be noticed, the net input to each neuron is a linear combination of inputs provided by other units and by an external source.

The weights in this linear combination give the strength of the connections: while negative weights represent inhibitory connections, positive coefficients reflect excitatory connections. Finally, the activation function performs a non-linear operation on the net input to produce an activation signal, which is then propagated to other neurons.

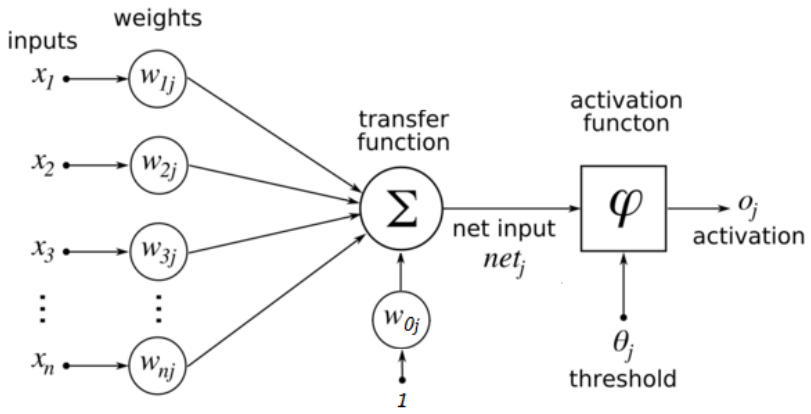


Fig. 2.8 Mathematical model of an artificial neuron

Artificial neural networks usually present a layered structure. Each layer consists of neurons that receive signals from units in a layer directly below and send their output to neurons in a layer directly above. No connections exist between neurons within a layer. Specifically, layers in an artificial neural network can be divided into three groups: (i) input layer, containing neurons that receive data only from external sources; (ii) output layer, consisting of units which send signals out of the neural network; and (iii) hidden layer, consisting of neurons that receive information from other units and send output signals to other neurons.

In the current application, a neural network with only three layers (i.e., one input layer, one output layer and only one hidden layer) is considered in order to avoid the overfitting phenomenon that could be induced by introducing many neurons. Specifically, the input layer consists of four neurons, one for each feature, while the output layer includes only one unit.

The signal provided by the output neuron is a number in the range of [0; 1] describing the probability of thermal plants having an initial ON status. The number of neurons in the hidden layer, a priori unknown, can be determined by testing different configurations. A network with three neurons in the hidden layer proved to be the best configuration in terms of classification error on the test set.

Using observations in the training set, the weights of the neurons connections are determined so as to obtain the desired outputs from the given set of inputs. The description of the method to set the weights is outside the scope of this work. We refer the reader to [46] for a detailed description of networks training. Fig. 2.9 illustrates the artificial neural network estimated on training data, which is then used to determine probabilities of thermal plants in the test set to have an initial ON status.

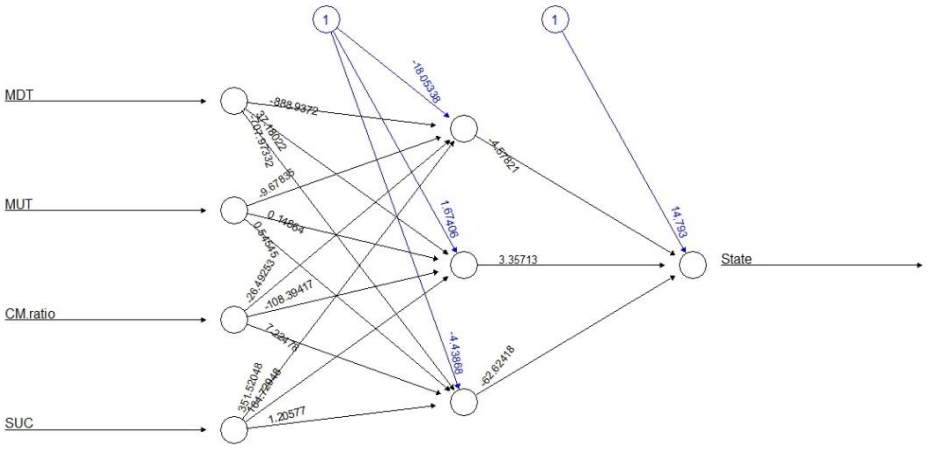


Fig. 2.9 Artificial neural network estimated on training data

2.5.3 Decision Trees

A decision tree is a classifier expressed as a recursive partition of the instance space. The decision tree consists of nodes that form a rooted tree with a node called a root that has no incoming edges [47]. All other nodes have exactly one incoming edge. A node with outgoing edges is called a test node and splits the instance space into two or more sub-spaces according to attributes values. Each node with no outgoing edges is called a leaf and it represents a group of data with specific attributes values. In a decision tree, each leaf is assigned to one class. Alternatively, the leaf may hold a probability vector indicating the probability of the target attribute having a certain value.

The construction of a decision tree from a given dataset, the so called *induction*, is usually performed by algorithms that apply a top-down recursive approach. First, the feature is found which best splits data into two (or more) groups. Data are then partitioned in subsets according to the splitting rule and the same approach is applied to each subset, until a stopping criterion is satisfied. Typical stopping criteria are: (i) the impossibility to further improve the classification; (ii) a minimum number of instances in a node; or (iii) the maximum depth of the tree.

At each iteration, algorithms for decision tree induction identify the splitting criterion by considering an impurity measure, which in many applications is the Gini index. Specifically, let us consider a classification problem with k groups. Let $p_i, 1 \leq i \leq k$, be the relative frequency of class i in a node of the decision tree. Then, the Gini index in that node is defined as follows

$$Gini = 1 - \sum_{i=1}^{k} (p_i)^2 \tag{2.58}$$

As can be easily understood, a pure node has a Gini index of 0, while the Gini index approaches 1 as the variability inside the node increases. Algorithms that use Gini index determine the splitting criterion in each node as the partition that most reduces the node impurity measured by equation (2.58). We refer the reader to [48] for a detailed description of impurity measures and decision trees training.

In our application we set at four the maximum depth of the tree, in order to avoid overfitting, and used the Gini index as impurity measure. The decision tree induced on the training set is shown in Fig. 2.10.

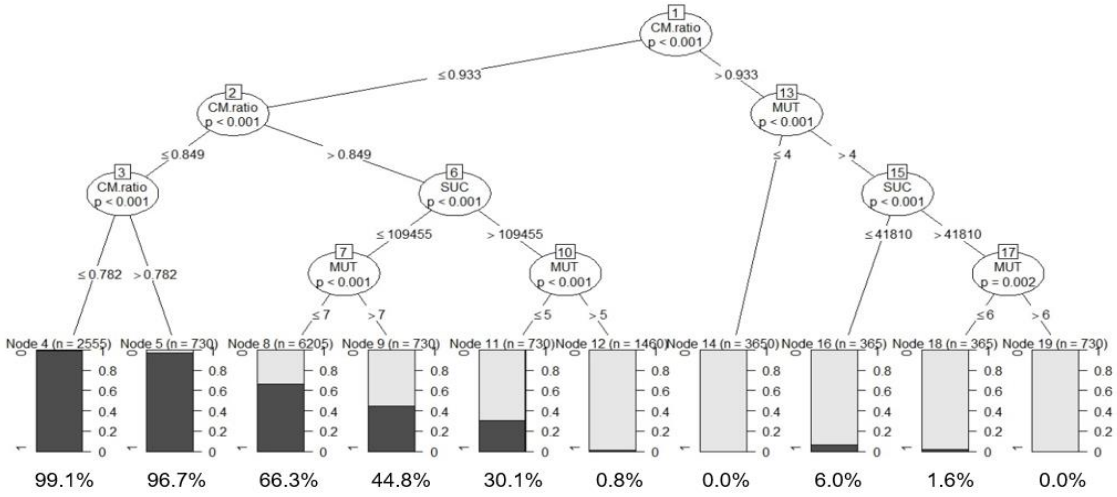


Fig. 2.10 Decision tree induced on training set

As can be noticed, the marginal cost ratio is the most relevant attribute in the considered classification, as it realizes a clear partition: while units with marginal cost ratios greater than 0.933 are usually offline, the initial status for thermal plants with marginal cost ratios lower than or equal to 0.849 is ON. Instead, for marginal cost ratios in the range (0.849 ; 0.933] the initial status is more uncertain and it depends also on the other attributes values. Specifically, while thermal units with high start-up costs are usually offline, thermal plants with lower start-up costs are more often committed at the beginning of the day, especially if they have enough flexibility.

Given the tree induced on training data, instances in the test set are classified simply by navigating them from the root of the tree down to a leaf, according to the outcome of the tests along the path.

2.5.4 Support Vector Machines

Support Vector Machines (SVMs) are a set of machine learning methods introduced in the late 1960s by Vapnik [49] to address classification and regression problems. Specifically, let us consider a classification problem. Let $(x_i, y_i),\ 1 \leq i \leq N$, be a training set with input data $x_i \in \mathbb{R}^n$ and corresponding binary labels $y_i \in \{-1; 1\}$ and let $\varphi(\cdot) : \mathbb{R}^n \to \mathbb{R}^m,\ m > n$ be a nonlinear function that maps empirical data into a higher dimensional space, the so-called feature space. The aim of the SVM classifier formulation is to determine a hyperplane in the feature space $w^T\varphi(x_i) + b = 0$ dividing the training set so that

$$y_i[w^T\varphi(x_i) + b] \geq 1 \qquad i = 1, \dots, N \tag{2.59}$$

To construct an optimal hyperplane, the following quadratic problem is solved

$$\begin{aligned} \min_{\omega,b,\xi} \quad & \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i \\ \text{subject to} \quad & y_i[w^T\varphi(x_i) + b] \geq 1 - \xi_i \qquad i = 1, \dots, N \end{aligned} \tag{2.60}$$

where ξ_i are nonnegative variables that allow misclassifications in the training set. Specifically, the minimization of $\|w\|^2$ in the objective function of problem (2.60) corresponds to the maximization of the margin of separation between classes. The positive constant C creates a trade-off between the classification error in the training set and the separation of the rest samples with maximum margin.

A way to solve (2.60) is via its dual formulation, which is the following quadratic programming problem:

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} y_i y_j K(x_i, x_j)\alpha_i\alpha_j \\ \text{subject to} \quad & \sum_{i=1}^{N}\alpha_i y_i = 0 \\ & 0 \leq \alpha_i \leq C \qquad i = 1, \dots, N \end{aligned} \tag{2.61}$$

where $K(x_i, x_j) = \varphi(x_i)^T\varphi(x_j)$ is a Kernel and Lagrange multipliers α_i are nonnegative real variables.

Finally, the non-linear SVM classifier becomes

$$y(x) = \text{sign}\left[\sum_{i=1}^{N}\alpha_i y_i K(x, x_i) + b\right] \tag{2.62}$$

with b being a real constant that follows from the KKT conditions. The nonzero Lagrange multipliers α_i are called support values. The corresponding data points are called support vectors and are located close to the decision boundary.

Several choices for the kernel $K(\cdot,\cdot)$ are possible [50]. In our application, we use a radial basis function (RBF) kernel, also known as Gaussian kernel:

$$K(x_i, x_j) = e^{-\gamma \|x_i - x_j\|^2} \tag{2.63}$$

Parameter γ in equation (2.63) as well as constant C in problem (2.60) are considered in the current analysis as tuning parameters: their values are determined by testing different options. After the tuning activity, parameter γ is set to 0.1, while constant C is set to 100 and the corresponding model trained on training data is used to predict the initial status of thermal plants in the test set according to equation (2.62).

2.5.5 Comparison of classification techniques

As previously explained, all the models trained on input data are then used to classify new instances in the test set. Specifically, in order to obtain more reliable information, we repeat 30 times the dataset division into the training and the test set. In this way, confidence intervals for the models performances can be computed. Table 2.5 reports for every technique the confidence interval for the classification accuracy, i.e., the percentage of instances in the test set correctly classified.

As can be noticed, decision trees are the most accurate method, with an average accuracy of 94.4%. Even the SVM classifier provide excellent results, having an average accuracy greater than 90% but with more variability than decision trees. The confidence interval width is maximum for artificial neural networks, which correctly classify on average less than 90% of observations in the test set. Finally, logistic regression model presents the lowest accuracy.

In addition to the highest accuracy, another advantage of decision trees, especially if compared to SVMs and neural networks, is the interpretability, since with this technique it is easy to visualize and understand classification rules.

Table 2.5 Accuracy of the different classifiers

Technique	Classification Accuracy (%)
Logistic Regression	82.9 ± 2.1
Artificial Neural Network	88.7 ± 3.9
Decision Tree	94.4 ± 1.2
SVM	91.2 ± 3.5

Thus, in the current application we select the decision tree as classifier to determine the ON/OFF status γ_{k0}^{c} of thermal power plant k at the beginning of representative day c according to a procedure that can be summarized as follows:

- A dataset containing historical commitment decisions is considered.
- For each thermal power plant in the dataset, four features are considered: marginal cost ratio, start-up cost, minimum up time and minimum down time.
- A decision tree with maximum depth equal to 4 is induced on the given dataset to estimate a classification rule to be used to determine the initial ON status according to features values.
- The decision tree is used to determine the probability π_k^{ON} of each thermal power plant k having an initial ON status.
- Parameters π_k^{ON} are used to set the probability of extracting 1 in the random selection between 1 (i.e., ON) and 0 (i.e., OFF). For each thermal power plant, this random selection is repeated for all representative days, in order to assign to each representative day $c \in \mathcal{C}$ a specific initial status γ_{k0}^{c}.

2.5.6 Evaluating the robustness of the classifier

The results described in the previous section show the accuracy of decision trees in correctly identifying the initial statuses of the thermal power plants in the test set. However, in the previous tests, instances in the training and in the test set were the same. The availability of historical data for the induction of the decision trees could be a very heavy drawback of this approach. For instance, when assessing future scenarios, new thermal power plants for which no observations are available are typically included in the analysis. Thus, more tests are needed to evaluate the robustness of decision trees, by analysing the ability of these classifiers to correctly identify the initial statuses of thermal power plants in scenarios that differ from the one used for the decision tree induction.

To this end, the decision tree estimated on the Italian historical data (*Base* scenario) has been applied to classify instances in two future scenarios for the Italian power system elaborated by CESI S.p.A.:

- *Gas-before-Coal* (GbC): in this scenario the coal price is expected to increase more than the gas price, leading to a switch in the merit order of thermal power plants.

- *Coal-before-Gas* (CbG): in this scenario the gas price is expected to slightly increase, while the coal price decreases.

Fig. 2.11 shows the average marginal production costs for three thermal technologies, namely coal, Combined Cycle Gas Turbine (CCGT) and Gas Turbine (GT), in scenarios Base, GbC and CbG. As can be noticed, the average marginal production cost for coal power plants changes from 59 €/MWh in the Base scenario, to 98 €/MWh in the GbC scenario, to 56 €/MWh in the CbG scenario. CCGT plants are characterized by a marginal cost equal to 65 €/MWh in the Base scenario, to 92 €/MWh in the GbC scenario and to 71 €/MWh in the CbG scenario. Finally, the average production cost for GT plants changes from 100 €/MWh in the Base scenario, to 136 €/MWh in the GbC scenario, to 105 €/MWh in the CbG scenario.

The initial statuses provided for the GbC and CbG scenarios by the decision tree estimated on the Base scenario have been compared with the initial statuses of thermal power plants obtained by solving the thermal unit commitment model on the two new scenarios. The results of this comparison are reported in Table 2.6. As can be observed, the proposed classifier shows a great robustness, as it estimates with high accuracy the initial statuses in scenarios which are significantly different from the dataset used for the decision tree induction. Indeed, classification accuracies (i.e., the percentage of instances correctly identified) in the new scenarios are roughly equal to the classification accuracy in the Base scenario, although production costs and merit order between thermal power plants are different.

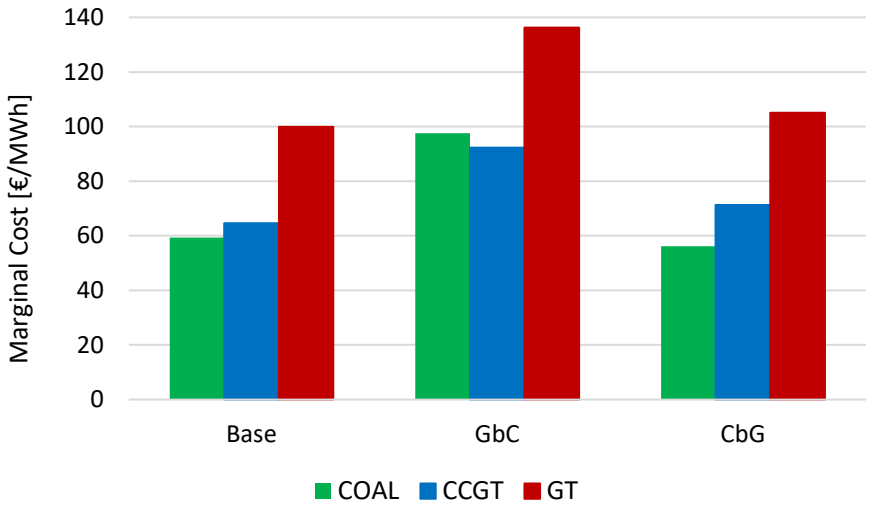


Fig. 2.11 Average marginal production cost for technology in different scenarios

Table 2.6 Classification accuracy of the decision tree in different scenarios

Scenario	Classification Accuracy (%)
Base	94.4
Gas-before-Coal	93.9
Coal-before-Gas	92.3

The robustness of the classifier is a very crucial result for generation and transmission expansion planning problem. Indeed, this problem is characterized by planning horizon of several decades, therefore operational costs for thermal power plants at the end of the planning horizon may dramatically differ from the costs at the beginning. However, as discussed in this paragraph, the application of the proposed procedure allows obtaining reliable results even when changes in the thermal merit order occur.

2.6 Evaluating short-term operations through representative days

In this section, we present the short-term operational model based on the procedure introduced in the previous sections. In particular, such a model evaluates power system operations through representative days while capturing the seasonality of the hydroelectric dispatch and providing an accurate estimate of the start-up decisions that occur in the first hours of representative days.

2.6.1 Notation

For the sake of clarity, in this paragraph we provide the full notation needed to formulate the short-term operational problem.

Sets

$\mathcal{Z}$	Set of zones, indexed by z
$\mathcal{K}$	Set of thermal power plants, indexed by k
$\Omega_z^k \subset \mathcal{K}$	Set of thermal power plants located in zone z
$\mathcal{H}$	Set of hydropower plants, indexed by h
Ω_z^h	Set of hydropower plants located in zone z
$\mathcal{D}$	Set of days in the planning horizon, indexed by d and d'
$\mathcal{D}_M \subset \mathcal{D}$	Set of days in the planning horizon in which the level of the long-term storage is checked

$\mathcal{C}$	Set of representative days, indexed by c
$\mathrm{Map}_{d,c}$	Cluster index, i.e., injective map of each day d to a representative day c
$\mathcal{T}$	Set of hours of representative days, from 1 to 24, indexed by t and τ
$\mathcal{L}$	Set of transmission lines, indexed by l
$rz(l)$	Receiving-end zone of transmission line l
$sz(l)$	Sending-end zone of transmission line l

Parameters

wg_c	[−]	Weight of representative day c
c_{ENP}	[€/MWh]	Penalty for energy not provided
c_{OG}	[€/MWh]	Penalty for over-generation
CM_k	[€/MWh]	Marginal production cost of thermal power plant k
$\underline{P}_k$	[MW]	Minimum power output of thermal power plant k
$\overline{P}_k$	[MW]	Maximum power produced by thermal plant k
SUC_k	[€]	Start-up cost of thermal power plant k
$\gamma^c_{k\,0}$	[−]	Initial ON/OFF status of thermal power plant k in representative day c
MUT_k	[h]	Minimum up time of thermal power plant k
MDT_k	[h]	Minimum down time of thermal power plant k
$Cvar_h$	[€/MWh]	Operating cost of hydropower plant h
$\bar{E}_h^{\mathrm{IN}}$	[MW]	Upper bound on hydropower plant h pumping power
$\bar{E}_h^{\mathrm{OUT}}$	[MW]	Upper bound on hydropower plant h power output
$\overline{sl}_h$	[MWh]	Upper bound on energy spillage from hydropower plant h
$Q_{h,c}$	[MWh]	Daily energy balance for hydropower plant h in representative day c
EPR_h	[h]	Maximum energy to power ratio (in hours) for hydropower plant h
$F^c_{h,t}$	[MWh]	Hourly energy inflows for hydropower plant h at time t of representative day c
λ_h^{IN}	[−]	Loss coefficient for hydro plant h pumping $\left(0 \le \lambda_h^{\mathrm{IN}} \le 1\right)$
λ_h^{OUT}	[−]	Loss coefficient for hydro plant h power generation $\left(\lambda_h^{\mathrm{OUT}} \ge 1\right)$
$E_{h\,0}$	[MWh]	Energy content of hydropower plant h at the beginning of planning horizon

M	[−]	Size of the temporal window in long-term storage constraints, set to 7 days
$\underline{F_l}$	[MW]	Minimum capacity of transmission line l
$\overline{F_l}$	[MW]	Maximum capacity of transmission line l
$NL_{z,t}^{c}$	[MW]	Net load in zone z in hour t of representative day c
$R_{z,t}^{c}$	[MW]	Reserve requirement for zone z in hour t of representative day c

Variables

$\gamma_{k,t}^{c}$	[−]	Commitment status of unit k at time t of representative day c, which is equal to 1 if the unit is online and 0 otherwise
$\alpha_{k,t}^{c}$	[−]	Start-up status of unit k at time t of representative day c, which is equal to 1 if the unit is started-up and 0 otherwise
$\beta_{k,t}^{c}$	[−]	Shut-down status of unit k at time t of representative day c, which is equal to 1 if the unit is shut-down and 0 otherwise
$p_{k,t}^{c}$	[MW]	Power production of thermal power plant k at time t of representative day c above its minimum output $\underline{P_k}$
$E_{h,t}^{\mathrm{IN},c}$	[MW]	Pumping power of hydro reservoir h in hour t of representative day c
$E_{h,t}^{\mathrm{OUT},c}$	[MW]	Power output of hydro reservoir h in hour t of representative day c
$sl_{h,t}^{c}$	[MWh]	Energy spillage from hydro reservoir h in hour t of representative day c in scenario w
$\hat{E}_{h}^{\mathrm{LT},d}$	[MWh]	Energy level of hydro reservoir h at the end of day d
$x_{l,t}^{c}$	[MW]	Power flow on transmission line l in hour t of representative day c
$ENP_{z,t}^{c}$	[MWh]	Energy not provided in zone z in hour t of representative day c
$OG_{z,t}^{c}$	[MWh]	Over-generation in zone z in hour t of representative day c

2.6.2 Mathematical formulation

In this section we introduce two formulations of the hydrothermal unit commitment problem with representative days, which differ for the approach used to capture seasonality of long-term storage. The first formulation, based on the preliminary hydroelectric analysis, is given by the following MILP model

$$\min\ z = \sum_{c\in\mathcal{C}} wg_c \sum_{t=1}^{24} \left(\sum_{k\in\mathcal{K}} CM_k \left(\underline{P}_k\, \gamma_{k,t}^c + p_{k,t}^c\right) + \sum_{k\in\mathcal{K}} SUC_k \alpha_{k,t}^c + \sum_{h\in\mathcal{H}} Cvar_h\, E_{h,t}^{\mathrm{OUT},c} + c_{ENP} \sum_{z\in Z} ENP_{z,t}^c + c_{OG} \sum_{z\in Z} OG_{z,t}^c \right) \tag{2.64}$$

subject to

$$p_{k,t}^c \le \gamma_{k,t}^c \left(\overline{P}_k - \underline{P}_k\right) \qquad k\in\mathcal{K}, c\in\mathcal{C}, 1\le t\le 24 \tag{2.65}$$

$$\gamma_{k,t}^c - \gamma_{k\,0}^c = \alpha_{k,t}^c - \beta_{k,t}^c \qquad k\in\mathcal{K}, c\in\mathcal{C}, t=1 \tag{2.66}$$

$$\gamma_{k,t}^c - \gamma_{k,t-1}^c = \alpha_{k,t}^c - \beta_{k,t}^c \qquad k\in\mathcal{K}, c\in\mathcal{C}, 2\le t\le 24 \tag{2.67}$$

$$\sum_{\tau=t-MUT_k+1}^{t} \alpha_{k,\tau}^c \le \gamma_{k,t}^c \qquad k\in\mathcal{K}, c\in\mathcal{C}, MUT_k\le t\le 24 \tag{2.68}$$

$$\sum_{\tau=t-MDT_k+1}^{t} \beta_{k,\tau}^c \le 1-\gamma_{k,t}^c \qquad k\in\mathcal{K}, c\in\mathcal{C}, MDT_k\le t\le 24 \tag{2.69}$$

$$E_{h,t}^{\mathrm{IN},c} \le \bar{E}_h^{\mathrm{IN}} \qquad h\in\mathcal{H}, c\in\mathcal{C}, 1\le t\le 24 \tag{2.70}$$

$$E_{h,t}^{\mathrm{OUT},c} \le \bar{E}_h^{\mathrm{OUT}} \qquad h\in\mathcal{H}, c\in\mathcal{C}, 1\le t\le 24 \tag{2.71}$$

$$\sum_{t=1}^{24} \left(\lambda_h^{\mathrm{OUT}} E_{h,t}^{\mathrm{OUT},c} - \lambda_h^{\mathrm{IN}} E_{h,t}^{\mathrm{IN},c}\right) = Q_{h,c} \qquad h\in\mathcal{H}, c\in\mathcal{C} \tag{2.72}$$

$$\underline{F}_l \le x_{l,t}^c \le \overline{F}_l \qquad l\in\mathcal{L}, c\in\mathcal{C}, 1\le t\le 24 \tag{2.73}$$

$$\sum_{k\in\Omega_z^k} \left(\underline{P}_k\, \gamma_{k,t}^c + p_{k,t}^c\right) + \sum_{l|rz(l)=z} x_{l,t}^c + \sum_{h\in\Omega_z^h} E_{h,t}^{\mathrm{OUT},c} + ENP_{z,t}^c = NL_{z,t}^c + \sum_{l|sz(l)=z} x_{l,t}^c + \sum_{h\in\Omega_z^h} E_{h,t}^{\mathrm{IN},c} + OG_{z,t}^c \qquad z\in\mathcal{Z}, c\in\mathcal{C}, 1\le t\le 24 \tag{2.74}$$

$$\sum_{k\in\Omega_z} \left[\left(\overline{P}_k - \underline{P}_k\right)\gamma_{k,t}^c - p_{k,t}^c\right] \ge R_{z,t}^c \qquad z\in\mathcal{Z}, c\in\mathcal{C}, 1\le t\le 24 \tag{2.75}$$

$$\gamma_{k,t}^{c}, \alpha_{k,t}^{c}, \beta_{k,t}^{c} \in \{0,1\} \qquad k \in \mathcal{K}, c \in \mathcal{C}, 1 \leq t \leq 24 \qquad (2.76)$$

$$p_{k,t}^{c} \geq 0 \qquad k \in \mathcal{K}, c \in \mathcal{C}, 1 \leq t \leq 24 \qquad (2.77)$$

$$E_{h,t}^{\mathrm{IN},c}, E_{h,t}^{\mathrm{OUT},c} \geq 0 \qquad h \in \mathcal{H}, c \in \mathcal{C}, 1 \leq t \leq 24 \qquad (2.78)$$

$$x_{l,t}^{c} \text{ free variable} \qquad l \in \mathcal{L}, c \in \mathcal{C}, 1 \leq t \leq 24 \qquad (2.79)$$

$$ENP_{z,t}^{c}, OG_{z,t}^{c} \geq 0 \qquad z \in \mathcal{Z}, c \in \mathcal{C}, 1 \leq t \leq 24. \qquad (2.80)$$

The objective function (2.64) computes the total costs by multiplying the daily operational costs in representative days by their weights. For each representative day operational costs include thermal power production costs, start-up costs, hydropower operational costs, penalties for energy not provided and penalties for over-generation.

Constraints (2.65)–(2.69) control the operations of thermal power plants as explained in Section 2.2.3. Specifically, parameters $\gamma_{k\,0}^{c}$ in equations (2.66) are initial statuses determined with the procedure summarized in Section 2.5.5. It is worth mentioning that representative days are kept separate in the optimization model: indeed, while equations (2.67) apply to all hours of representative days except the first one, minimum up/down time constraints (2.68)/(2.69) are enforced on the hours in the range from $MUT_k/MDT_k$to the final (the 24th) of the day.

Equations (2.70)–(2.72) control the operations of the hydropower plants by imposing upper bounds to the pumping power and the hydroelectric production and by setting the daily hydroelectric energy balance to the value determined by the preliminary hydroelectric model, as explained in Section 2.4.1. The values of the daily energy balances $Q_{h,c}$ associated to representative days $c \in \mathcal{C}$, which appear in (2.72), are determined by solving the preliminary hydroelectric model.

Constraints (2.73) control power flows on transmission lines. Equations (2.74) ensure net load supply. Constraints (2.75) are the spinning reserve constraints. Finally, equations (2.76)–(2.80) define the optimization variables.

When modeling long-term storage by the alternative approach based on inter-day balance equations, constraints (2.72) are replaced by the following (see Section 2.4.2 above):

$$\hat{E}_h^{\mathrm{LT},d} = E_{h_0} + \sum_{d'=d-M+1}^{d} \sum_{c\in \mathrm{Map}_{d',c}} \sum_{t=1}^{24} \left(F_{h,t}^{c} + \lambda_h^{\mathrm{IN}} E_{h,t}^{c,\mathrm{IN}} - \lambda_h^{\mathrm{OUT}} E_{h,t}^{c,\mathrm{OUT}} - sl_{h,t}^{c}\right) \qquad h \in \mathcal{H}, d = M \tag{2.81}$$

$$\hat{E}_h^{\mathrm{LT},d} = \hat{E}_h^{\mathrm{LT},d-M} + \sum_{d'=d-M+1}^{d} \sum_{c\in \mathrm{Map}_{d',c}} \sum_{t=1}^{24} \left(F_{h,t}^{c} + \lambda_h^{\mathrm{IN}} E_{h,t}^{c,\mathrm{IN}} - \lambda_h^{\mathrm{OUT}} E_{h,t}^{c,\mathrm{OUT}} - sl_{h,t}^{c}\right) \qquad h \in \mathcal{H}, d \in \mathcal{D}_M, d > M \tag{2.82}$$

$$0 \le \hat{E}_h^{\mathrm{LT},d} \le EPR_h \bar{E}_h^{\mathrm{IN}} \qquad h \in \mathcal{H}, d \in \mathcal{D}_M \tag{2.83}$$

$$0 \le sl_{h,t}^{c} \le \overline{sl}_h \qquad h \in \mathcal{H}, 1 \le t \le 24, c \in \mathcal{C} \tag{2.84}$$

Thus, with inter-day balance equations approach the power system operation model is given by (2.64)–(2.71) and (2.73)–(2.84).

2.7 Case studies and results

To assess the proposed procedure performance in evaluating short-term operations, two tests have been conducted on scenarios elaborated by CESI S.p.A.

2.7.1 Tests on a reduced Italian scenario

The first scenario is a simplified representation of the power system in Southern Italy in year 2017, which consists of three market zones (Central-South, South and Sicily) connected in a tree network. As the test focus was on the system thermal component, a power system consisting of 48 thermal power plants and no hydropower plants has been considered in the test.

An hourly unit commitment problem has been applied to this scenario, so as to obtain a benchmark to evaluate the accuracy of the method proposed in this chapter. To identify representative days, we applied the procedure described in Section 2.3.2 fixing a threshold of 1%. Fig. 2.12 shows the system average mean absolute percentage error in load duration curves approximation for different numbers of representative days. As can be noticed, the higher the number of representative days, the better the approximation of the original curve. Moreover, by using seven representative days the input threshold is satisfied, with the average mean absolute percentage error being 0.97%.

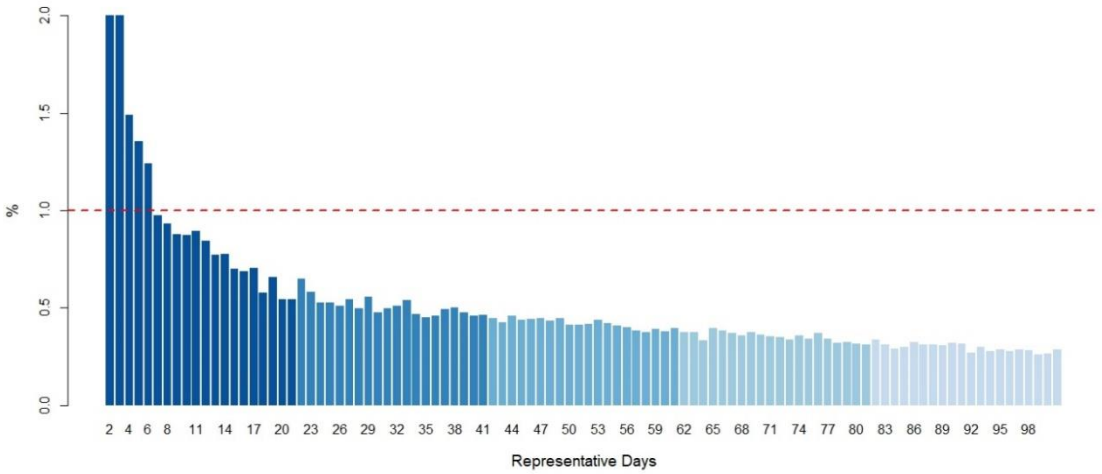


Fig. 2.12 System average mean absolute percentage error in load duration curves approximation for different numbers of representative days

Fig. 2.13 illustrates the seven representative days in the Central-South zone. The dashed and dotted lines represent extreme days with the highest and the lowest demand values, respectively; the solid lines show the daily profiles identified by the k-medoids algorithm.

Since no hydropower plants are included in the scenario for Southern Italy, we did not have to address the problem of capturing the seasonality of long-term storage. In order to determine the initial statuses of thermal plants in each representative day, we considered the thermal power plants commitment in years 2015 and 2016. We then estimated on this data the decision tree already introduced in Fig. 2.10, which assigns to each thermal power plant a specific probability to be online in the initial hour of the representative days according to the attributes values.

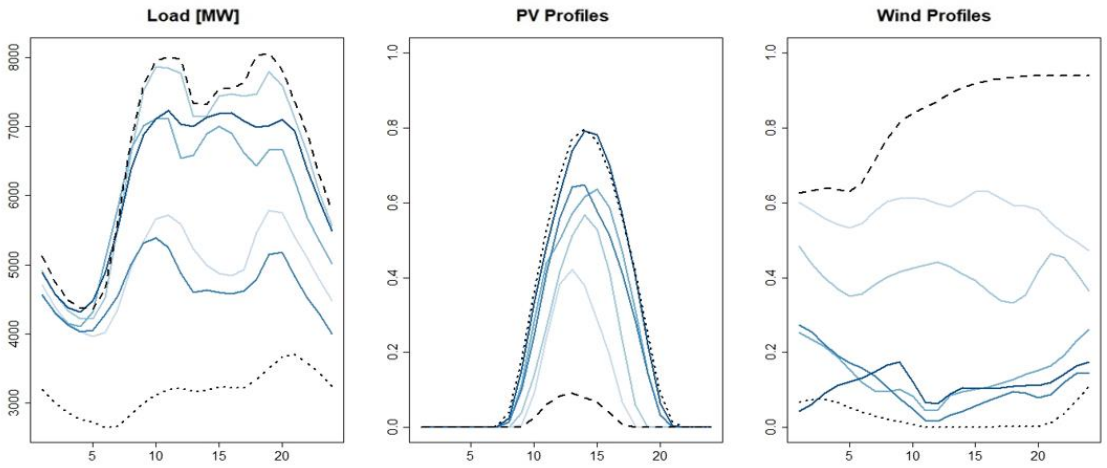


Fig. 2.13 Daily load, solar and wind profiles in the Central-South in the seven representative days

Table 2.7 Comparison between the hourly and the clustered unit commitment model for the Italian scenario

Model	Days	Total Costs [10^9€]	Production Costs [10^9€]	Start-up Costs [10^9€]	Total Error [%]	Solution time [seconds]
Hourly	365	4.751	4.735	0.016	–	3334
Clustered	7	4.752	4.733	0.019	0.01%	9

As described in Section 2.5.3, by performing for each thermal power plant a random extraction between ON and OFF with probabilities set by the decision tree, binary parameters $\gamma^c_{k_0}$ describing the initial ON or OFF status of thermal power plant k in representative day c have been determined.

Table 2.7 compares costs and computational times for the hourly model and the clustered model. In the tests, the penalty for energy not provided was set to 10000 €/MWh, while the penalty for over-generation was set to 200 €/MWh. Results for the Italian scenario have been obtained with an ASUS laptop with a 3 GHz Intel Core i7-5500U Processor and 4 GB of RAM using solver Gurobi under GAMS 24.7.4. As can be noticed, the use of representative days provides a huge computational saving, since the operational problem can be solved in only 9 seconds, while the solution time for the complete hourly formulation is 3334 seconds. Moreover, the analysis of the costs values shows how the clustered model closely replicate the output provided by the complete hourly model, being the total percentage error equal to only 0.01%.

The huge reduction of computational times, which are reduced of a factor 370, and the very small deviation from the optimal results observed on the case study demonstrate the goodness of the proposed procedure, which allows obtaining very accurate estimates of power systems operational decisions while keeping the problem computationally tractable. However, in the considered scenario only the thermal component has been tested: an additional analysis is therefore needed to test the accuracy of the proposed method for hydrothermal power systems. Such a test is described in the following paragraph.

2.7.2 Tests on the European scenario

As a case study to test the accuracy of the proposed method for power systems including also the hydroelectric component, we considered a representation of the European power system in year 2017. Specifically, such a scenario consists in 55 market zones connected by a capacitated transmission network, 1962 thermal power plants and 105 equivalent hydropower plants, divided into 37 run-of-the-river (ROR) hydroelectric plants and 68

hydroelectric plants with reservoir. Being the ROR hydroelectric plants non dispatchable power sources, the corresponding production was subtracted from the net load to obtain the zonal demands for electricity that have to be satisfied with either dispatchable power sources (i.e., thermal power plants and hydroelectric plants with reservoir) or energy import from neighbouring zones.

Given the high dimensionality of the scenario, solving the yearly hydrothermal unit commitment problem with an hourly time resolution is computationally infeasible. Thus, we obtained an approximation of the optimal hourly solution by applying the following procedure. We considered the planning horizon divided into twelve months. For each reservoir we considered the energy content at the end of each month in the previous year (i.e., 2016). Finally, we solved for each month of the planning horizon the hydrothermal unit commitment problem with hourly resolution by fixing the initial and final energy contents of reservoirs in each month to the values observed in the previous year.

Such a model has been compared with the clustered model (2.64)–(2.80) obtained as follows. First, we identified representative days by applying the procedure described in Section 2.3.2 fixing a threshold of 5% given the higher number of market zones in the scenario, obtaining 10 representative days.

As regards to the hydrothermal component, to include the long-term dispatching decisions we applied the approach described in Section 2.4.1. To estimate the linear relationship between market prices and net load we considered a dataset including historical data for years 2015 and 2016. Specifically, we chose to run a preliminary model for the hydroelectric dispatch rather than connecting representative days for the following reasons:

- In such an application, no other constraints link representative days. Thus, keeping representative days temporally disconnected allows obtaining a huge computational saving, since the yearly operational problem can be solved by solving in an independent and parallel fashion the operational problems defined for each representative day.
- The scenario analysed is very similar to scenarios used in the regression analysis, thus information provided by analysing years 2015 and 2016 can be considered reliable to predict the behaviour of the power system in year 2017.

To better consider the availability of hydroelectric energy in representative days, we applied on the hydroelectric preliminary model results the post-processing activity described in Section 2.4.1.4 assigning to each hydropower plant in each representative

day two different values for the daily hydroelectric energy balance. Thus, the final number of representative day considered in the analysis equals 20.

In order to determine the initial statuses of thermal plants in each representative day, we applied the decision tree shown in Fig. 2.10 to assign to each thermal power plant a specific probability to be ON in the initial hour of the representative days. Indeed, the analysis described in Section 2.5.6 showed the robustness of such a decision tree, which provides accurate estimates of initial statuses also for scenarios consistently different from the training set on which the decision tree has been estimated.

Table 2.8 compares costs and computational times for the hourly model and the clustered model. In the tests, penalties for energy not provided and over-generation were set to 10000 €/MWh and 200 €/MWh, respectively, while the hydropower production was supposed to be costless.

Table 2.8 Comparison between the hourly and the clustered unit commitment model for the European scenario

Model	Days	Total Costs [10^9€]	Production Costs [10^9€]	Start-up Costs [10^9€]	Total Error [%]	Solution time [minutes]
Hourly	365	47.318	47.280	0.038	–	2883.12
Clustered	20	47.443	47.399	0.044	0.27%	5.40

We solved the two models on a computer with two 2.10 GHz Intel® Xeon® Platinum 8160 CPU Processors and 128 GB of RAM, using solver Gurobi under GAMS 24.7.4. As can be noticed, also for the European scenario the use of representative days provides a huge computational saving, while only marginally distorting the optimal results. Indeed, the operational problem with representative days is solved in 5.40 minutes (i.e., 5 minutes and 24 seconds), while the solution time for the complete hourly formulation takes more than 48 hours. Moreover, the clustered model provides accurate estimates of the system costs, being the total percentage error equal to only 0.27%.

2.8 Chapter conclusions

In this chapter, the problem of evaluating power system short-term operations has been discussed. First, different formulations for the thermal unit commitment problem have been introduced and compared in terms of computational times. Our numerical experiments showed that the formulation introduced in [29] is the most efficient in this

application due to its tightness, i.e., the small distance between optimal integer and relaxed solutions.

Then, a novel method to select representative days has been presented. Specifically, the proposed method is designed to capture correlations among production and load, as well as spatial correlations among different system zones.

The use of representative days to evaluate power system operational conditions on medium-term or long-term planning horizons raises the crucial issue regarding how to properly model the seasonality of hydroelectric dispatch. Two different approaches have been proposed to address this issue. First, we presented a QMIP model that determines the optimal daily dispatching of water resources over a medium-term or long-term planning horizon according to the zonal net load values and the linear relationship between energy price and net load. Then, we introduced some constraints that link representative days to create continuity in storage across the entire time horizon by means of the cluster index, i.e., the association of each day in the planning horizon to a representative day.

The use of representative days raises also the crucial issue regarding how to properly apply thermal unit commitment constraints without overestimating start-up costs. This problem has been addressed by introducing different techniques to assign to each thermal power plant an initial ON/OFF status in representative days. Experimental results showed that decision trees are the best classifier in the considered application, providing accurate estimates of initial statuses also in scenarios very different from the dataset used for the tree induction.

Finally, tests performed on both the Italian and the European scenarios demonstrated the goodness of the proposed method, which allows obtaining very accurate estimates of the operational costs, while dramatically reducing computational times.

Provided a comprehensive description of the proposed model to evaluate power systems short-term operations, the following step is represented by the integration of this analysis in the expansion planning framework. This research activity is described in the following chapter, which designs an optimization model to plan long-term investment decisions in the power sector. The distinct feature of such a model is the very detailed representation of the short-term operations, which is needed to accurately address all the challenges related to integrating high shares of intermittent renewable energy sources.

Chapter 3

Planning investments in the power sector to reach decarbonisation targets

3.1 Research motivation

The previous chapter introduced the optimization model developed to provide an accurate representation of power systems short-term operations, while keeping the problem computationally tractable. This chapter deals with the following research activity, consisting in the integration of such an analysis in the expansion planning framework, by formulating an optimization model to plan long-term investment decisions in the power sector to reach decarbonisation targets, which include an increasing penetration of renewable power sources and a reduction of CO_2 emissions. Such a model optimizes strategic decisions including retirement of existing capacity and investments in new generation, transmission and storage facilities, as well as operational decisions.

The structure of the chapter is as follows. Section 3.2 describes the assumptions we introduced in our analysis and formulates the GTEP problem as a MILP model. Results concerning the expansion planning of the Italian power system to reach by 2030 decarbonisation targets are presented in Section 3.3. Finally, Section 3.4 concludes the chapter.

3.2 Modeling framework

In this section, we present our approach to GTEP analysis by introducing a MILP model specifically designed to reach in the power sector decarbonisation targets set by the European Commission. The distinct feature of this model is a high level of both technical and temporal detail, which is required to properly evaluate all the challenges related to integrating high shares of renewables to reach decarbonisation targets. This section

presents the modeling assumptions as well as the notation and the mathematical formulation of the optimization model.

3.2.1 Modeling assumptions

In this chapter, we introduce an optimization model developed in collaboration with CESI S.p.A. to support government authorities in generation and transmission expansion planning. Such a model is based on a centralized approach: by assuming the perspective of a single central entity, we plan the joint expansion of generation and transmission facilities so as to minimize the total system costs. To support regulators in searching for optimal policies, our approach to GTEP analysis is focused on the inclusion of as many engineering details as possible, while neglecting interactions between the different agents involved in the liberalized power sector, as already mentioned in Section 1.2.2.

As regards to the electricity system modeling, in our analysis we consider the power system consisting of a set $\mathcal{Z}$ of zones, grouped into macro-areas, with the set of macro-areas denoted by $\mathcal{M}$. The structure of the power system at the beginning of the planning horizon is described by set $\mathcal{L}_E$ of transmission lines connecting zones, set $\mathcal{K}_E$ of thermal power plants, set $\mathcal{H}_E$ of hydropower plants, and parameters sol_{z_0} and $wind_{z_0}$ representing the solar power capacity and the wind power capacity, respectively, installed in each system zone $z \in \mathcal{Z}$. The decisions to be taken concern decommissioning of existing thermal power plants as well as building of new thermal units, new transmission lines, new hydropower plants, new wind and solar power capacity, and new batteries capacity, which are supposed to have a negligible installed capacity at the beginning of the planning horizon. Capacity decisions are defined for every year of the planning horizon, with the set of years denoted by $\mathcal{Y}$.

Decommissioning of existing thermal power plants are modeled trough a set of binary variables representing the decision to decommission a given power plant in a given year of the planning horizon. Set of existing thermal power plants is partitioned in two different subsets: while set $\mathcal{K}_{E1}$ includes power plants to be mandatorily decommissioned (e.g., coal plants), set $\mathcal{K}_{E2}$ groups existing thermal power plants that may be optionally decommissioned.

Investments in new thermal power plants are modeled by introducing set $\mathcal{K}_C$ of candidate facilities and binary variables representing the decision to build candidate power plants in a given year of the planning horizon. Candidate thermal power plants may present different investment priorities. Thus, set $\mathcal{K}_C$ is partitioned into subsets $\mathcal{K}_{C1}$

and $\mathcal{K}_{C2}$ to distinguish between candidate projects to be mandatorily or optionally constructed. Moreover, candidate projects include both the construction of new power plants and the upgrade of existing units to model, for instance, the possibility to convert some existing coal power plants into combined cycle gas turbine units.

Similarly to thermal generation capacity expansion, also investments in new transmission lines and new hydropower plants are modeled by introducing set of discrete facilities $\mathcal{L}_C$ and $\mathcal{H}_C$ and binary variables. Set $\mathcal{L}_C$ is partitioned in two subsets according to the mandatory or optional investment decisions on candidate transmission lines. Instead, all investments in new hydropower plants are modeled as optional decisions. This modeling choice is strictly related to the characteristics of the power systems analysed in our applications. However, the proposed model could be easily modified to include in the analysis also hydropower plants to be mandatorily constructed.

Moreover, earliest and latest dates are considered for the decommissioning of existing thermal power plants and the building of new thermal units, new transmission lines and new hydropower plants.

As opposed to thermal units, hydropower plants and transmission facilities, since individual components of renewable generation parks present small sizes, the investments in new wind and solar power generation are modeled by introducing continuous variables. Moreover, in the analysis we consider that the total solar and wind power capacities installed in each zone have to respect bounds set by the Energy Plan for every year of the planning horizon. Similarly to RES capacity expansion, also investments in batteries capacity are represented by continuous variables. According to the characteristics of the power systems analysed in our applications, no installed capacity for batteries is considered at the beginning of the planning horizon. However, the model proposed could be easily modified to include also for batteries an installed capacity at the beginning of the planning horizon, as it is for solar and wind power technology.

As regards to short-term operation modeling, we consider in every year y of the planning horizon a small set $\mathcal{C}^y$ of days that could be representative of the operational conditions in all days $\mathcal{D}^y$ of year y. Representative days for the first year of the planning horizon are determined by applying the procedure described in Section 2.3.2. Representative days for the following years are derived by applying annual growth factors to load profiles in the representative days of the first year.

As explained in the previous chapter, when working with representative days, two crucial issues have to be addressed: (i) capturing the seasonality of hydroelectric dispatch; and (ii) determining the initial ON/OFF status of thermal power plants in representative days. As regards to the long-term hydroelectric dispatch, we decided to

integrate into the expansion planning framework the approach described in Section 2.4.2, which is based on the application of inter-day balance constraints rather than executing a preliminary hydroelectric model. Such a choice is due to the following reasons:

- Representative days in the proposed formulation are already interconnected by some constraints (e.g., the yearly upper bounds on CO_2 emissions), thus keeping representative days temporally disconnected in modeling the long-term hydro operation does not provide a huge computational saving.
- In our applications, planning horizon of several decades are applied, thus future system conditions, especially at the end of the planning horizon, could be very different from those observed in the past.

Inter-day balance constraints are imposed by controlling every 7 days the energy level of the reservoirs and by considering the association of each day d to a representative day c, which is included in the cluster index $\mathrm{Map}_{d,c}$. Moreover, since all hydropower plants considered in the analysis are supposed to have monthly or yearly cycles, to reduce the computational burden the operation of hydropower plants is modeled by only considering inter-day constraints, while intra-day equations are not imposed. To ensure the respect of storage bounds in every hour of the planning horizon, the iterative procedure introduced in Section 2.4.2 is applied. Instead, batteries are supposed to have cycles lower than 24 hours, which implies the application of only intra-day balance constraints, but no inter-day constraints.

As regards to the determination of initial statuses, we applied the procedure summarized in Section 2.5.5, by assigning to each thermal power plant an initial status in representative days $\gamma_{k_0}^c$ according to the probability provided by a decision tree estimated on historical data. It is worth mentioning that for each thermal power plant, a different probability to be ON is computed for every year of the planning horizon. Indeed, since fuel costs and emission costs may change throughout the years of the planning horizon, the same thermal power plant could present different values for the marginal cost ratio in different years and thus different probabilities to be ON at the beginning of the day.

Given the long-term horizon, a transportation approach is introduced to model the operation of the transmission network, imposing only transmission limits on power flows, without including in the model voltage variables. Although transportation models do not provide a perfect representation of load flows, this choice is justified by the

computational burden of the expansion planning problem, related to the long-term scope, and by the zonal representation adopted to represent power systems.

Finally, in this first formulation of the investment planning problem, no long-term uncertainties are included in the analysis, while the demand for electricity is supposed to be inelastic. These two restrictions will be removed in the following chapters.

3.2.2 Notation

To formulate the GTEP problem, the following notation is introduced.

Sets

$\mathcal{Y}$	Set of years, indexed by y and i
$\mathcal{Z}$	Set of zones, indexed by z
$\mathcal{M}$	Set of macro-areas, indexed by m
$\mathcal{K}_E$	Set of existing thermal power plants
$\mathcal{K}_C$	Set of candidate thermal power plants
$\mathcal{K}$	Set of thermal power plants ($\mathcal{K} = \mathcal{K}_E \cup \mathcal{K}_C$), indexed by k and k'
$\Omega_z^k \subset \mathcal{K}$	Set of thermal power plants located in zone z
$\mathcal{K}_{E1} \subset \mathcal{K}_E$	Set of existing thermal power plants to be mandatorily decommissioned
$\mathcal{K}_{E2} \subset \mathcal{K}_E$	Set of existing thermal power plants that may be optionally decommissioned
$\mathcal{K}_{C1} \subset \mathcal{K}_C$	Set of candidate thermal power plants to be mandatorily constructed
$\mathcal{K}_{C2} \subset \mathcal{K}_C$	Set of candidate thermal power plants that may be optionally constructed
$\mathcal{AK}_j \subset \mathcal{K}_C$	j-th group of associate candidate thermal power plants
$J^{\mathcal{AK}}$	Set of groups of associate candidate thermal power plants
$\mathcal{MEK}_j \subset \mathcal{K}_C$	j-th group of mutually exclusive candidate thermal power plants
$J^{\mathcal{MEK}}$	Set of groups of mutually exclusive candidate thermal power plants
$\mathcal{L}_E$	Set of existing transmission lines
$\mathcal{L}_C$	Set of candidate transmission lines
$\mathcal{L}$	Set of transmission lines ($\mathcal{L} = \mathcal{L}_E \cup \mathcal{L}_C$), indexed by l and l'
$\mathcal{L}_{C1} \subset \mathcal{L}_C$	Set of candidate transmission lines to be mandatorily constructed
$\mathcal{L}_{C2} \subset \mathcal{L}_C$	Set of candidate transmission lines that may be optionally constructed
$\mathcal{AL}_j \subset \mathcal{L}_C$	j-th group of associate candidate transmission lines
$J^{\mathcal{AL}}$	Set of groups of associate candidate transmission lines

$\mathcal{MEL}_j \subset \mathcal{L}_C$	j-th group of mutually exclusive candidate transmission lines
$J^{\mathcal{MEL}}$	Set of groups of mutually exclusive candidate transmission lines
$\mathcal{H}_E$	Set of existing hydropower plants
$\mathcal{H}_C$	Set of candidate hydropower plants
$\mathcal{H}$	Set of hydropower plants ($\mathcal{H} = \mathcal{H}_E \cup \mathcal{H}_C$), indexed by h
$\Omega_z^h \subset \mathcal{H}$	Set of hydropower plants located in zone z
$\mathcal{B}$	Set of batteries, indexed by b
$\Omega_z^b \subset \mathcal{B}$	Set of batteries located in zone z
$\mathcal{F}$	Set of fuels, indexed by f
$\Phi_{z,f} \subset \Omega_z^k$	Set of thermal power plants located in zone z using fuel f
$\mathcal{C}^y$	Set of representative days of year y, indexed by c
$\mathcal{D}^y$	Set of all days of year y, indexed by d and d'
$\mathcal{D}_M^y \subset \mathcal{D}^y$	Set of days of year y in which the level of the long-term storage is checked
$\mathcal{T}$	Set of hours, from 1 to 24, indexed by t and τ
$ma(z)$	macro-area that contains zone z
$UP(k)$	upgrade project of existing thermal power plant k
$rz(l)$	Receiving-end zone of transmission line l
$sz(l)$	Sending-end zone of transmission line l
$fuel(k)$	Fuel used in thermal power plant k
$\mathrm{Map}_{d,c}$	Cluster index, i.e., injective map of each day d to a representative day c

Parameters

y_0	[−]	Reference year to which all investment costs are discounted
r	[−]	Annual discount rate
c_{ENP}	[€/MWh]	Penalty for energy not provided
c_{OG}	[€/MWh]	Penalty for over-generation
wg_c	[−]	Weight of representative day c
DC_k	[€]	Decommissioning cost of existing thermal power plant $k \in \mathcal{K}_E$
IC_k^{The}	[€]	Investment cost of candidate thermal power plant $k \in \mathcal{K}_C$
FC_k	[€]	Annual fixed costs of thermal power plant $k \in \mathcal{K}$
$CM_{k,y}$	[€/MWh]	Marginal production cost of thermal power plant k in year y

$\underline{\tau}_k$	[−]	Earliest date for construction/decommission of thermal power plant k
$\overline{\tau}_k$	[−]	Latest date for construction/decommission of thermal power plant k
$\underline{P}_k$	[MW]	Minimum power output of thermal power plant k
$\overline{P}_k$	[MW]	Maximum power produced by thermal plant k
SUC_k	[€]	Start-up cost of thermal power plant k
MUT_k	[h]	Minimum up time of thermal power plant k
MDT_k	[h]	Minimum down time of thermal power plant k
$\gamma^c_{k\,0}$	[−]	Initial ON/OFF status of thermal power plant k in representative day c
HR_k	[Gcal/MWh]	Heat rate of thermal power plants of cluster k
OM_k	[€/MWh]	Operative and maintenance cost of thermal power plants of cluster k
$IC^{\mathrm{Sol}}_{z,y}$	[€/MW]	Investment cost of new solar power capacity in zone z in year y
$sol_{z,0}$	[MW]	Solar power capacity installed in zone z at the beginning of the planning horizon
$\underline{PV}_{z,y}$	[MW]	Lower bound for solar power capacity in zone z in year y
$\overline{PV}_{z,y}$	[MW]	Upper bound for solar power capacity in zone z in year y
$IC^{\mathrm{Wind}}_{z,y}$	[€/MW]	Investment cost of new wind power capacity in zone z in year y
$wind_{z,0}$	[MW]	Wind power capacity installed in zone z at the beginning of the planning horizon
$\underline{W}_{z,y}$	[MW]	Lower bound for wind power capacity in zone z in year y
$\overline{W}_{z,y}$	[MW]	Upper bound for wind power capacity in zone z in year y
$D^c_{z,t}$	[MW]	Load in zone z in hour t of representative day c
$R^c_{z,t}$	[MW]	Reserve requirement for zone z in hour t of representative day c
$\mu^c_{z,t}$	[MWh/MW]	Solar power capacity factor for zone z in hour t of representative day c
$\rho^c_{z,t}$	[MWh/MW]	Wind power capacity factor for zone z in hour t of representative day c
$IC^{\mathrm{Hyd}}_{h,y}$	[€]	Investment cost of candidate hydropower plant $h \in \mathcal{H}_C$ in year y

$\underline{\tau}_h$	[−]	Earliest date for construction of hydropower plant h
$\overline{\tau}_h$	[−]	Latest date for construction of hydropower plant h
$Cvar_h$	[€/MWh]	Operating cost of hydropower plant h
$\bar{E}_h^{\mathrm{IN}}$	[MW]	Upper bound on hydropower plant h pumping power
$\bar{E}_h^{\mathrm{OUT}}$	[MW]	Upper bound on hydropower plant h power output
$\overline{sl}_h$	[MWh]	Upper bound on energy spillage from hydropower plant h
$F_{h,t}^c$	[MWh]	Hourly energy inflows for hydropower plant h at time t of representative day c
λ_h^{IN}	[−]	Loss coefficient for hydro plant h pumping $\left(0 \le \lambda_h^{\mathrm{IN}} \le 1\right)$
λ_h^{OUT}	[−]	Loss coefficient for hydro plant h power generation $\left(\lambda_h^{\mathrm{OUT}} \ge 1\right)$
$E_{h\,0}$	[MWh]	Energy content of hydropower plant h at the beginning of planning horizon
EPR_h	[h]	Maximum energy to power ratio (in hours) for hydropower plant h
M	[−]	Size of the temporal window in long-term storage constraints, set to 7 days
$IC_{b,y}^{\mathrm{Batt}}$	[€/MW]	Investment cost for battery b in year y
$Cvar_b$	[€/MWh]	Operating cost of battery b
$\overline{CAP}_b^{\mathrm{Batt}}$	[MW]	Upper bound on battery b installed power
$E_{b\,0}^c$	[MWh]	Initial energy content of battery b in representative day c
λ_b	[−]	Loss coefficient for energy stored by battery b $(0 \le \lambda_b \le 1)$
λ_b^{IN}	[−]	Loss coefficient for battery b charge $\left(0 \le \lambda_b^{\mathrm{IN}} \le 1\right)$
λ_b^{OUT}	[−]	Loss coefficient for battery b discharge $\left(\lambda_b^{\mathrm{OUT}} \ge 1\right)$
$\bar{E}_b^{\mathrm{IN}}$	[MW]	Upper bound on battery b charge
$\bar{E}_b^{\mathrm{OUT}}$	[MW]	Upper bound on battery b discharge
EPR_b	[h]	Maximum energy to power ratio (in hours) for battery b
IC_l^{Line}	[€]	Investment cost of candidate transmission line $l \subset \mathcal{L}_C$
$\underline{\tau}_l$	[−]	Earliest date for construction of candidate transmission line l
$\overline{\tau}_l$	[−]	Latest date for construction of candidate transmission line l
$\underline{F}_l$	[MW]	Minimum capacity of transmission line l
$\overline{F}_l$	[MW]	Maximum capacity of transmission line l

$ECnt_f$	[Gcal/M.U[1]]	Energy content of fuel f
co_{2_f}	[ton/Gcal]	CO_2 emission factor of fuel f
$\overline{FA}_{f,m,y}$	[M.U]	Upper bound on availability of fuel f in macro-area m in year y
Pr_y^f	[€/Gcal]	Price of fuel f in year y
$\overline{CO_2}_{m,y}$	[ton]	CO_2 emission limit for macro-area m in year y
$\varphi_{m,y}$	[−]	Lower bound for renewables penetration in macro-area m in year y
Pr_y^{CO2}	[€/ton]	Emission price in year y

Variables

$\delta_{k,y}^-$	[−]	1: thermal power plant $k \in \mathcal{K}_E$ is decommissioned in year y; 0: otherwise
$\delta_{k,y}^+$	[−]	1: thermal power plant $k \in \mathcal{K}_C$ is built in year y; 0: otherwise
$\delta_{h,y}$	[−]	1: hydro power plant $h \in \mathcal{H}_C$ is built in year y; 0: otherwise
$\delta_{l,y}$	[−]	1: transmission line $l \in \mathcal{L}_C$ is built in year y; 0: otherwise
$\theta_{k,y}^-$	[−]	1: thermal power plant $k \in \mathcal{K}_E$ is decommissioned within year y; 0: otherwise
$\theta_{k,y}^+$	[−]	1: thermal power plant $k \in \mathcal{K}_C$ is built within year y; 0: otherwise
$\theta_{h,y}$	[−]	1: hydro power plant $h \in \mathcal{H}_C$ is built within year y; 0: otherwise
$\theta_{l,y}$	[−]	1: transmission line $l \in \mathcal{L}_C$ is built within year y; 0: otherwise
$sol_{z,y}$	[MW]	New solar capacity installed in zone z in year y
$wind_{z,y}$	[MW]	New wind capacity installed in zone z in year y
$cap_{b,y}^{Batt}$	[MW]	Storage capacity of battery b installed in year y
$\gamma_{k,t}^c$	[−]	1: thermal power plant k is ON in hour t of representative day c; 0: otherwise
$\alpha_{k,t}^c$	[−]	1: thermal power plant k is started-up in hour t of representative day c; 0: otherwise
$\beta_{k,t}^c$	[−]	1: thermal power plant k is shut-down in hour t of representative day c; 0: otherwise
$p_{k,t}^c$	[MW]	Power production of thermal power plant k at time t of representative day c above its minimum output $\underline{P}_k$

[1] By M.U we mean the specific measure unit of fuel f, i.e., 10^6 m^3 for natural gas and ktons for coal and oil

$E_{h,t}^{\mathrm{IN},c}$	[MW]	Pumping power of hydro reservoir h in hour t of representative day c
$E_{h,t}^{\mathrm{OUT},c}$	[MW]	Power output of hydro reservoir h in hour t of representative day c
$sl_{h,t}^{c}$	[MWh]	Energy spillage from hydro reservoir h in hour t of representative day c
$\hat{E}_{h}^{\mathrm{LT},d}$	[MWh]	Energy level of hydro reservoir h at the end of day d
$E_{b,t}^{\mathrm{IN},c}$	[MW]	Charge of battery b in hour t of representative day c
$E_{b,t}^{\mathrm{OUT},c}$	[MW]	Discharge of battery b in hour t of representative day c
$E_{b,t}^{c}$	[MWh]	Energy level of battery b in hour t of representative day c
$x_{l,t}^{c}$	[MW]	Power flow on line l in hour t of representative day c
$ENP_{z,t}^{c}$	[MWh]	Energy not provided in zone z in hour t of representative day c
$OG_{z,t}^{c}$	[MWh]	Over-generation in zone z in hour t of representative day c
$RES_{z,t}^{c}$	[MWh]	Renewable generation in zone z in hour t of representative day c

3.2.3 Mathematical formulation

The GTEP model can be formulated as the following MILP model

$$
\begin{aligned}
\min\ z \;=\;& \sum_{y\in\mathcal{Y}}\left(\sum_{k\in\mathcal{K}_E}\frac{DC_k\,\delta_{k,y}^{-}}{(1+r)^{y-y_0}}+\sum_{k\in\mathcal{K}_C}\frac{IC_k^{\mathrm{The}}\,\delta_{k,y}^{+}}{(1+r)^{y-y_0}}\right)+\\
&+\sum_{y\in\mathcal{Y}}\left(\sum_{z\in\mathcal{Z}}\frac{IC_{z,y}^{\mathrm{Sol}}\,sol_{z,y}}{(1+r)^{y-y_0}}+\sum_{z\in\mathcal{Z}}\frac{IC_{z,y}^{\mathrm{Wind}}\,wind_{z,y}}{(1+r)^{y-y_0}}\right)+\\
&+\sum_{y\in\mathcal{Y}}\sum_{h\in\mathcal{H}_C}\frac{IC_{h,y}^{\mathrm{Hyd}}\,\delta_{h,y}}{(1+r)^{y-y_0}}+\\
&+\sum_{y\in\mathcal{Y}}\sum_{b\in\mathcal{B}}\frac{IC_{b,y}^{\mathrm{Batt}}\,cap_{b,y}^{\mathrm{Batt}}}{(1+r)^{y-y_0}}+\\
&+\sum_{y\in\mathcal{Y}}\sum_{l\in\mathcal{L}_C}\frac{IC_{l}^{\mathrm{Line}}\,\delta_{l,y}}{(1+r)^{y-y_0}}+\\
&+\sum_{y\in\mathcal{Y}}\left[\sum_{k\in\mathcal{K}_E}FC_k\left(1-\theta_{k,y}^{-}\right)+\sum_{k\in\mathcal{K}_C}FC_k\,\theta_{k,y}^{+}\right]+\\
&+\sum_{y\in\mathcal{Y}}\sum_{c\in\mathcal{C}^{\mathcal{Y}}}wg_c\sum_{t=1}^{24}\left[\sum_{k\in\mathcal{K}}\left(CM_{k,y}\left(\underline{P}_k\,\gamma_{k,t}^{c}+p_{k,t}^{c}\right)+SUC_k\alpha_{k,t}^{c}\right)\right.
\end{aligned}
$$

$$+\sum_{h\in\mathcal{H}} Cvar_h\, E_{h,t}^{\mathrm{OUT},c} + \sum_{b\in\mathcal{B}} Cvar_b\, E_{b,t}^{\mathrm{OUT},c}$$

$$+ c_{ENP}\sum_{z\in Z} ENP_{z,t}^{c} + c_{OG}\sum_{z\in Z} OG_{z,t}^{c}\Bigg] \tag{3.1}$$

subject to

$$\delta_{k,y}^{-} = 0 \qquad k\in\mathcal{K}_E, y\notin\left[\underline{\tau}_k,\overline{\tau}_k\right] \tag{3.2}$$

$$\sum_{y\in\mathcal{Y}}\delta_{k,y}^{-} = 1 \qquad k\in\mathcal{K}_{E1} \tag{3.3}$$

$$\sum_{y\in\mathcal{Y}}\delta_{k,y}^{-} \le 1 \qquad k\in\mathcal{K}_{E2} \tag{3.4}$$

$$\delta_{k,y}^{+} = 0 \qquad k\in\mathcal{K}_C, y\notin\left[\underline{\tau}_k,\overline{\tau}_k\right] \tag{3.5}$$

$$\sum_{y\in\mathcal{Y}}\delta_{k,y}^{+} = 1 \qquad k\in\mathcal{K}_{C1} \tag{3.6}$$

$$\sum_{y\in\mathcal{Y}}\delta_{k,y}^{+} \le 1 \qquad k\in\mathcal{K}_{C2} \tag{3.7}$$

$$\sum_{y\in\mathcal{Y}}\delta_{k,y}^{+} = \sum_{y\in\mathcal{Y}}\delta_{k',y}^{+} \qquad k\in\mathcal{AK}_j, k'\in\mathcal{AK}_j, j\in J^{\mathcal{AK}} \tag{3.8}$$

$$\sum_{k\in\mathcal{MEK}_j}\sum_{y\in\mathcal{Y}}\delta_{k,y}^{+} \le 1 \qquad j\in J^{\mathcal{MEK}} \tag{3.9}$$

$$\underline{PV}_{z,y} \le sol_{z,0} + \sum_{i=1}^{y} sol_{z,i} \le \overline{PV}_{z,y} \qquad z\in\mathcal{Z}, y\in\mathcal{Y} \tag{3.10}$$

$$\underline{W}_{z,y} \le wind_{z,0} + \sum_{i=1}^{y} wind_{z,i} \le \overline{W}_{z,y} \qquad z\in\mathcal{Z}, y\in\mathcal{Y} \tag{3.11}$$

$$\delta_{h,y} = 0 \qquad h\in\mathcal{H}_C, y\notin\left[\underline{\tau}_h,\overline{\tau}_h\right] \tag{3.12}$$

$$\sum_{y\in\mathcal{Y}}\delta_{h,y} \le 1 \qquad h\in\mathcal{H}_C \tag{3.13}$$

$$\sum_{y\in\mathcal{Y}} cap_{b,y}^{\mathrm{Batt}} \le \overline{CAP}_b^{\mathrm{Batt}} \qquad b\in\mathcal{B} \tag{3.14}$$

$$\delta_{l,y} = 0 \qquad l\in\mathcal{L}_C, y\notin\left[\underline{\tau}_l,\overline{\tau}_l\right] \tag{3.15}$$

$$\sum_{y\in\mathcal{Y}}\delta_{l,y} = 1 \qquad l\in\mathcal{L}_{C1} \tag{3.16}$$

$$\sum_{y\in\mathcal{Y}}\delta_{l,y} \le 1 \qquad l\in\mathcal{L}_{C2} \tag{3.17}$$

$$\sum_{y\in\mathcal{Y}}\delta_{l,y}=\sum_{y\in\mathcal{Y}}\delta_{l\prime,y} \qquad l\in\mathcal{AL}_j, l'\in\mathcal{AL}_j, j\in J^{\mathcal{AL}} \qquad (3.18)$$

$$\sum_{l\in\mathcal{MEL}_j}\sum_{y\in\mathcal{Y}}\delta_{l,y}\leq 1 \qquad j\in J^{\mathcal{MEL}} \qquad (3.19)$$

$$\theta^{-}_{k,y}=\sum_{i=1}^{y}\delta^{-}_{k,i} \qquad k\in\mathcal{K}_E, y\in\mathcal{Y} \qquad (3.20)$$

$$\theta^{+}_{k,y}=\sum_{i=1}^{y}\delta^{+}_{k,i} \qquad k\in\mathcal{K}_C, y\in\mathcal{Y} \qquad (3.21)$$

$$\theta_{h,y}=\sum_{i=1}^{y}\delta_{h,i} \qquad h\in\mathcal{H}_C, y\in\mathcal{Y} \qquad (3.22)$$

$$\theta_{l,y}=\sum_{i=1}^{y}\delta_{l,i} \qquad l\in\mathcal{L}_C, y\in\mathcal{Y} \qquad (3.23)$$

$$\gamma^{c}_{k,t}\leq 1-\theta^{-}_{k,y} \qquad k\in\mathcal{K}_E, 1\leq t\leq 24, c\in\mathcal{C}^{y}, y\in\mathcal{Y} \qquad (3.24)$$

$$\gamma^{c}_{k,t}\leq \theta^{+}_{k,y} \qquad k\in\mathcal{K}_C, 1\leq t\leq 24, c\in\mathcal{C}^{y}, y\in\mathcal{Y} \qquad (3.25)$$

$$\gamma_{k,t}\leq 1-\theta^{+}_{k\prime,y} \qquad k\in\mathcal{K}_E, k'=UP(k), 1\leq t\leq 24, c\in\mathcal{C}^{y}, y\in\mathcal{Y} \qquad (3.26)$$

$$p^{c}_{k,t}\leq\gamma^{c}_{k,t}\left(\overline{P}_k-\underline{P}_k\right) \qquad k\in\mathcal{K}, 1\leq t\leq 24, c\in\mathcal{C}^{y}, y\in\mathcal{Y} \qquad (3.27)$$

$$\gamma^{c}_{k,t}-\gamma^{c}_{k_0}=\alpha^{c}_{k,t}-\beta^{c}_{k,t} \qquad k\in\mathcal{K}, t=1, c\in\mathcal{C}^{y}, y\in\mathcal{Y} \qquad (3.28)$$

$$\gamma^{c}_{k,t}-\gamma^{c}_{k,t-1}=\alpha^{c}_{k,t}-\beta^{c}_{k,t} \qquad k\in\mathcal{K}, 2\leq t\leq 24, c\in\mathcal{C}^{y}, y\in\mathcal{Y} \qquad (3.29)$$

$$\sum_{\tau=t-MUT_k+1}^{t}\alpha^{c}_{k,\tau}\leq\gamma^{c}_{k,t} \qquad k\in\mathcal{K}, MUT_k\leq t\leq 24, c\in\mathcal{C}^{y}, y\in\mathcal{Y} \qquad (3.30)$$

$$\sum_{\tau=t-MDT_k+1}^{t}\beta^{c}_{k,\tau}\leq 1-\gamma^{c}_{k,t} \qquad k\in\mathcal{K}, MDT_k\leq t\leq 24, c\in\mathcal{C}^{y}, y\in\mathcal{Y} \qquad (3.31)$$

$$E^{\mathrm{IN},c}_{h,t}\leq\bar{E}^{\mathrm{IN}}_{h} \qquad h\in\mathcal{H}_E, 1\leq t\leq 24, c\in\mathcal{C}^{y}, y\in\mathcal{Y} \qquad (3.32)$$

$$E^{\mathrm{OUT},c}_{h,t}\leq\bar{E}^{\mathrm{OUT}}_{h} \qquad h\in\mathcal{H}_E, 1\leq t\leq 24, c\in\mathcal{C}^{y}, y\in\mathcal{Y} \qquad (3.33)$$

$$E^{\mathrm{IN},c}_{h,t}\leq\bar{E}^{\mathrm{IN}}_{h}\theta_{h,y} \qquad h\in\mathcal{H}_C, 1\leq t\leq 24, c\in\mathcal{C}^{y}, y\in\mathcal{Y} \qquad (3.34)$$

$$E^{\mathrm{OUT},c}_{h,t}\leq\bar{E}^{\mathrm{OUT}}_{h}\theta_{h,y} \qquad h\in\mathcal{H}_C, 1\leq t\leq 24, c\in\mathcal{C}^{y}, y\in\mathcal{Y} \qquad (3.35)$$

$$\frac{E^{\mathrm{IN},c}_{h,t}}{\bar{E}^{\mathrm{IN}}_{h}}+\frac{E^{\mathrm{OUT},c}_{h,t}}{\bar{E}^{\mathrm{OUT}}_{h}}\leq 1 \qquad h\in\mathcal{H}, 1\leq t\leq 24, c\in\mathcal{C}^{y}, y\in\mathcal{Y} \qquad (3.36)$$

$$sl_{h,t}^{c} \leq \bar{sl}_{h} \qquad h \in \mathcal{H}, 1 \leq t \leq 24, c \in \mathcal{C}^{y}, y \in \mathcal{Y} \qquad (3.37)$$

$$\hat{E}_{h}^{\mathrm{LT},d} = E_{h_0} + \sum_{d'=d-M+1}^{d} \sum_{c \in \mathrm{Map}_{d',c}} \sum_{t=1}^{24} \left(F_{h,t}^{c} + \lambda_{h}^{\mathrm{IN}} E_{h,t}^{c,\mathrm{IN}} - \lambda_{h}^{\mathrm{OUT}} E_{h,t}^{c,\mathrm{OUT}} - sl_{h,t}^{c} \right)$$
$$h \in \mathcal{H}, d = M \qquad (3.38)$$

$$\hat{E}_{h}^{\mathrm{LT},d} = \hat{E}_{h}^{\mathrm{LT},d-M} + \sum_{d'=d-M+1}^{d} \sum_{c \in \mathrm{Map}_{d',c}} \sum_{t=1}^{24} \left(F_{h,t}^{c} + \lambda_{h}^{\mathrm{IN}} E_{h,t}^{c,\mathrm{IN}} - \lambda_{h}^{\mathrm{OUT}} E_{h,t}^{c,\mathrm{OUT}} - sl_{h,t}^{c} \right)$$
$$h \in \mathcal{H}, d \in \mathcal{D}_{M}^{y}, y \in \mathcal{Y}, d > M \qquad (3.39)$$

$$\hat{E}_{h}^{\mathrm{LT},d} \leq EPR_{h} \bar{E}_{h}^{\mathrm{IN}} \qquad h \in \mathcal{H}, d \in \mathcal{D}_{M}^{y}, y \in \mathcal{Y} \qquad (3.40)$$

$$\hat{E}_{h}^{\mathrm{LT},d} = E_{h_0} \qquad h \in \mathcal{H}, d = |\mathcal{D}^{y}|, y \in \mathcal{Y} \qquad (3.41)$$

$$E_{b,t}^{\mathrm{IN},c} \leq \sum_{i=1}^{y} cap_{b,i}^{\mathrm{Batt}} \qquad h \in \mathcal{B}, 1 \leq t \leq 24, c \in \mathcal{C}^{y}, y \in \mathcal{Y} \qquad (3.42)$$

$$E_{b,t}^{\mathrm{OUT},c} \leq \sum_{i=1}^{y} cap_{b,i}^{\mathrm{Batt}} \qquad h \in \mathcal{B}, 1 \leq t \leq 24, c \in \mathcal{C}^{y}, y \in \mathcal{Y} \qquad (3.43)$$

$$E_{b,t}^{c} \leq EPR_{b} \sum_{i=1}^{y} cap_{b,i}^{\mathrm{Batt}} \qquad b \in \mathcal{B}, 1 \leq t \leq 24, c \in \mathcal{C}^{y}, y \in \mathcal{Y} \qquad (3.44)$$

$$E_{b,t}^{c} = (1 - \lambda_{b}) E_{b,t-1}^{c} + \lambda_{b}^{\mathrm{IN}} E_{b,t}^{\mathrm{IN},c} - \lambda_{b}^{\mathrm{OUT}} E_{b,t}^{\mathrm{OUT},c}$$
$$b \in \mathcal{B}, 2 \leq t \leq 24, c \in \mathcal{C}^{y}, y \in \mathcal{Y} \qquad (3.45)$$

$$E_{b,t}^{c} = (1 - \lambda_{b}) E_{b_0}^{c} + \lambda_{b}^{\mathrm{IN}} E_{b,t}^{\mathrm{IN},c} - \lambda_{b}^{\mathrm{OUT}} E_{b,t}^{\mathrm{OUT},c}$$
$$b \in \mathcal{B}, t = 1, c \in \mathcal{C}^{y}, y \in \mathcal{Y} \qquad (3.46)$$

$$E_{b,t}^{c} = E_{b_0}^{c} \qquad b \in \mathcal{B}, t = 24, c \in \mathcal{C}^{y}, y \in \mathcal{Y} \qquad (3.47)$$

$$\underline{F_l} \le x_{l,t}^c \le \overline{F}_l \qquad l \in \mathcal{L}_E, 1 \le t \le 24, c \in \mathcal{C}^y, y \in \mathcal{Y} \qquad (3.48)$$

$$\theta_{l,y}\,\underline{F_l} \le x_{l,t}^c \le \theta_{l,y}\,\overline{F}_l \qquad l \in \mathcal{L}_C, 1 \le t \le 24, c \in \mathcal{C}^y, y \in \mathcal{Y} \qquad (3.49)$$

$$\sum_{k\epsilon\Omega_z^k}\left(\underline{P_k}\,\gamma_{k,t}^c + p_{k,t}^c\right) + \mu_{z,t}^c\left(sol_{z,0} + \sum_{i=1}^{y} sol_{z,i}\right) + \rho_{z,t}^c\left(wind_{z,0} + \sum_{i=1}^{y} wind_{z,i}\right)$$

$$+ \sum_{l|rz(l)=z} x_{l,t}^c + \sum_{h\epsilon\Omega_z^h} E_{h,t}^{\mathrm{OUT},c} + \sum_{b\epsilon\Omega_z^b} E_{b,t}^{\mathrm{OUT},c} + ENP_{z,t}^c =$$

$$= D_{z,t}^c + \sum_{l|sz(l)=z} x_{l,t}^c + \sum_{h\epsilon\Omega_z^h} E_{h,t}^{\mathrm{IN},c} + \sum_{b\epsilon\Omega_z^b} E_{b,t}^{\mathrm{IN},c} + OG_{z,t}^c$$

$$z \in \mathcal{Z}, 1 \le t \le 24, c \in \mathcal{C}^y, y \in \mathcal{Y} \qquad (3.50)$$

$$\sum_{k\epsilon\Omega_z}\left[\left(\overline{P}_k - \underline{P_k}\right)\gamma_{k,t}^c - p_{k,t}^c\right] \ge R_{z,t}^c \qquad z \in \mathcal{Z}, 1 \le t \le 24, c \in \mathcal{C}^y, y \in \mathcal{Y} \qquad (3.51)$$

$$\sum_{z|ma(z)=m}\ \sum_{k\in\Phi_{z,f}}\ \sum_{c\in\mathcal{C}^y} wg_c \sum_{t=1}^{24} \frac{HR_k(\underline{P_k}\,\gamma_{k,t}^c + p_{k,t}^c)}{ECnt_f} \le \overline{FA}_{f,m,y}$$

$$f \in \mathcal{F}, m \in \mathcal{M}, y \in \mathcal{Y} \qquad (3.52)$$

$$\sum_{z|ma(z)=m}\ \sum_{f\in\mathcal{F}}\ \sum_{k\in\Phi_{z,f}}\ \sum_{c\in\mathcal{C}^y} wg_c \sum_{t=1}^{24} HR_k(\underline{P_k}\,\gamma_{k,t}^c + p_{k,t}^c)\,co_{2_f} \le \overline{CO}_{2_{m,y}}$$

$$m \in \mathcal{M}, y \in \mathcal{Y} \qquad (3.53)$$

$$RES_{z,t}^c = \mu_{z,t}^c\left(sol_{z,0} + \sum_{i=1}^{y} sol_{z,i}\right) + \rho_{z,t}^c\left(wind_{z,0} + \sum_{i=1}^{y} wind_{z,i}\right) + \sum_{h\epsilon\Omega_z^h} E_{h,t}^{\mathrm{OUT},c}$$

$$z \in \mathcal{Z}, 1 \le t \le 24, c \in \mathcal{C}^y, y \in \mathcal{Y} \qquad (3.54)$$

$$\sum_{z|ma(z)=m}\ \sum_{c\in\mathcal{C}^y} wg_c \sum_{t=1}^{24} RES_{z,t}^c \ge \varphi_{m,y}\left(\sum_{z|ma(z)=m}\ \sum_{c\in\mathcal{C}^y} wg_c \sum_{t=1}^{24} D_{z,t}^c\right)$$

$$m \in \mathcal{M}, y \in \mathcal{Y} \qquad (3.55)$$

$$\delta_{k,y}^-, \theta_{k,y}^- \in \{0,1\} \qquad k \in \mathcal{K}_E, y \in \mathcal{Y} \qquad (3.56)$$

$$\delta_{k,y}^+, \theta_{k,y}^+ \in \{0,1\} \qquad k \in \mathcal{K}_C, y \in \mathcal{Y} \qquad (3.57)$$

$$\delta_{h,y}, \theta_{h,y} \in \{0,1\} \qquad h \in \mathcal{H}_C, y \in \mathcal{Y} \qquad (3.58)$$

$$\delta_{l,y}, \theta_{l,y} \in \{0,1\} \qquad l \in \mathcal{L}_C, y \in \mathcal{Y} \qquad (3.59)$$

$$sol_{z,y}, wind_{z,y} \geq 0 \qquad z \in \mathcal{Z}, y \in \mathcal{Y} \qquad (3.60)$$

$$cap_{b,y}^{\text{Batt}} \geq 0 \qquad b \in \mathcal{B}, y \in \mathcal{Y} \qquad (3.61)$$

$$\gamma_{k,t}^{c}, \alpha_{k,t}^{c}, \beta_{k,t}^{c} \in \{0,1\} \qquad k \in \mathcal{K}, 1 \leq t \leq 24, c \in \mathcal{C}^{y}, y \in \mathcal{Y} \qquad (3.62)$$

$$p_{k,t}^{c} \geq 0 \qquad k \in \mathcal{K}, 1 \leq t \leq 24, c \in \mathcal{C}^{y}, y \in \mathcal{Y} \qquad (3.63)$$

$$E_{h,t}^{\text{IN},c}, E_{h,t}^{\text{OUT},c}, sl_{h,t}^{c} \geq 0 \qquad h \in \mathcal{H}, 1 \leq t \leq 24, c \in \mathcal{C}^{y}, y \in \mathcal{Y} \qquad (3.64)$$

$$\hat{E}_{h}^{\text{LT},d} \geq 0 \qquad h \in \mathcal{H}, d \in \mathcal{D}_{M}^{y}, y \in \mathcal{Y} \qquad (3.65)$$

$$E_{b,t}^{\text{IN},c}, E_{b,t}^{\text{OUT},c}, E_{b,t}^{c} \geq 0 \qquad b \in \mathcal{B}, 1 \leq t \leq 24, c \in \mathcal{C}^{y}, y \in \mathcal{Y} \qquad (3.66)$$

$$x_{l,t}^{c} \text{ free variable} \qquad l \in \mathcal{L}, 1 \leq t \leq 24, c \in \mathcal{C}^{y}, y \in \mathcal{Y} \qquad (3.67)$$

$$ENP_{z,t}^{c}, OG_{z,t}^{c}, RES_{z,t}^{c} \geq 0 \qquad z \in \mathcal{Z}, 1 \leq t \leq 24, c \in \mathcal{C}^{y}, y \in \mathcal{Y}. \qquad (3.68)$$

The objective function (3.1) comprises the seven terms below:

1. $\sum_{y \in \mathcal{Y}} \left(\sum_{k \in \mathcal{K}_E} \frac{DC_k \, \delta_{k,y}^{-}}{(1+r)^{y-y_0}} + \sum_{k \in \mathcal{K}_C} \frac{IC_k^{\text{The}} \, \delta_{k,y}^{+}}{(1+r)^{y-y_0}} \right)$ are the annualized decommissioning costs of existing thermal power plants and investment costs in new thermal power generation;
2. $\sum_{y \in \mathcal{Y}} \left(\sum_{z \in \mathcal{Z}} \frac{IC_{z,y}^{\text{Sol}} \, sol_{z,y}}{(1+r)^{y-y_0}} + \sum_{z \in \mathcal{Z}} \frac{IC_{z,y}^{\text{Wind}} \, wind_{z,y}}{(1+r)^{y-y_0}} \right)$ are the annualized investment costs in new solar and wind capacity;
3. $\sum_{y \in \mathcal{Y}} \sum_{h \in \mathcal{H}_C} \frac{IC_{h,y}^{\text{Hyd}} \, \delta_{h,y}}{(1+r)^{y-y_0}}$ are the annualized investment costs in new hydropower plants;
4. $\sum_{y \in \mathcal{Y}} \sum_{b \in \mathcal{B}} \frac{IC_{b,y}^{\text{Batt}} \, cap_{b,y}^{Batt}}{(1+r)^{y-y_0}}$ are the annualized investment costs in new batteries capacity;
5. $\sum_{y \in \mathcal{Y}} \sum_{l \in \mathcal{L}_C} \frac{IC_{l}^{\text{Line}} \, \delta_{l,y}}{(1+r)^{y-y_0}}$ are the annualized investment costs in new transmission lines;

6. $\sum_{y\in\mathcal{Y}}\left[\sum_{k\in\mathcal{K}_E} FC_k\left(1-\theta^{-}_{k,y}\right)+\sum_{k\in\mathcal{K}_C} FC_k\,\theta^{+}_{k,y}\right]$ are the fixed costs for the available thermal power plants, i.e., not decommissioned existing plants and already constructed candidate plants;
7. $\sum_{y\in\mathcal{Y}}\sum_{c\in\mathcal{C}^y} wg_c\sum_{t=1}^{24}\left[\sum_{k\in\mathcal{K}}\left(CM_{k,y}\left(\underline{P}_k\,\gamma^c_{k,t}+p^c_{k,t}\right)+SUC_k\alpha^c_{k,t}\right)+\sum_{h\in\mathcal{H}} Cvar_h\,E^{\mathrm{OUT},c}_{h,t}+\sum_{b\in\mathcal{B}} Cvar_b\,E^{\mathrm{OUT},c}_{b,t}+c_{ENP}\sum_{z\in Z} ENP^c_{z,t}+c_{OG}\sum_{z\in Z} OG^c_{z,t}\right]$ are the operational costs.

Specifically, item 7 above considers for each representative day the sum of production costs, start-up costs, hydro and batteries operational costs and penalties for energy not provided and over-generation. Production costs are supposed to be linear functions of the power output, being $CM_{k,y}$ the slopes of these linear relationships. The marginal cost of thermal plant k in year y is computed as:

$$CM_{k,y}=O\&M_k+HR_k\left(Pr^f_y+co_{2_f}Pr^{\mathrm{CO2}}_y\right)\qquad k\in\mathcal{K},y\in\mathcal{Y},f=fuel(k)\qquad(3.69)$$

with $O\&M_k$ [€/MWh] being the operative and maintenance costs of plant k, HR_k [Gcal/MWh] the heat rate of thermal power plant k, Pr^f_y [€/Gcal] the price in year y of fuel f used by unit k, co_{2_f} [ton/Gcal] the CO_2 emission factor of fuel f used by unit k and Pr^{CO2}_y [€/ton] the emission cost in year y.

In the proposed model there are three groups of constraints, namely investment constraints (3.2)–(3.23), operational constraints (3.24)–(3.51) and target constraints (3.52)–(3.55). Investment constraints model investment decisions, considering the project priorities and the existence of logical relations between some investment decisions. In particular, assignment constraints (3.2) impose earliest and latest dates for decommissioning of existing thermal power plants, equalities (3.3) enforce the decommissioning of mandatory thermal units and constraints (3.4) model decisions regarding optional decommissioning.

The group of constraints (3.5)–(3.7) work similarly with investment decisions on candidate thermal power plants. Indeed, while equations (3.5) impose earliest and latest dates for construction of new thermal power plants, constraints (3.6) and (3.7) model investment decisions on mandatory and optional projects, respectively. The associate project constraints (3.8) indicate that a group of projects is subject to a single investment decision, that is, either all projects in the group $\mathcal{AK}_j$ are built, or none. Inequalities (3.9) are the mutually exclusive project's constraints and they ensure that only one unit (or none) of the projects in each group ($\mathcal{MEK}_j$) is built. Inequalities (3.10) and (3.11) impose

lower and upper bounds on solar and wind installed capacity, respectively. Equations (3.12) impose a temporal window for construction of new hydropower plants, while constraints (3.13) model investments in new hydropower plants as optional decisions. Inequalities (3.14) impose an upper bound on batteries installed capacity at the end of the planning horizon. The group of inequalities (3.15)–(3.19) model investment decisions on new transmission lines. Specifically, equations (3.15) impose a temporal window for the introduction of a new transmission line, while constraints (3.16) and (3.17) model, respectively, mandatory and optional decisions. Equations (3.18) are the associate lines constraints, indicating that a group of lines is subject to a single investment decision, while the mutually exclusive lines constraints (3.19) ensure that only one interconnection (or none) of the transmission lines in a given group ($\mathcal{MEL}_j$) is built. Finally, equations (3.20)–(3.23) determine the values of the binary variables $\theta^-_{k,y}$, $\theta^+_{k,y}$, $\theta_{h,y}$ and $\theta_{l,y}$ that express if decommissioning decisions for existing thermal power plants and investment decisions for new thermal power plants, new hydropower plants, and new transmission lines, respectively, have been made within every year y of the planning horizon.

Operational constraints model the technical conditions for operating thermal and hydropower plants, power transmission and storages and consider the flexibility provided to the energy system by the hydro-thermal dispatch and the storage units. In particular, the block of equations (3.24)–(3.31) models the thermal component of the energy system. Constraints (3.24) ensure consistency between the binary variables representing the commitment status and those representing the decommissioning decisions, forcing the existing thermal power plants decommissioned within year y ($\theta^-_{k,y} = 1$) to be offline in all hours after decommission. Constraints (3.25) and (3.26) enforce the consistency between the binary variables representing the commitment status and those representing investment decisions. Indeed, inequalities (3.25) impose that projects built within year y ($\theta^+_{k,y} = 1$) can be used to supply load, while thermal units not yet constructed ($\theta^+_{k,y} = 0$) are forced to be offline ($\gamma^c_{k,t} = 0$) in all hours of year y. Inequalities (3.26) model the reinforcements of existing thermal power plants. Specifically, let $k' = UP(k)$ denote the new project that replaces the existing unit k when it starts operating: building project k' within year y ($\theta^+_{k1,y} = 1$) implies the permanent offline status of unit k ($\gamma^c_{k,t} = 0$). As already explained in Section 2.6.2, inequalities (3.27) state that the power output $p^c_{k,t}$ above the minimum power output $\underline{P}_k$ is either bounded above by $\overline{P}_k - \underline{P}_k$, if unit k is online ($\gamma^c_{k,t} = 1$), or zero if unit k is offline ($\gamma^c_{k,t} = 0$).

Constraints (3.28) and (3.29) enforce consistency between the binary variables that represent start-up, shut down and status in adjacent hours. Specifically, the parameter $\gamma^{c}_{k_0}$ in constraints (3.28) represents the status of unit k at the beginning of representative day c, whose value is determined by applying a classification tree trained on historical data. Inequalities (3.30) and (3.31) are the minimum up time and down time constraints, respectively.

Constraints (3.32)–(3.42) model the operation of hydropower plants. Specifically, constraints (3.32) and (3.33) bound the pumping power and the power output of the existing hydroelectric reservoirs below their respective upper limits. If candidate hydropower plant $h \in \mathcal{H}_C$ is built within year y, inequalities (3.34) and (3.35) define the upper bounds to pumping power and power output of new hydroelectric reservoirs, otherwise set the corresponding variables to zero. Constraints (3.36) limit for each reservoir the power production and the pumping power in the same hour, while inequalities (3.37) impose an upper bound to the energy spillage from reservoirs. Equations (3.38)–(3.40) create the continuity in storage across the entire time horizon by checking at regular intervals (i.e., every M hours) that all the energy charged and discharged since the previous check point plus the total energy at the previous check point are within bounds. Equations (3.41) enforce the equality between energy levels of each reservoir h at the beginning and the end of the planning horizon.

Constraints (3.42)–(3.47) model the operation of batteries. Specifically, inequalities (3.42)–(3.44) impose upper bounds to charge, discharge and stored energy and enforce consistency between the values of investment and operational variables. Energy balances (3.45) apply to all hours but the first of the representative days and they ensure that the energy stored by battery b at the end of hour t equals the energy stored at the end of hour $t-1$ (reduced by the loss coefficient $\lambda_b \leq 1$), plus the energy injected in b (reduced by the coefficient $\lambda_b^{\mathrm{IN}} \leq 1$), minus the energy released from the battery (reduced by the coefficient $\lambda_b^{\mathrm{OUT}} \geq 1$). Equations (3.46) impose the energy balance for the first hour of each representative day, while constraints (3.47) state the equality for each battery b between energy levels at the beginning and the end of each representative day.

Inequalities (3.48) restrict the power flows on the existing transmission lines. Constraints (3.49) impose lower and upper bounds to the power exchanges among zones and enforce consistency between the power flows on candidate transmission lines and the binary variables related to investment decisions, not allowing power flows on candidate lines which have not been built ($\theta_{l,y} = 0$). The zonal balance equations (3.50) impose equality between energy sources and uses in every zone and every hour of each

representative day. Indeed, the left-hand side of these equations represents the hourly energy sources of zone z (given by thermal, solar and wind generation, incoming energy flows, hydro generation and energy released by batteries) and the right-hand side describes the energy uses (represented by the load, outgoing energy flows, pumping power and energy absorbed by batteries). The variables $ENP_{z,t}^{c}$ and $OG_{z,t}^{c}$ allow detecting and evaluating problems in the simulated system that can cause a mismatch between supply and demand. Inequalities (3.51) ensure the fulfilment of zonal reserve requirements provided by thermal power plants.

Target constraints (3.52)–(3.55) model conditions required to promote sustainable development of energy systems. Specifically, inequalities (3.52) impose limits on thermal power generation employing fossil fuels whose availability could be limited in time. These constraints are imposed for each macro-area m, each year y and each fuel f. In particular, they are computed by multiplying the daily consumption of fuel f in all the zones belonging to macro-area m by the weight of each cluster, to take into account the different occurrences of representative days. Inequalities (3.53) impose limits for thermal energy production due to CO_2 emissions and they present a structure very similar to constraints (3.52), as also, in this case, the total daily emission in each representative day is multiplied by the cluster's weight. Equations (3.54) compute the zonal hourly renewable production, by considering solar, wind and hydro generation. Constraints (3.55) control the renewables penetration, forcing the total renewable generation in macro-area m in year y to cover at least ratio $\varphi_{m,y}$ of the total yearly demand for electricity. Finally, constraints (3.56)–(3.68) define the optimization variables.

3.3 Evaluating the achievement by 2030 of the 55% renewable penetration target for the Italian power system

The proposed model has been applied to plan the expansion of the Italian power system to achieve by 2030 a 55% renewable penetration target and a sustainable reduction of CO_2 emissions, according to the goals set by the European Commission. This section describes the simulation assumptions and the results obtained by addressing this research question.

3.3.1 Data for the Italian power system

The simulation assumptions summarized in this paragraph are based on public information provided by ENTSO-E, ENTSOG, Terna and GME [50, 51, 52, 53, 54]. In our tests, we considered a scenario elaborated by CESI S.p.A. to represent the Italian power system. Such a scenario includes 174 existing thermal power plants, 26 candidate thermal units, 18 existing equivalent hydropower plants, 5 candidate pumping power plants, 9 existing transmission lines, 15 candidate lines, and 14 candidate batteries.

The market analysis considers the Italian electric power system divided into seven interconnected market zones: North (ITn), Central-North (ITcn), Central-South (ITcs), South (ITs), Calabria (ITcal), Sicily (ITsic), and Sardinia (ITsar). The neighbouring countries are modelled as four equivalent areas: Montenegro (ME), Greece (GR), Tunisia (TN) and one single zone named Europe (EU) that summarizes the power flow at the northern Italian border. Each of these equivalent areas is characterized by a set of equivalent power units whose bidding can model a dynamic Import/Export with the Italian power system. Fig. 3.1 shows the existing interconnections as well as a set of candidate new interconnections among which the least cost generation and transmission expansion tool can choose.

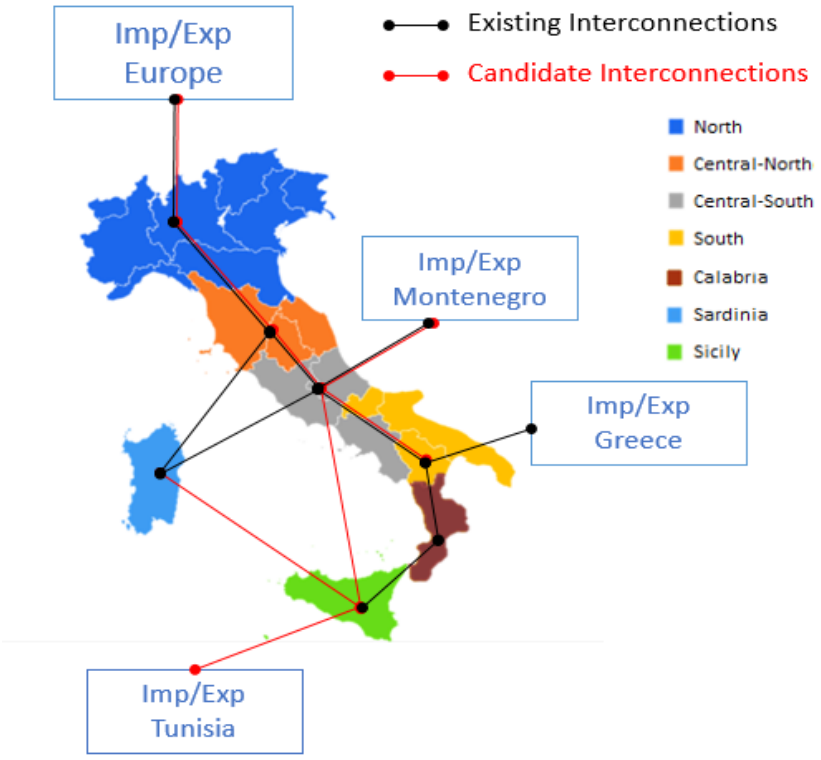


Fig. 3.1 Existing and candidate interconnections in the Italian power system

Table 3.1 Number of representative days for different threshold values

Input threshold	10%	5%	2.5%	1%
Number of representative days	3	5	9	18

On such a system, we performed a GTEP analysis under a 11-year planning horizon, from 2020 up to 2030. For modeling short-term operation, we have applied to the Italian System the procedure described in Section 2.3.2 by fixing at 5% the threshold for the system average mean absolute percentage error in load duration curve approximation, obtaining five representative days for each year of the planning horizon. The value of 5% for the threshold has been chosen after several tests performed on the Italian scenario, which showed that this choice was a good balance between computational costs and approximation accuracy. By modifying the threshold, the number of representative days to be used changes as well, as shown in Table 3.1.

Fig. 3.2 illustrates the five representative days in year 2020 for the North zone, which has the highest electricity demand. As can be noticed, each of the representative days is characterized by 24 values for load, solar capacity factors, and wind capacity factors and by a specific weight.

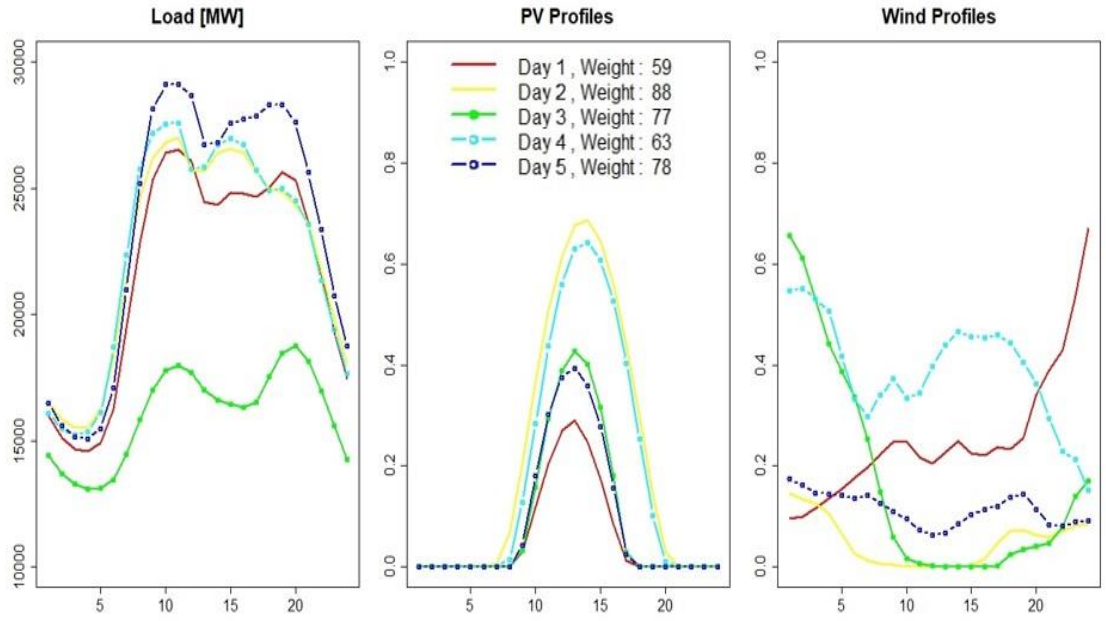
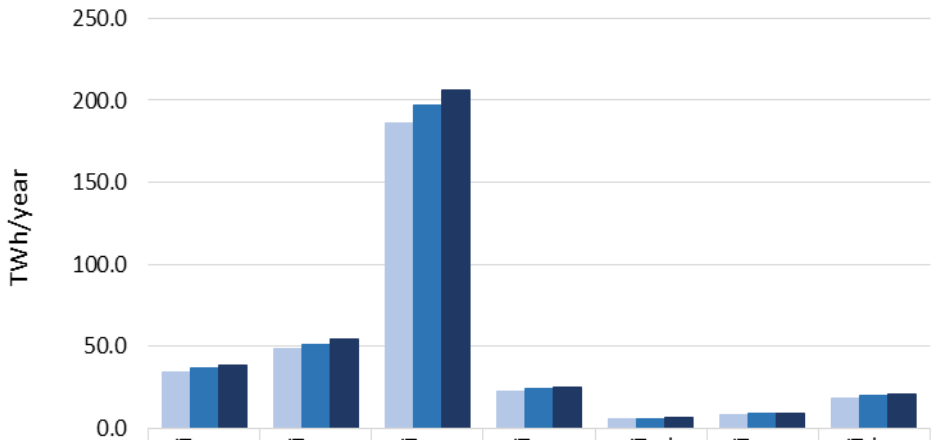


Fig. 3.2 Daily load, solar and wind profiles for the North zone in the five representative days in year 2020



	ITcn	ITcs	ITn	ITs	ITcal	ITsar	ITsic
2020	34.4	48.2	185.8	22.8	5.8	8.5	18.6
2025	36.6	51.2	197.0	24.2	6.1	9.0	19.7
2030	38.3	54.1	206.1	25.2	6.4	9.4	20.6

Fig. 3.3 Zonal load [TWh/year]

Representative days for each of the following 10 years of the planning horizon have been obtained by applying an annual average demand growth of 1%. For the sake of brevity, Fig. 3.3 shows the yearly zonal load only in years 2020, 2025 and 2030, i.e., at the beginning, the middle and the end of the planning horizon.

As far as the renewable installed capacity growth is concerned, a lower bound of 55% of RES penetration (calculated as the ratio between the renewable production and the expected demand) has been imposed for year 2030. It is worth mentioning that this value includes the production from hydro sources. Based on this lower bound set by the Italian Energy Plan, the tool can decide how much to install and what is the best generation mix; however, it has to respect the minimum and maximum targets of photovoltaic and wind capacity set for every expansion plan year, according to the Italian Energy Plan.

Regarding the thermal fleet, based on data provided by the Italian TSO, the existing set of power plants has been implemented together with all related technical information and decommissioning details.

Table 3.2 Installed, outgoing and incoming capacity [GW]

	CCGT	COAL	OIL	GT	**TOT**
Installed in 2020	37.75	8.15	1.42	2.57	**49.89**
Outgoing capacity 2020/2030	3.57	8.15	1.42	1.45	**14.59**
Incoming capacity 2025/2030	6.35	-	-	1.99	**8.34**

The achievement of the challenging RES penetration targets requires the availability of an appropriate reserve margin that can be provided by thermal generation, therefore a set of candidate Combined Cycle Power Plants and Open Cycle Power Plants (both fuelled by natural gas) has been considered to analyse how the system needs in terms of generation capacity and flexibility can be satisfied. Table 3.2 resumes the above-mentioned assumptions.

As regards to fuel consumption, we have considered a CO_2 emission cap of 70 Mtons for each year of the planning horizon, according to levels set by the European Commission, aimed at reducing the impact of the electricity sector on greenhouse gas emissions. In terms of the storage system, we have considered the possibility of adding new pumping units in the southern regions of Italy to the existing hydro pumping fleet and/or the possibility of including in the system Lithium-ion batteries (cheaper but with less storage capacity) or Sodium-ion batteries (a little more expensive but with higher storage capacity). Table 3.3 resumes the technical data related to candidate storage projects. The investment cost is assumed to decrease during the period 2020-2030, with 9% rate until 2025 and with 5% until 2030.

For the economic factors, the values relating to Investment Costs (IC), Fixed Costs and Decommissioning Costs are shown below, both for traditional plants and photovoltaic and wind plants (see Table 3.4).

Table 3.3 Technical data of storage systems

	Candidate Pumping Units					**Sodium-ion Batteries**	**Lithium-ion Batteries**
	Central South	South	Calabria	Sardinia	Sicily		
Maximum energy to power ratio [h]	14	14	14	14	14	6	4
$\overline{E}^{\text{OUT}}$ [MW]	1000	450	1250	800	480	600	600
$\overline{E}^{\text{IN}}$ [MW]	1000	450	1250	800	480	600	600
IC@2020 [€/kWh]	82	82	82	82	82	400	350
Variable Cost [€/MWh]	2	2	2	2	2	30	20

Table 3.4 Economic factors for traditional power plants

	CCGT	COAL	OIL	GT	PV	Wind
IC [€/kW]	800	-	-	400	1000	1300
Fixed Costs [€/MW/year]	10	32	32	5	0	0
Decommissioning [€/kW]	6	20	20	3	-	-

Both PV and wind investment costs are supposed to decrease along the planning horizon due to technology development. Specifically, while PV investment cost in the current analysis goes from 1000 €/kW in 2020 to 600 €/kW in 2030, the wind investment cost is supposed to decrease from 1300 €/kW in 2020 to 900 €/kW in 2030.

Table 3.5 summarizes the assumptions on prices introduced in our analysis. It is worth mentioning that codes Gas, Gasoil, and Coal in Table 3.5 represent the fuels used by Italian thermal plants, while code EUmix refers to the expected fuel cost in the foreign countries interconnected with the Italian market zones. Specifically, this cost reflects the expected variation of the generation mix in foreign countries, considering several factors such as increasing renewable penetrations and gas consumption, decreasing coal consumption and nuclear phase-out.

Table 3.5 Expected fuel and CO_2 prices

	2020	2025	2030
Gas [€/Gcal]	23.43	28.15	32.86
Gasoil [€/Gcal]	90.00	89.27	88.53
Coal [€/Gcal]	9.50	10.72	11.93
EUmix [€/Gcal]	19.68	26.78	33.88
CO_2 [€/ton]	19.00	28.67	38.33

3.3.2 Results and discussion

We solved the proposed model on an ASUS laptop with a 3 GHz Intel Core i7-5500U Processor and 4 GB of RAM using solver Gurobi under GAMS 24.7.4. The total time needed to solve the problem is 2,884 seconds, corresponding to 48 minutes and 4 seconds. Table 3.6 provides more information about computational costs, specifying size and solution time of the problem.

Table 3.6 Size and solution time of the optimization model

# Constraints	# Decision Variables	# Discrete Variables	# Non Zero Elements	CPU Time [seconds]
1,581,805	1,314,831	274,830	6,507,341	2,884

The system costs for the whole expansion planning period are shown in Table 3.7. As can be observed, there is a remarkable difference between investment and operational costs: while the sum of investment, decommissioning and fixed costs represents 24.8% of total costs, operating costs account for 75.2% of total costs. Specifically, the most relevant cost for the system is related to the production costs of thermoelectric power plants, which include O&M and fuel consumption, representing 99.7% of operational costs and 75.0% of total costs. Also the installation of new renewable capacity has a significant impact on total system costs, with the solar and wind power capacity expansion costs representing 15.5% and 5.3% respectively of total costs.

Table 3.7 System costs breakdown for expansion plan period

Costs	M€	%
Thermal Capacity Expansion	3,234	1.57%
Wind Capacity Expansion	10,998	5.32%
Solar Capacity Expansion	32,015	15.50%
Transmission Capacity Expansion	522	0.25%
Pumping Units Capacity Expansion	3,159	1.53%
Batteries Capacity Expansion	1,091	0.53%
Decommissioning Costs	207	0.10%
Thermal Fixed Costs	4	0.002%
Thermal Production Cost	154,875	74.98%
Start-Up Costs	286	0.14%
Hydro Operation Costs	103	0.05%
Batteries Operation Costs	61	0.03%
Penalties for Overgeneration	0	0%
Penalties for Energy Not Provided	0	0%
Total System Costs	**206,555**	**100%**

Table 3.8 Renewable generation capacity expansion [GW] divided by source and implementation year

Expansion capacity [GW]	PV	Wind
2020	6.280	2.760
2021	11.765	1.766
2022	3.537	0.772
2023	3.512	0.776
2024	2.706	0.776
2025	2.706	0.776
2026	2.706	0.776
2027	2.706	0.776
2028	2.706	0.776
2029	2.706	0.645
2030	2.706	0.538
Total	**44.041**	**11.136**

Table 3.8 shows the additional capacity of wind and PV installed in order to reach the RES penetration target in year 2030. The RES installed capacity consists of 44.041 GW of PV and 11.136 GW of wind power: this unbalance may be explained by the lower costs of the PV technology with respect to the wind technology.

As far as interconnection projects are concerned, new national and international cross border lines must be implemented in year 2025 to better exploit the stochastic renewable energy sources and compensate for the decommissioning of some Italian thermoelectric power plants. The selected interconnections are listed in Table 3.9.

Table 3.9 Candidate interconnections selected by the model

From	To	Transmission Limits	Year of intervention
Tunisia	Sicily	[–600 MW ; 600 MW]	2025
Central-South	Central-North	[–150 MW ; 150 MW]	2025

Table 3.10 Installed capacity of energy storage systems [MW]

Zone	Technology	Installed Capacity [MW]
North	Lithium-Ion Batteries	600
Central-South	Pumping Unit	1000
South	Pumping Unit	450
Calabria	Lithium-Ion Batteries	600
Calabria	Pumping Unit	1250
Sicily	Pumping Unit	480
Sardinia	Pumping Unit	800

Moreover, the tool couples the installed RES capacity with energy storage systems, installing throughout the planning period 1.2 GW of batteries and 3.98 GW of pumping units. A list that summarizes the installed capacity according to technology and zone is reported inTable 3.10. As can be noticed, as regards to batteries capacity, the model suggests installing Lithium-Ion batteries due to the lower investment costs with respect to Sodium-Ion batteries.

In the list of thermoelectric candidate projects, ten CCGT power plants have been selected as thermoelectric expansion capacity, starting operation in 2025 to ensure the availability of energy reserve margins. The new thermal power plants introduced in the system are mainly located in the North, Central-South, South, and Sardinia zones. On the contrary, the decommissioning of coal power plants has been planned for years 2020 and 2025 according to decarbonisation targets.

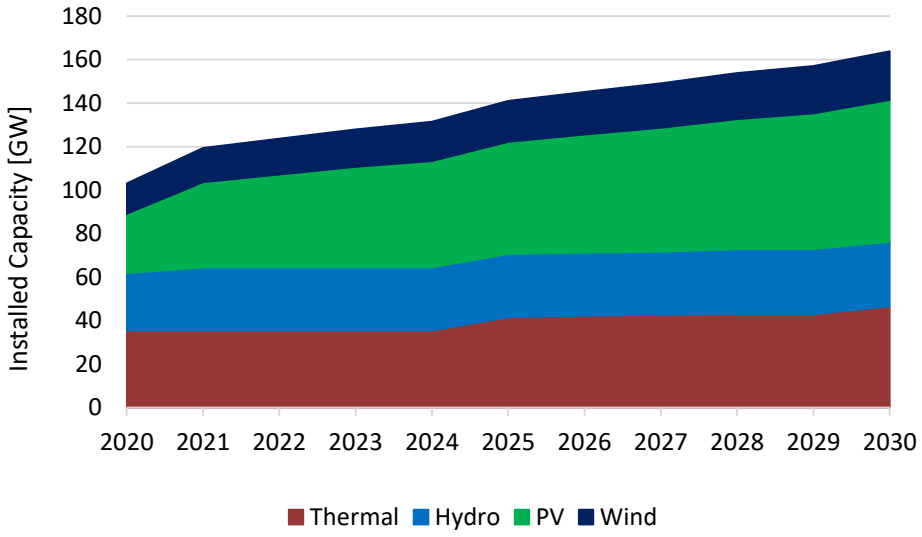


Fig. 3.4 Evolution of the Italian capacity mix over the planning horizon

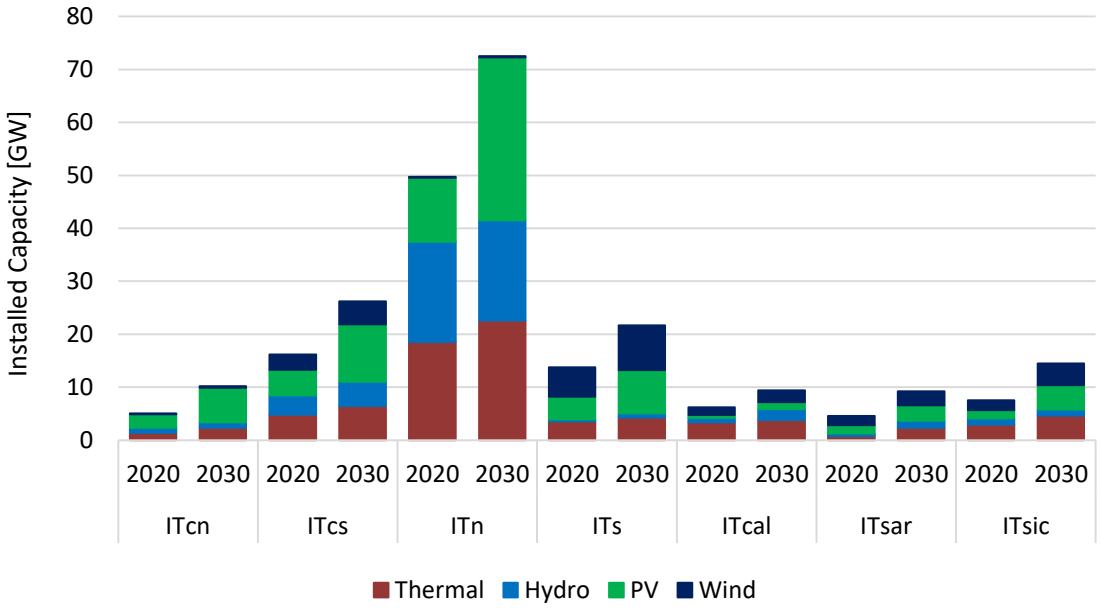


Fig. 3.5 Installed capacity by source for each market zone at the beginning and the end of the horizon

Fig. 3.4 shows the evolution over the planning horizon of the total installed capacity in the Italian power system divided by source. As can be noticed, to reach the 55% target for renewable penetration, the proposed model suggests to install large shares of solar and wind power capacities, while thermal and hydro power capacities only marginally change during the planning horizon.

Fig. 3.5 shows the distribution of the installed capacity between the Italian market zones, highlighting the differences between the beginning and the end of the planning period. As can be observed, half of the total solar expansion capacity, as well as most of the thermoelectric expansion capacity, is installed in the North, due to its high electricity demand, which is almost 60% of the Italian load. Instead, the wind expansion is mainly located in southern regions, which are characterized by the highest wind capacity factors.

Fig. 3.6 shows the evolution over the planning horizon of the total energy divided by source produced in the Italian power system. As can be observed, the thermal component represents the main generation source in every year of the planning horizon. Specifically, the total power generation from thermal units does not significantly change over the years, being equal to 183 TWh in 2020 and to 179 TWh in 2030. Due to the installation of new pumping units, the production from hydro resources increases from 60 TWh in 2020 to 69 TWh at the end of the planning horizon. The installation of large shares of RES capacity increases the generation from both wind power plants (from 28 TWh in 2020 to 44 TWh in 2030) and solar power plants (from 36 TWh in 2020 to 85 TWh in 2030).

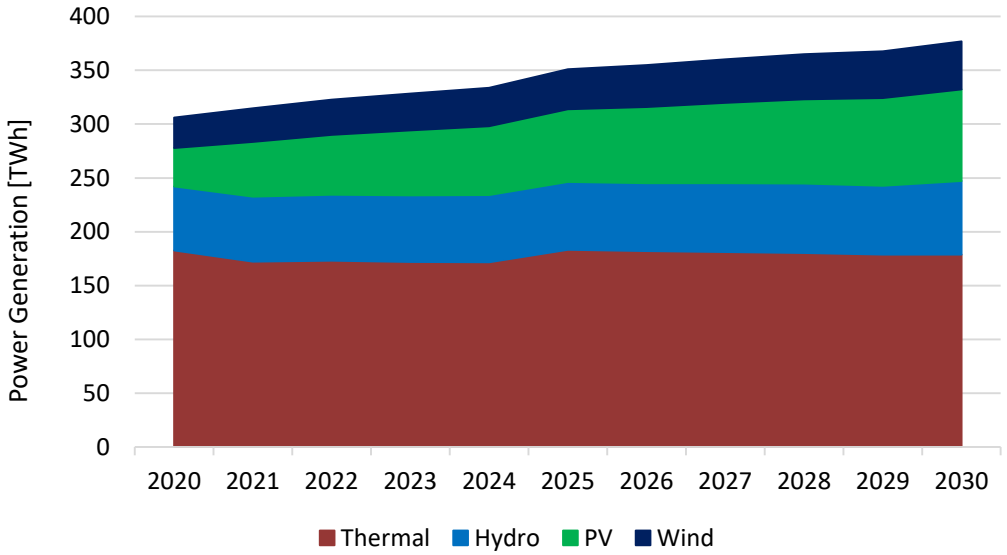


Fig. 3.6 Evolution of the Italian generation mix over the planning horizon

Moreover, the total yearly load for the Italian power system in 2030 corresponds to 360 TWh, while the total generation from renewable sources in 2030 equals 198 TWh, hence the target of reaching 55% of renewable penetration by 2030 is fully achieved. Finally, Fig. 3.7 illustrates for each Italian zone the power generation divided by energy source at the beginning and the end of the planning horizon. As can be noticed, most of the hydroelectric and solar power generation is located in the North zone, while the energy production from wind power plants is mainly located in the southern regions of the country.

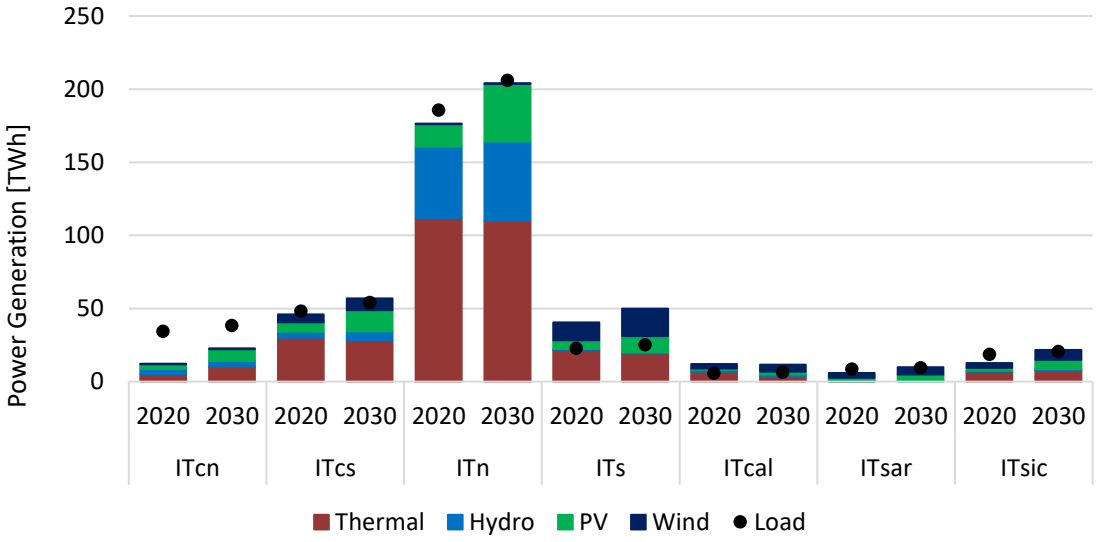


Fig. 3.7 Generation by source for each market zone at the beginning and the end of the horizon

3.4 Chapter conclusions

In this chapter, an optimization model to plan investment decisions to reach the decarbonisation targets in the power sector has been introduced. First, we have presented our approach to the GTEP problem, which is based on a centralized approach: by assuming the perspective of a single central entity, we plan the joint expansion of generation and transmission facilities so as to minimize the total system costs. Such an approach is justified by the objective of our research, which is to develop a computational tool that could support regulators in searching for optimal policies, focusing on the inclusion of as many engineering details as possible, while neglecting market aspects.

Then, a deterministic model for the GTEP analysis has been presented. Such a model co-optimizes strategic and operational decisions for transmission, generation and storage facilities, providing a very detailed representation of the power system and including constraints to limit the CO_2 emissions and to increase the generation from renewable power sources.

The proposed model has been applied to plan the expansion of the Italian power system to achieve by 2030 a 55% renewable penetration target and a sustainable reduction of CO_2 emissions, according to the goals set by the European Commission. The simulation assumptions and the results obtained by addressing this research question have also been discussed in the chapter.

The main limitation of the proposed analysis is the deterministic assumption on fossil fuel prices and emission costs, which strongly influences the results of the simulation. However, given the long-term horizon, this hypothesis could be very restrictive. Indeed, since expansion plans are usually provided for long-term horizons, system conditions at the end of the planning horizons are generally uncertain. The long-term uncertainty should be accurately included in the decision making process to provide more reliable decisions to the actors involved in the GTEP analysis. The inclusion of the long-term uncertainty in the expansion planning framework is addressed in the following chapter.

Chapter 4

Including long-term uncertainties in the expansion planning framework

4.1 Research motivation

Investment decisions in the power sector are characterized by a high level of uncertainty. Indeed, since expansion plans are usually provided for a long-term planning horizon, the future system conditions are generally uncertain at the time the expansion plans are decided. Different sources of uncertainty may affect planning decisions and must be considered in the decision-making process. They can be divided into two groups, namely short-term and long-term uncertainties. Specifically, short-term uncertainties include the stochastic production from intermittent renewable energy sources and the demand variability throughout the hours of the day and the days of the week. These uncertainties are also known as random uncertainties, since they have an underlying probability distribution, which can be approximated from historical data. Instead, long-term uncertainty refers to long-term dynamics, including future values of investment costs, fossil fuel prices and availability, long-term increase/decrease in demand, change in geographical demand distribution, and policy constraints, such as carbon prices. These uncertainties are also known as non-random uncertainties, since they are not usually probability-based: their probabilities are generally obtained based on expert judgement.

To obtain more robust results when planning the joint generation and transmission expansion, the different sources of uncertainty have to be considered in the analysis. While short-term uncertainty can be captured by accurately selecting the representative days, long-term uncertainty can be included in the decision making framework by developing a two-stage stochastic programming MILP model to plan the investment decisions in the power sector. In a stochastic programming approach, the future realization of every uncertain parameter is described by a set of scenarios. Thus, the inclusion of the long-term uncertainty in the analysis further increases the complexity of the problem, making the real-scale GTEP problems computationally intractable. To solve such problems, efficient solution algorithms that exploit the decomposable structure of

two-stage stochastic programming models have to be applied. Therefore, in this chapter also a comprehensive description of the multi-cut Benders decomposition algorithm applied to the proposed stochastic model is provided.

The structure of the chapter is as follows. Section 4.2 reviews the existing literature for expansion planning under uncertainty. Section 4.3 describes the assumptions introduced in the analysis and formulates the expansion planning problem as a two-stage stochastic MILP model. The solution algorithm is presented in Section 4.4. A case study concerning the Italian power system is introduced in Section 4.5. Finally, Section 4.6 concludes the chapter.

4.2 Literature review

Several approaches have been developed in the literature to accurately represent the uncertainty framework in the planning decisions, including fuzzy decision approach, chance-constrained models, Monte Carlo simulation, robust optimization, adaption programming and stochastic programming. In particular, given the relevant penetration of intermittent renewable power sources, many recent studies focus on the influence of these uncertainties on generation and transmission planning. For instance, in [56] an algorithm is developed for multi-objective optimization transmission expansion planning considering wind farm generation and combining Monte Carlo simulation and Point Estimation Method to investigate the effects of network uncertainties. Reference [57] proposes a chance constrained formulation to tackle the uncertainties of load and wind turbine generators in transmission network expansion planning. Load and wind power generation are considered the main sources of uncertainty also in [58], where authors propose an efficient approach for probabilistic transmission expansion planning, dealt with by a Benders decomposition algorithm combined with a Monte Carlo method.

Other studies deal with long-term uncertainties. For instance, authors in [59] address the problem of transmission expansion planning under uncertainty in an electric energy system, considering future demand growth and the availability of generation facilities as main uncertainty sources. A robust optimization model is used to derive the investment decisions that minimize the system's total costs by anticipating the worst case realization of the uncertain parameters within an uncertainty set. Instead, reference [60] proposes a robust generation and transmission expansion planning model, including flexible AC transmission systems (FACTS) devices and considering the uncertainty related to the

annual net load duration curve. A robust optimization approach is also adopted in [61] and [62] to address generation expansion and transmission expansion, respectively. Specifically, reference [61] introduces a multi-year robust methodology for expansion planning, modelling the uncertainties associated with forecasted electricity load demand, as well as estimated investment and operation costs, through distribution-free bounded intervals producing polyhedral uncertainty sets. In [62] two optimization criteria for the transmission expansion planning problem under the robust optimization paradigm are studied, where maximum cost and maximum regret of the expansion plan over all uncertainties are minimized, respectively.

Adaptation programming represents another method to cope with uncertainty. As described in [63], adaptation programming designs a flexible system by minimizing the sum of investment and operational cost and of system future adaptation cost to the conditions of other identified scenarios. Reference [64] further explores the adaptation programming method by applying this kind of model to a small system over a 40 year planning horizon, considering wind and solar build cost, carbon taxes, demand and peak demand growth, fuel prices and transmission costs as uncertainty sources.

Among all the techniques developed to include uncertainty in the expansion planning framework, the most widely used is stochastic programming, a methodology introduced in the 1950s that uses a set of scenarios to model the future realization of the uncertain parameters in the considered planning horizon [65]. In recent years, several studies in the field of stochastic programming have been carried out, leading to the development of two classes of methods: two-stage and multi-stage models. In a typical two-stage stochastic model, the investment decisions represent first stage decisions, which are made before any uncertainty is revealed. Operational decisions are instead second stage decisions, made after realization of parameter values. For instance, in [66] a two-stage stochastic programming model for joint generation and transmission expansion is presented, considering as random events the demand, the equivalent availability of the generating plants and the transmission capacity factor of the transmission lines. Reference [67] presents a stochastic two-stage optimisation model to evaluate interregional grid reinforcements in Great Britain. The same approach is also used in [68] to determine the type and quantity of power plants to be constructed in each year of an extended planning horizon, considering uncertainty regarding future demand and fuel prices. Authors in [69] propose a two-stage stochastic generation expansion model, where the long-term wind power uncertainty is represented by a set of scenarios.

The two-stage approach can be extended to a multi-stage method, constructing models that are both more flexible and complex. As explained in [70], in multi-stage

approaches expansion decisions are made at different stages, i.e., different points in time of the planning horizon. The expansion decisions at each stage depend on the scenario realization of the previous periods, but they do not depend on the future scenario realizations. Examples of multi-stage models are represented by [71], [72] and [24]. For instance, in [71] a multi-stage multi-scale linear stochastic model to optimize electricity generation, storage and transmission investments over a long planning horizon is presented. Both long-term uncertainties, such as investment and fuel-cost changes and long-run demand-growth rates, and short-term uncertainties, such as hour-to-hour demand and renewable-availability uncertainty, are considered in this analysis and the progressive hedging algorithm is applied to decompose the original model by scenario. Reference [72] deals with wind power investments considering three major issues: the production variability and uncertainty of wind power facilities, the eventual future decline in wind power investment costs and the significant financial risk involved in such investment decisions. Recognizing the previous important issues, this paper proposes a risk-constrained multi-stage stochastic programming model to make optimal investment decisions on wind power facilities along a multi-stage horizon. Finally, in [24] authors study how uncertain future renewable penetration levels impact the electricity system and try to quantify effects for the Central European power market, by applying a multi-stage stochastic investment and dispatch model to analyse the effects on investment choices, electricity generation and system costs. Although multi-stage approach better represents long-term dynamics than two-stage method, the complexity of the problem in multi-stage models is further increased. Finding a right balance between modeling accuracy and computation tractability remains an open research topic.

Table 4.1 reports a list of important features for expansion planning models, such as the modeling approach, the inclusion of investment decisions in new generation, storage and transmission facilities, the temporal detail in power system's operation evaluation, the inclusion of unit commitment constraints, the long-term policies considered in the expansion planning framework and the number of periods in which investment decisions can be made along the planning horizon. According to these characteristics, some relevant works in the literature and the model proposed in this chapter are compared.

Table 4.1 Comparison between relevant works in the literature and the proposed model. RO = Robust Optimization; AP = Adaptation Programming; SP = Stochastic Programming; LDC = Load Duration Curve

Ref.	Modeling Approach	Generation Expansion			Storage Expansion		Transmission Expansion	Operation Evaluation	UC Constraints	Policy Goals	Investment Periods
		Thermal	Wind	Solar	Long Term	Short Term					
[60]	RO	Yes	Yes	No	No	No	Yes	LDC	No	None	5
[63]	AP	Yes	Yes	Yes	Yes	No	No	LDC	No	RES penetration	20
[64]	AP	Yes	Yes	Yes	No	No	Yes	LDC	No	None	2
[67]	2-Stage SP	Yes	Yes	No	No	No	Yes	Sampled hours	No	None	2
[68]	2-Stage SP	Yes	Yes	No	No	No	No	LDC	No	None	10
[69]	2-Stage SP	Yes	No	No	No	No	No	Sampled weeks	Yes	None	1
[71]	Multi-Stage SP	Yes	Yes	Yes	No	Yes	Yes	Rep. days	No	None	4
[24]	Multi-Stage SP	Yes	No	No	Yes	Yes	No	Rep. days	No	None	4
[23]	2-Stage SP	Yes	Yes	No	No	No	No	Rep. days	Yes	RES penetration	1
This chapter	2-Stage SP	Yes	Yes	Yes	Yes	Yes	Yes	Rep. days	Yes	RES penetration CO_2 emissions Fuels availability	21

4.3 Modeling framework

In this section, we present our approach to include long-term uncertainties in the GTEP analysis by introducing a two-stage stochastic MILP model. The distinct feature of the proposed model is a detailed representation of the short-term operation, which is required to properly evaluate all the challenges related to integrating high shares of renewables to reach decarbonisation targets. The motivations leading to the choice of a two-stage stochastic programming approach as well as the modeling assumptions, the notation and the mathematical formulation of the optimization model are presented in the following subsections.

4.3.1 Uncertainty modeling

In our approach, we focus on the uncertainty of fuel prices and CO_2 price, because of the important role they play in the generation and transmission expansion planning by affecting the merit order of thermal power plants and the economic viability of renewable generation. To deal with these long-term uncertainties, different scenarios for prices are introduced, with the set of scenarios denoted by $\mathcal{W}$. The joint GTEP problem is then formulated as a two-stage stochastic MILP model, being decommissioning of thermal power plants and investments in new generation, transmission and storage facilities first-stage variables and operating decisions second-stage variables. The choice of a two-stage model rather than a multi-stage approach is first motivated by the objective of our analysis, which consists in supporting government authorities in defining a single optimal investment trajectory rather than multiple investment plans. The choice of a two-stage approach is also motivated by computational restrictions. Indeed, as can be noticed in Table 4.1, existing multi-stage models represent power systems short-term operation with a lower level of accuracy to the proposed model. For instance, no long-term storage facilities or unit commitment constraints are included in references [24] and [71], while the proposed model provides an accurate modeling of short-term dynamics by considering the unit commitment constraints on a plant-by-plant level to properly evaluate all the challenges related to integrating high shares of renewables. Moreover, to reduce the computational burden, existing multi-stage models consider only a limited number of periods in which investment decisions can be made along the planning horizon. Our proposed model differs from multi-stage models by allowing investment decisions in each year of the planning horizon.

4.3.2 Notation

For the sake of clarity, in this paragraph we provide the full notation needed to formulate the GTEP problem as a two-stage stochastic MILP model.

Sets

$\mathcal{Y}$	Set of years, indexed by y and i
$\mathcal{Z}$	Set of zones, indexed by z
$\mathcal{M}$	Set of macro-areas, indexed by m
$\mathcal{K}_E$	Set of existing thermal power plants
$\mathcal{K}_C$	Set of candidate thermal power plants
$\mathcal{K}$	Set of thermal power plants ($\mathcal{K} = \mathcal{K}_E \cup \mathcal{K}_C$), indexed by k and k'
$\Omega_z^k \subset \mathcal{K}$	Set of thermal power plants located in zone z
$\mathcal{K}_{E1} \subset \mathcal{K}_E$	Set of existing thermal power plants to be mandatorily decommissioned
$\mathcal{K}_{E2} \subset \mathcal{K}_E$	Set of existing thermal power plants that may be optionally decommissioned
$\mathcal{K}_{C1} \subset \mathcal{K}_C$	Set of candidate thermal power plants to be mandatorily constructed
$\mathcal{K}_{C2} \subset \mathcal{K}_C$	Set of candidate thermal power plants that may be optionally constructed
$\mathcal{AK}_j \subset \mathcal{K}_C$	j-th group of associate candidate thermal power plants
$J^{\mathcal{AK}}$	Set of groups of associate candidate thermal power plants
$\mathcal{MEK}_j \subset \mathcal{K}_C$	j-th group of mutually exclusive candidate thermal power plants
$J^{\mathcal{MEK}}$	Set of groups of mutually exclusive candidate thermal power plants
$\mathcal{L}_E$	Set of existing transmission lines
$\mathcal{L}_C$	Set of candidate transmission lines
$\mathcal{L}$	Set of transmission lines ($\mathcal{L} = \mathcal{L}_E \cup \mathcal{L}_C$), indexed by l and l'
$\mathcal{L}_{C1} \subset \mathcal{L}_C$	Set of candidate transmission lines to be mandatorily constructed
$\mathcal{L}_{C2} \subset \mathcal{L}_C$	Set of candidate transmission lines that may be optionally constructed
$\mathcal{AL}_j \subset \mathcal{L}_C$	j-th group of associate candidate transmission lines
$J^{\mathcal{AL}}$	Set of groups of associate candidate transmission lines
$\mathcal{MEL}_j \subset \mathcal{L}_C$	j-th group of mutually exclusive candidate transmission lines
$J^{\mathcal{MEL}}$	Set of groups of mutually exclusive candidate transmission lines
$\mathcal{H}_E$	Set of existing hydropower plants
$\mathcal{H}_C$	Set of candidate hydropower plants

$\mathcal{H}$	Set of hydropower plants ($\mathcal{H} = \mathcal{H}_E \cup \mathcal{H}_C$), indexed by h
$\Omega_z^h \subset \mathcal{H}$	Set of hydropower plants located in zone z
$\mathcal{B}$	Set of batteries, indexed by b
$\Omega_z^b \subset \mathcal{B}$	Set of batteries located in zone z
$\mathcal{F}$	Set of fuels, indexed by f
$\Phi_{z,f} \subset \Omega_z^k$	Set of thermal power plants located in zone z using fuel f
$\mathcal{C}^y$	Set of representative days of year y, indexed by c
$\mathcal{D}^y$	Set of all days of year y, indexed by d and d'
$\mathcal{D}_M^y \subset \mathcal{D}^y$	Set of days of year y in which the level of the long-term storage is checked
$\mathcal{T}$	Set of hours, from 1 to 24, indexed by t and τ
$\mathcal{W}$	Set of scenarios, indexed by w
$ma(z)$	macro-area that contains zone z
$UP(k)$	upgrade project of existing thermal power plant k
$rz(l)$	Receiving-end zone of transmission line l
$sz(l)$	Sending-end zone of transmission line l
$fuel(k)$	Fuel used in thermal power plant k
$\mathrm{Map}_{d,c}$	Cluster index, i.e., injective map of each day d to a representative day c

Parameters

y_0	[−]	Reference year to which all investment costs are discounted
r	[−]	Annual discount rate
$prob_w$	[−]	Probability of scenario w
c_{ENP}	[€/MWh]	Penalty for energy not provided
c_{OG}	[€/MWh]	Penalty for over-generation
wg_c	[−]	Weight of representative day c
DC_k	[€]	Decommissioning cost of existing thermal power plant $k \in \mathcal{K}_E$
IC_k^{The}	[€]	Investment cost of candidate thermal power plant $k \in \mathcal{K}_C$
FC_k	[€]	Annual fixed costs of thermal power plant $k \in \mathcal{K}$
$CM_{k,y,w}$	[€/MWh]	Marginal production cost of thermal power plant k in year y in scenario w

$\underline{\tau}_k$	[−]	Earliest date for construction/decommission of thermal power plant k
$\overline{\tau}_k$	[−]	Latest date for construction/decommission of thermal power plant k
$\underline{P}_k$	[MW]	Minimum power output of thermal power plant k
$\overline{P}_k$	[MW]	Maximum power produced by thermal plant k
SUC_k	[€]	Start-up cost of thermal power plant k
MUT_k	[h]	Minimum up time of thermal power plant k
MDT_k	[h]	Minimum down time of thermal power plant k
γ^c_{k,w_0}	[−]	Initial ON/OFF status of thermal power plant k in representative day c in scenario w
HR_k	[Gcal/MWh]	Heat rate of thermal power plants of cluster k
OM_k	[€/MWh]	Operative and maintenance cost of thermal power plants of cluster k
$IC^{\mathrm{Sol}}_{z,y}$	[€/MW]	Investment cost of new solar power capacity in zone z in year y
$sol_{z,0}$	[MW]	Solar power capacity installed in zone z at the beginning of the planning horizon
$\underline{PV}_{z,y}$	[MW]	Lower bound for solar power capacity in zone z in year y
$\overline{PV}_{z,y}$	[MW]	Upper bound for solar power capacity in zone z in year y
$IC^{\mathrm{Wind}}_{z,y}$	[€/MW]	Investment cost of new wind power capacity in zone z in year y
$wind_{z,0}$	[MW]	Wind power capacity installed in zone z at the beginning of the planning horizon
$\underline{W}_{z,y}$	[MW]	Lower bound for wind power capacity in zone z in year y
$\overline{W}_{z,y}$	[MW]	Upper bound for wind power capacity in zone z in year y
$D^c_{z,t}$	[MW]	Load in zone z in hour t of representative day c
$R^c_{z,t}$	[MW]	Reserve requirement for zone z in hour t of representative day c
$\mu^c_{z,t}$	[MWh/MW]	Solar power capacity factor for zone z in hour t of representative day c
$\rho^c_{z,t}$	[MWh/MW]	Wind power capacity factor for zone z in hour t of representative day c

$IC_{h,y}^{\mathrm{Hyd}}$	[€]	Investment cost of candidate hydropower plant $h \in \mathcal{H}_C$ in year y
$\underline{\tau}_h$	[−]	Earliest date for construction of hydropower plant h
$\overline{\tau}_h$	[−]	Latest date for construction of hydropower plant h
$Cvar_h$	[€/MWh]	Operating cost of hydropower plant h
$\bar{E}_h^{\mathrm{IN}}$	[MW]	Upper bound on hydropower plant h pumping power
$\bar{E}_h^{\mathrm{OUT}}$	[MW]	Upper bound on hydropower plant h power output
$\overline{sl}_h$	[MWh]	Upper bound on energy spillage from hydropower plant h
$F_{h,t}^c$	[MWh]	Hourly energy inflows for hydropower plant h at time t of representative day c
λ_h^{IN}	[−]	Loss coefficient for hydro plant h pumping $\left(0 \le \lambda_h^{\mathrm{IN}} \le 1\right)$
λ_h^{OUT}	[−]	Loss coefficient for hydro plant h power generation $\left(\lambda_h^{\mathrm{OUT}} \ge 1\right)$
E_{h_0}	[MWh]	Energy content of hydropower plant h at the beginning of planning horizon
EPR_h	[h]	Maximum energy to power ratio (in hours) for hydropower plant h
M	[−]	Size of the temporal window in long-term storage constraints, set to 7 days
$IC_{b,y}^{\mathrm{Batt}}$	[€/MW]	Investment cost for battery b in year y
$Cvar_b$	[€/MWh]	Operating cost of battery b
$\overline{CAP}_b^{\mathrm{Batt}}$	[MW]	Upper bound on battery b installed capacity
$E_{b_0}^c$	[MWh]	Initial energy content of battery b in representative day c
λ_b	[−]	Loss coefficient for energy stored by battery b $(0 \le \lambda_b \le 1)$
λ_b^{IN}	[−]	Loss coefficient for battery b charge $\left(0 \le \lambda_b^{in} \le 1\right)$
λ_b^{OUT}	[−]	Loss coefficient for battery b discharge $(\lambda_b^{out} \ge 1)$
$\bar{E}_b^{\mathrm{IN}}$	[MW]	Upper bound on battery b charge
$\bar{E}_b^{\mathrm{OUT}}$	[MW]	Upper bound on battery b discharge
EPR_b	[h]	Maximum energy to power ratio (in hours) for battery b
IC_l^{Line}	[€]	Investment cost of candidate transmission line $l \subset \mathcal{L}_C$
$\underline{\tau}_l$	[−]	Earliest date for construction of candidate transmission line l

$\overline{\tau}_l$	[−]	Latest date for construction of candidate transmission line l
$\underline{F_l}$	[MW]	Minimum capacity of transmission line l
$\overline{F}_l$	[MW]	Maximum capacity of transmission line l
$ECnt_f$	[Gcal/M. U]	Energy content of fuel f
co_{2_f}	[ton/Gcal]	CO_2 emission factor of fuel f
$\overline{FA}_{f,m,y}$	[M. U]	Upper bound on availability of fuel f in macro-area m in year y
$Pr_{y,w}^{f}$	[€/Gcal]	Price of fuel f in year y in scenario w
$\overline{CO_2}_{m,y}$	[ton]	CO_2 emission limit for macro-area m in year y
$\varphi_{m,y}$	[−]	Lower bound for renewables penetration in macro-area m in year y
$Pr_{y,w}^{\mathrm{CO2}}$	[€/ton]	Emission price in year y in scenario w

Variables

1) First-stage variables

$\delta_{k,y}^{-}$	[−]	1: thermal power plant $k \in \mathcal{K}_E$ is decommissioned in year y; 0: otherwise
$\delta_{k,y}^{+}$	[−]	1: thermal power plant $k \in \mathcal{K}_C$ is built in year y; 0: otherwise
$\delta_{h,y}$	[−]	1: hydro power plant $h \in \mathcal{H}_C$ is built in year y; 0: otherwise
$\delta_{l,y}$	[−]	1: transmission line $l \in \mathcal{L}_C$ is built in year y; 0: otherwise
$\theta_{k,y}^{-}$	[−]	1: thermal power plant $k \in \mathcal{K}_E$ is decommissioned within year y; 0: otherwise
$\theta_{k,y}^{+}$	[−]	1: thermal power plant $k \in \mathcal{K}_C$ is built within year y; 0: otherwise
$\theta_{h,y}$	[−]	1: hydro power plant $h \in \mathcal{H}_C$ is built within year y; 0: otherwise
$\theta_{l,y}$	[−]	1: transmission line $l \in \mathcal{L}_C$ is built within year y; 0: otherwise
$sol_{z,y}$	[MW]	New solar capacity installed in zone z in year y
$wind_{z,y}$	[MW]	New wind capacity installed in zone z in year y
$cap_{b,y}^{\mathrm{Batt}}$	[MW]	Storage capacity of battery b installed in year y

2) Second-stage variables

$\gamma_{k,t,w}^{c}$	[−]	1: thermal power plant k is ON in hour t of representative day c in scenario w; 0: otherwise
$\alpha_{k,t,w}^{c}$	[−]	1: thermal power plant k is started-up in hour t of representative day c in scenario w; 0: otherwise
$\beta_{k,t,w}^{c}$	[−]	1: thermal power plant k is shut-down in hour t of representative day c in scenario w; 0: otherwise
$p_{k,t,w}^{c}$	[MW]	Power production of thermal power plant k at time t of representative day c above its minimum output $\underline{P}_k$ in scenario w
$E_{h,t,w}^{\mathrm{IN},c}$	[MW]	Pumping power of hydro reservoir h in hour t of representative day c in scenario w
$E_{h,t,w}^{\mathrm{OUT},c}$	[MW]	Power output of hydro reservoir h in hour t of representative day c in scenario w
$sl_{h,t,w}^{c}$	[MWh]	Energy spillage from hydro reservoir h in hour t of representative day c in scenario w
$\hat{E}_{h,w}^{\mathrm{LT},d}$	[MWh]	Energy level of hydro reservoir h at the end of day d in scenario w
$E_{b,t,w}^{\mathrm{IN},c}$	[MW]	Charge of battery b in hour t of representative day c in scenario w
$E_{b,t,w}^{\mathrm{OUT},c}$	[MW]	Discharge of battery b in hour t of representative day c in scenario w
$E_{b,t,w}^{c}$	[MWh]	Energy level of battery b in hour t of representative day c in scenario w
$x_{l,t,w}^{c}$	[MW]	Power flow on transmission line l in hour t of representative day c in scenario w
$ENP_{z,t,w}^{c}$	[MWh]	Energy not provided in zone z in hour t of representative day c in scenario w
$OG_{z,t,w}^{c}$	[MWh]	Over-generation in zone z in hour t of representative day c in scenario w
$RNP_{z,t,w}^{c}$	[MWh]	Reserve not provided in zone z in hour t of representative day c in scenario w
$RES_{z,t,w}^{c}$	[MWh]	Renewable generation in zone z in hour t of representative day c in scenario w

4.3.3 Mathematical formulation

The GTEP model introduced in Section 3.2.3 can be modified as follows to include long-term uncertainty on prices:

$$
\begin{aligned}
\min\ z \;=\; & \sum_{y\in\mathcal{Y}}\left(\sum_{k\in\mathcal{K}_E}\frac{DC_k\,\delta^-_{k,y}}{(1+r)^{y-y_0}}+\sum_{k\in\mathcal{K}_C}\frac{IC_k^{\mathrm{The}}\,\delta^+_{k,y}}{(1+r)^{y-y_0}}\right)+\\
& +\sum_{y\in\mathcal{Y}}\left(\sum_{z\in\mathcal{Z}}\frac{IC_{z,y}^{\mathrm{Sol}}\,sol_{z,y}}{(1+r)^{y-y_0}}+\sum_{z\in\mathcal{Z}}\frac{IC_{z,y}^{\mathrm{Wind}}\,wind_{z,y}}{(1+r)^{y-y_0}}\right)+\\
& +\sum_{y\in\mathcal{Y}}\sum_{h\in\mathcal{H}_C}\frac{IC_{h,y}^{\mathrm{Hyd}}\,\delta_{h,y}}{(1+r)^{y-y_0}}+\\
& +\sum_{y\in\mathcal{Y}}\sum_{b\in\mathcal{B}}\frac{IC_{b,y}^{\mathrm{Batt}}\,cap_{b,y}^{\mathrm{Batt}}}{(1+r)^{y-y_0}}+\\
& +\sum_{y\in\mathcal{Y}}\sum_{l\in\mathcal{L}_C}\frac{IC_l^{\mathrm{Line}}\,\delta_{l,y}}{(1+r)^{y-y_0}}+\\
& +\sum_{y\in\mathcal{Y}}\left[\sum_{k\in\mathcal{K}_E}FC_k\left(1-\theta^-_{k,y}\right)+\sum_{k\in\mathcal{K}_C}FC_k\,\theta^+_{k,y}\right]+\\
& +\sum_{w\in\mathcal{W}}prob_w\left[\sum_{y\in\mathcal{Y}}\sum_{c\in\mathcal{C}^y}wg_c\sum_{t=1}^{24}\left(\sum_{k\in\mathcal{K}}CM_{k,y,w}\left(\underline{P}_k\,\gamma^c_{k,t,w}+p^c_{k,t,w}\right)\right.\right.\\
& \qquad +\sum_{k\in\mathcal{K}}SUC_k\alpha^c_{k,t,w}+\sum_{h\in\mathcal{H}}Cvar_h\,E^{\mathrm{OUT},c}_{h,t,w}+\sum_{b\in\mathcal{B}}Cvar_b\,E^{\mathrm{OUT},c}_{b,t,w}\\
& \qquad \left.\left.+c_{ENP}\sum_{z\in Z}ENP^c_{z,t,w}+c_{OG}\sum_{z\in Z}OG^c_{z,t,w}\right)\right] \qquad (4.1)
\end{aligned}
$$

subject to

$$\delta^-_{k,y}=0 \qquad k\in\mathcal{K}_E, y\notin\left[\underline{\tau}_k,\overline{\tau}_k\right] \qquad (4.2)$$

$$\sum_{y\in\mathcal{Y}}\delta^-_{k,y}=1 \qquad k\in\mathcal{K}_{E1} \qquad (4.3)$$

$$\sum_{y\in\mathcal{Y}}\delta^-_{k,y}\le 1 \qquad k\in\mathcal{K}_{E2} \qquad (4.4)$$

$$\delta^+_{k,y}=0 \qquad k\in\mathcal{K}_C, y\notin\left[\underline{\tau}_k,\overline{\tau}_k\right] \qquad (4.5)$$

$$\sum_{y\in\mathcal{Y}} \delta^{+}_{k,y} = 1 \qquad k \in \mathcal{K}_{C1} \tag{4.6}$$

$$\sum_{y\in\mathcal{Y}} \delta^{+}_{k,y} \leq 1 \qquad k \in \mathcal{K}_{C2} \tag{4.7}$$

$$\sum_{y\in\mathcal{Y}} \delta^{+}_{k,y} = \sum_{y\in\mathcal{Y}} \delta^{+}_{k',y} \qquad k \in \mathcal{AK}_j, k' \in \mathcal{AK}_j, j \in J^{\mathcal{AK}} \tag{4.8}$$

$$\sum_{k\in\mathcal{MEK}_j} \sum_{y\in\mathcal{Y}} \delta^{+}_{k,y} \leq 1 \qquad j \in J^{\mathcal{MEK}} \tag{4.9}$$

$$\underline{PV}_{z,y} \leq sol_{z,0} + \sum_{i=1}^{y} sol_{z,i} \leq \overline{PV}_{z,y} \qquad z \in \mathcal{Z}, y \in \mathcal{Y} \tag{4.10}$$

$$\underline{W}_{z,y} \leq wind_{z,0} + \sum_{i=1}^{y} wind_{z,i} \leq \overline{W}_{z,y} \qquad z \in \mathcal{Z}, y \in \mathcal{Y} \tag{4.11}$$

$$\delta_{h,y} = 0 \qquad h \in \mathcal{H}_C, y \notin \left[\underline{\tau}_h, \overline{\tau}_h\right] \tag{4.12}$$

$$\sum_{y\in\mathcal{Y}} \delta_{h,y} \leq 1 \qquad h \in \mathcal{H}_C \tag{4.13}$$

$$\sum_{y\in\mathcal{Y}} cap^{\text{Batt}}_{b,y} \leq \overline{CAP}^{\text{Batt}}_{b} \qquad b \in \mathcal{B} \tag{4.14}$$

$$\delta_{l,y} = 0 \qquad l \in \mathcal{L}_C, y \notin \left[\underline{\tau}_l, \overline{\tau}_l\right] \tag{4.15}$$

$$\sum_{y\in\mathcal{Y}} \delta_{l,y} = 1 \qquad l \in \mathcal{L}_{C1} \tag{4.16}$$

$$\sum_{y\in\mathcal{Y}} \delta_{l,y} \leq 1 \qquad l \in \mathcal{L}_{C2} \tag{4.17}$$

$$\sum_{y\in\mathcal{Y}} \delta_{l,y} = \sum_{y\in\mathcal{Y}} \delta_{l',y} \qquad l \in \mathcal{AL}_j, l' \in \mathcal{AL}_j, j \in J^{\mathcal{AL}} \tag{4.18}$$

$$\sum_{l\in\mathcal{MEL}_j} \sum_{y\in\mathcal{Y}} \delta_{l,y} \leq 1 \qquad j \in J^{\mathcal{MEL}} \tag{4.19}$$

$$\theta^{-}_{k,y} = \sum_{i=1}^{y} \delta^{-}_{k,i} \qquad k \in \mathcal{K}_E, y \in \mathcal{Y} \tag{4.20}$$

$$\theta^{+}_{k,y} = \sum_{i=1}^{y} \delta^{+}_{k,i} \qquad k \in \mathcal{K}_C, y \in \mathcal{Y} \tag{4.21}$$

$$\theta_{h,y} = \sum_{i=1}^{y} \delta_{h,i} \qquad h \in \mathcal{H}_C, y \in \mathcal{Y} \tag{4.22}$$

$$\theta_{l,y} = \sum_{i=1}^{y} \delta_{l,i} \qquad l \in \mathcal{L}_C, y \in \mathcal{Y} \tag{4.23}$$

$$\gamma^{c}_{k,t,w} \leq 1 - \theta^{-}_{k,y} \qquad k \in \mathcal{K}_E, 1 \leq t \leq 24, c \in \mathcal{C}^{y}, y \in \mathcal{Y}, w \in \mathcal{W} \tag{4.24}$$

$$\gamma_{k,t,w}^{c} \leq \theta_{k,y}^{+} \qquad k \in \mathcal{K}_C, 1 \leq t \leq 24, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W} \tag{4.25}$$

$$\gamma_{k,t,w} \leq 1 - \theta_{k',y}^{+} \qquad k \in \mathcal{K}_E, k' = UP(k), 1 \leq t \leq 24, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W} \tag{4.26}$$

$$p_{k,t,w}^{c} \leq \gamma_{k,t,w}^{c}\left(\overline{P}_k - \underline{P}_k\right) \qquad k \in \mathcal{K}, 1 \leq t \leq 24, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W} \tag{4.27}$$

$$\gamma_{k,t,w}^{c} - \gamma_{k,w_0}^{c} = \alpha_{k,t,w}^{c} - \beta_{k,t,w}^{c} \qquad k \in \mathcal{K}, t = 1, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W} \tag{4.28}$$

$$\gamma_{k,t,w}^{c} - \gamma_{k,t-1,w}^{c} = \alpha_{k,t,w}^{c} - \beta_{k,t,w}^{c} \qquad k \in \mathcal{K}, 2 \leq t \leq 24, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W} \tag{4.29}$$

$$\sum_{\tau=t-MUT_k+1}^{t} \alpha_{k,\tau,w}^{c} \leq \gamma_{k,t,w}^{c} \qquad k \in \mathcal{K}, MUT_k \leq t \leq 24, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W} \tag{4.30}$$

$$\sum_{\tau=t-MDT_k+1}^{t} \beta_{k,\tau,w}^{c} \leq 1 - \gamma_{k,t,w}^{c} \qquad k \in \mathcal{K}, MDT_k \leq t \leq 24, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W} \tag{4.31}$$

$$E_{h,t,w}^{\mathrm{IN},c} \leq \bar{E}_h^{\mathrm{IN}} \qquad h \in \mathcal{H}_E, 1 \leq t \leq 24, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W} \tag{4.32}$$

$$E_{h,t,w}^{\mathrm{OUT},c} \leq \bar{E}_h^{\mathrm{OUT}} \qquad h \in \mathcal{H}_E, 1 \leq t \leq 24, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W} \tag{4.33}$$

$$E_{h,t,w}^{\mathrm{IN},c} \leq \bar{E}_h^{\mathrm{IN}} \theta_{h,y} \qquad h \in \mathcal{H}_C, 1 \leq t \leq 24, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W} \tag{4.34}$$

$$E_{h,t,w}^{\mathrm{OUT},c} \leq \bar{E}_h^{\mathrm{OUT}} \theta_{h,y} \qquad h \in \mathcal{H}_C, 1 \leq t \leq 24, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W} \tag{4.35}$$

$$\frac{E_{h,t,w}^{\mathrm{IN},c}}{\bar{E}_h^{\mathrm{IN}}} + \frac{E_{h,t,w}^{\mathrm{OUT},c}}{\bar{E}_h^{\mathrm{OUT}}} \leq 1 \qquad h \in \mathcal{H}, 1 \leq t \leq 24, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W} \tag{4.36}$$

$$sl_{h,t,w}^{c} \leq \bar{sl}_h \qquad h \in \mathcal{H}, 1 \leq t \leq 24, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W} \tag{4.37}$$

$$\hat{E}_{h,w}^{\mathrm{LT},d} = E_{h_0} + \sum_{d'=d-M+1}^{d} \sum_{c \in \mathrm{Map}_{d',c}} \sum_{t=1}^{24} \left(F_{h,t}^{c} + \lambda_h^{\mathrm{IN}} E_{h,t,w}^{c,\mathrm{IN}} - \lambda_h^{\mathrm{OUT}} E_{h,t,w}^{c,\mathrm{OUT}} - sl_{h,t,w}^{c}\right) \qquad h \in \mathcal{H}, d = M, w \in \mathcal{W} \tag{4.38}$$

$$\hat{E}_{h,w}^{\mathrm{LT},d} = \hat{E}_{h,w}^{\mathrm{LT},d-M} + \sum_{d'=d-M+1}^{d} \sum_{c \in \mathrm{Map}_{d',c}} \sum_{t=1}^{24} \left(F_{h,t}^{c} + \lambda_h^{\mathrm{IN}} E_{h,t,w}^{c,\mathrm{IN}} - \lambda_h^{\mathrm{OUT}} E_{h,t,w}^{c,\mathrm{OUT}} - sl_{h,t,w}^{c}\right) \qquad h \in \mathcal{H}, d \in \mathcal{D}_M^y, y \in \mathcal{Y}, d > M, w \in \mathcal{W} \tag{4.39}$$

$$\hat{E}_{h,w}^{\mathrm{LT},d} \leq EPR_h \bar{E}_h^{\mathrm{IN}} \qquad h \in \mathcal{H}, d \in \mathcal{D}_M^y, y \in \mathcal{Y}, w \in \mathcal{W} \tag{4.40}$$

$$\hat{E}_{h,w}^{\mathrm{LT},d} = E_{h\,0} \qquad h \in \mathcal{H}, d = |\mathcal{D}^y|, y \in \mathcal{Y}, w \in \mathcal{W} \tag{4.41}$$

$$E_{b,t,w}^{\mathrm{IN},c} \leq \sum_{i=1}^{y} cap_{b,i}^{\mathrm{Batt}} \qquad h \in \mathcal{B}, 1 \leq t \leq 24, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W} \tag{4.42}$$

$$E_{b,t,w}^{\mathrm{OUT},c} \leq \sum_{i=1}^{y} cap_{b,i}^{\mathrm{Batt}} \qquad h \in \mathcal{B}, 1 \leq t \leq 24, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W} \tag{4.43}$$

$$E_{b,t,w}^{c} \leq EPR_b \sum_{i=1}^{y} cap_{b,i}^{\mathrm{Batt}} \qquad b \in \mathcal{B}, 1 \leq t \leq 24, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W} \tag{4.44}$$

$$E_{b,t,w}^{c} = (1-\lambda_b)\, E_{b,t-1,w}^{c} + \lambda_b^{\mathrm{IN}}\, E_{b,t,w}^{\mathrm{IN},c} - \lambda_b^{\mathrm{OUT}} E_{b,t,w}^{\mathrm{OUT},c}$$
$$b \in \mathcal{B}, 2 \leq t \leq 24, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W} \tag{4.45}$$

$$E_{b,t,w}^{c} = (1-\lambda_b)\, E_{b\,0}^{c} + \lambda_b^{\mathrm{IN}}\, E_{b,t,w}^{\mathrm{IN},c} - \lambda_b^{\mathrm{OUT}} E_{b,t,w}^{\mathrm{OUT},c}$$
$$b \in \mathcal{B}, t = 1, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W} \tag{4.46}$$

$$E_{b,t,w}^{c} = E_{b\,0}^{c} \qquad b \in \mathcal{B}, t = 24, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W} \tag{4.47}$$

$$\underline{F}_l \leq x_{l,t,w}^{c} \leq \overline{F}_l \qquad l \in \mathcal{L}_E, 1 \leq t \leq 24, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W} \tag{4.48}$$

$$\theta_{l,y}\, \underline{F}_l \leq x_{l,t,w}^{c} \leq \theta_{l,y}\, \overline{F}_l \qquad l \in \mathcal{L}_C, 1 \leq t \leq 24, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W} \tag{4.49}$$

$$\sum_{k \epsilon \Omega_z^k} \left(\underline{P}_k\, \gamma_{k,t,w}^{c} + p_{k,t,w}^{c}\right) + \mu_{z,t}^{c} \left(sol_{z,0} + \sum_{i=1}^{y} sol_{z,i} \right) + \rho_{z,t}^{c} \left(wind_{z,0} + \sum_{i=1}^{y} wind_{z,i} \right)$$

$$+ \sum_{l|rz(l)=z} x_{l,t,w}^{c} + \sum_{h \epsilon \Omega_z^h} E_{h,t,w}^{\mathrm{OUT},c} + \sum_{b \epsilon \Omega_z^b} E_{b,t,w}^{\mathrm{OUT},c} + ENP_{z,t,w}^{c} =$$

$$= D_{z,t,w}^{c} + \sum_{l|sz(l)=z} x_{l,t,w}^{c} + \sum_{h \epsilon \Omega_z^h} E_{h,t,w}^{\mathrm{IN},c} + \sum_{b \epsilon \Omega_z^b} E_{b,t,w}^{\mathrm{IN},c} + OG_{z,t,w}^{c}$$
$$z \in \mathcal{Z}, 1 \leq t \leq 24, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W} \tag{4.50}$$

$$\sum_{k\epsilon\Omega_z}\left[\left(\overline{P}_k - \underline{P}_k\right)\gamma^c_{k,t,w} - p^c_{k,t,w}\right] \geq R^c_{z,t}$$

$$z \in \mathcal{Z}, 1 \leq t \leq 24, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W} \qquad (4.51)$$

$$\sum_{z|ma(z)=m}\ \sum_{k\in\Phi_{z,f}}\ \sum_{c\in\mathcal{C}^y} wg_c \sum_{t=1}^{24}\frac{HR_k(\underline{P}_k\,\gamma^c_{k,t,w} + p^c_{k,t,w})}{ECnt_f} \leq \overline{FA}_{f,m,y}$$

$$f \in \mathcal{F}, m \in \mathcal{M}, y \in \mathcal{Y}, w \in \mathcal{W} \qquad (4.52)$$

$$\sum_{z|ma(z)=m}\ \sum_{f\in\mathcal{F}}\sum_{k\in\Phi_{z,f}}\sum_{c\in\mathcal{C}^y} wg_c \sum_{t=1}^{24} HR_k(\underline{P}_k\,\gamma^c_{k,t,w} + p^c_{k,t,w})\,co_{2\,f} \leq \overline{CO}_{2\,m,y}$$

$$m \in \mathcal{M}, y \in \mathcal{Y}, w \in \mathcal{W} \qquad (4.53)$$

$$RES^c_{z,t,w} = \mu^c_{z,t}\left(sol_{z,0} + \sum_{i=1}^{y} sol_{z,i}\right) + \rho^c_{z,t}\left(wind_{z,0} + \sum_{i=1}^{y} wind_{z,i}\right) + \sum_{h\epsilon\Omega^h_z} E^{\mathrm{OUT},c}_{h,t,w}$$

$$z \in \mathcal{Z}, 1 \leq t \leq 24, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W} \qquad (4.54)$$

$$\sum_{z|ma(z)=m}\sum_{c\in\mathcal{C}^y} wg_c \sum_{t=1}^{24} RES^c_{z,t,w} \geq \varphi_{m,y}\left(\sum_{z|ma(z)=m}\sum_{c\in\mathcal{C}^y} wg_c \sum_{t=1}^{24} D^c_{z,t}\right)$$

$$m \in \mathcal{M}, y \in \mathcal{Y}, w \in \mathcal{W} \qquad (4.55)$$

$$\delta^-_{k,y}, \theta^-_{k,y} \in \{0,1\} \qquad k \in \mathcal{K}_E, y \in \mathcal{Y} \qquad (4.56)$$

$$\delta^+_{k,y}, \theta^+_{k,y} \in \{0,1\} \qquad k \in \mathcal{K}_C, y \in \mathcal{Y} \qquad (4.57)$$

$$\delta_{h,y}, \theta_{h,y} \in \{0,1\} \qquad h \in \mathcal{H}_C, y \in \mathcal{Y} \qquad (4.58)$$

$$\delta_{l,y}, \theta_{l,y} \in \{0,1\} \qquad l \in \mathcal{L}_C, y \in \mathcal{Y} \qquad (4.59)$$

$$sol_{z,y}, wind_{z,y} \geq 0 \qquad z \in \mathcal{Z}, y \in \mathcal{Y} \qquad (4.60)$$

$$cap^{\mathrm{Batt}}_{b,y} \geq 0 \qquad b \in \mathcal{B}, y \in \mathcal{Y} \qquad (4.61)$$

$$\gamma^c_{k,t,w}, \alpha^c_{k,t,w}, \beta^c_{k,t,w} \in \{0,1\} \qquad k \in \mathcal{K}, 1 \leq t \leq 24, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W} \qquad (4.62)$$

$$p^c_{k,t,w} \geq 0 \qquad k \in \mathcal{K}, 1 \leq t \leq 24, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W} \qquad (4.63)$$

$$E^{\mathrm{IN},c}_{h,t,w}, E^{\mathrm{OUT},c}_{h,t,w}, sl^{c}_{h,t,w} \geq 0 \qquad h \in \mathcal{H}, 1 \leq t \leq 24, c \in \mathcal{C}^{y}, y \in \mathcal{Y}, w \in \mathcal{W} \qquad (4.64)$$

$$\hat{E}^{\mathrm{LT},d}_{h,w} \geq 0 \qquad h \in \mathcal{H}, d \in \mathcal{D}^{y}_{M}, y \in \mathcal{Y}, w \in \mathcal{W} \qquad (4.65)$$

$$E^{\mathrm{IN},c}_{b,t,w}, E^{\mathrm{OUT},c}_{b,t,w}, E^{c}_{b,t,w} \geq 0 \qquad b \in \mathcal{B}, 1 \leq t \leq 24, c \in \mathcal{C}^{y}, y \in \mathcal{Y}, w \in \mathcal{W} \qquad (4.66)$$

$$x^{c}_{l,t,w} \text{ free variable} \qquad l \in \mathcal{L}, 1 \leq t \leq 24, c \in \mathcal{C}^{y}, y \in \mathcal{Y}, w \in \mathcal{W} \qquad (4.67)$$

$$ENP^{c}_{z,t,w}, OG^{c}_{z,t,w}, RES^{c}_{z,t,w} \geq 0 \qquad z \in \mathcal{Z}, 1 \leq t \leq 24, c \in \mathcal{C}^{y}, y \in \mathcal{Y}, w \in \mathcal{W}. \qquad (4.68)$$

The objective function (4.1) minimizes total system costs, computed as the sum of first-stage costs and the expected value of second-stage costs. Specifically, first-stage costs include decommissioning costs of existing thermal power plants, investment costs in new generation, transmission and storage facilities, and fixed costs for the available thermal power plants. Instead, second-stage costs are the operational costs, including thermal production costs, start-up costs, hydro and batteries operational costs and penalties for energy not provided and over-generation.

Production costs are supposed to be linear functions of the power output, being $CM_{k,y,w}$ the slopes of these linear relationships. The marginal cost of thermal plant k in year y under scenario w is computed as:

$$CM_{k,y,w} = O\&M_{k} + HR_{k}(Pr^{fuel(k)}_{y,w} + co_{2_{fuel(k)}} Pr^{\mathrm{CO2}}_{y,w}) \qquad k \in \mathcal{K}, y \in \mathcal{Y}, w \in \mathcal{W}. \qquad (4.69)$$

Constraints of the model can be divided into two groups, namely first-stage constraints (4.2)–(4.23) and second-stage constraints (4.24)–(4.55). Specifically, first-stage constraints do not depend on the scenario realization and they control investment decisions as in the deterministic model by considering the different project priorities and the existence of logical relations between some investment decisions. Instead, second-stage constraints are imposed for each scenario and they model the technical conditions for operating thermal and hydropower plants, power transmission and storages, considering the flexibility provided to the energy system by the hydro-thermal dispatch and the storage units. We refer the reader to Section 3.2.3 for a detailed description of these equations. Finally, constraints (4.56)–(4.61) and (4.62)–(4.68) define first-stage and second-stage variables, respectively.

4.4 Solution algorithm

Given the long-term planning horizon and the high level of temporal and technical detail, the proposed two-stage stochastic programming model results computationally intractable even for a small number of scenarios. To obtain a solution, in this work, we apply a multi-cut Benders decomposition algorithm. Specifically, Benders decomposition is a method introduced in the 1960s [73] that allows solving a linear programming problem with complicating variables in a distributed manner at the cost of iterations [74]. In recent years, this algorithm has been widely applied to two-stage stochastic programming models [75, 76, 77]. Indeed, given their particular structure, two-stage stochastic programming models are suited for Benders decomposition application, being the first-stage variables the complicating variables: fixing the first-stage variables, the stochastic model decomposes into a set of independent and easy to solve subproblems. In the literature, there are several examples of power systems planning models solved through Benders decomposition. Pioneering work in this area is due to Bloom [78], who in 1983 proposed the application of generalized Benders' decomposition in a model for planning least-cost investments in electricity generating capacity subject to probabilistic reliability constraints. Since then, many relevant contributions have been produced. For instance, in [79] an enhanced Benders decomposition algorithm for two-stage stochastic linear problems is presented and applied to a large-scale dynamic generation and transmission expansion planning model for the European power system. In [23], authors implement a Benders decomposition algorithm to solve a two-stage stochastic generation expansion model, whose first stage determines the long-term expansion and short-term unit commitment decisions, while the second stage models the real-time operation. In [80] a Benders decomposition algorithm is used to solve a network-constrained AC unit commitment problem under uncertainty. Finally, references [81] and [82] apply Benders decomposition algorithm to solve large-scale transmission expansion planning problems.

In our model, the first-stage variables represent investment and decommissioning decisions and include $\theta^{-}_{k,y}$, $\theta^{+}_{k,y}$, $\theta_{h,y}$, $\theta_{l,y}$, $sol_{z,y}$, $wind_{z,y}$, and $cap^{\mathrm{Batt}}_{b,y}$. If these variables are fixed, the original problem decomposes into a set of independent subproblems, one per year and scenario, each representing the operation in the second stage. Benders decomposition replaces the two-stage stochastic problem with an iterative collection of smaller problems. At each iteration, the so-called master problem is solved first to determine suitable values for the first-stage variables. Once the investment schedule is

determined, the subproblems are solved. The number of these subproblems equals the number of years of the planning horizon times the number of scenarios. Finally, the dual information of the subproblems is sent to the master problem employing a cut to update the master problem solution. The next paragraphs provide the formulation of the master problem and the subproblems, as well as a more detailed description of the implemented algorithm.

4.4.1 Master problem

As previously mentioned, the master problem aims to provide values of the first-stage variables by solving at each iteration j the following MILP model, whose optimization variables are $\delta^{-}_{k,y}$, $\theta^{-}_{k,y}$, $\delta^{+}_{k,y}$, $\theta^{+}_{k,y}$, $sol_{z,y}$, $wind_{z,y}$, $\delta_{h,y}$, $\theta_{h,y}$, $cap^{\text{Batt}}_{b,y}$, $\delta_{l,y}$, $\theta_{l,y}$, and σ_w.

$$
\begin{aligned}
\min\ z_{down} \quad = \quad & \sum_{y\in\mathcal{Y}}\left(\sum_{k\in\mathcal{K}_E}\frac{DC_k\delta^{-}_{k,y}}{(1+r)^{y-y_0}}+\sum_{k\in\mathcal{K}_C}\frac{IC^{\text{The}}_{k}\delta^{+}_{k,y}}{(1+r)^{y-y_0}}\right)+\\
+ \quad & \sum_{y\in\mathcal{Y}}\left(\sum_{z\in\mathcal{Z}}\frac{IC^{\text{Sol}}_{z,y}\,sol_{z,y}}{(1+r)^{y-y_0}}+\sum_{z\in\mathcal{Z}}\frac{IC^{\text{Wind}}_{z,y}\,wind_{z,y}}{(1+r)^{y-y_0}}\right)+\\
+ \quad & \sum_{y\in\mathcal{Y}}\sum_{h\in\mathcal{H}_C}\frac{IC^{\text{Hyd}}_{h,y}\,\delta_{h,y}}{(1+r)^{y-y_0}}+\\
+ \quad & \sum_{y\in\mathcal{Y}}\sum_{b\in\mathcal{B}}\frac{IC^{\text{Batt}}_{b,y}\,cap^{\text{Batt}}_{b,y}}{(1+r)^{y-y_0}}+\\
+ \quad & \sum_{y\in\mathcal{Y}}\sum_{l\in\mathcal{L}_C}\frac{IC^{\text{Line}}_{l}\,\delta_{l,y}}{(1+r)^{y-y_0}}+\\
+ \quad & \sum_{y\in\mathcal{Y}}\left(\sum_{k\in\mathcal{K}_C}FC_k\cdot\theta^{+}_{k,y}+\sum_{k\in\mathcal{K}_E}FC_k\left(1-\theta^{-}_{k,y}\right)\right)+\\
+ \quad & \sum_{w\in\mathcal{W}}prob_w\,\sigma_w
\end{aligned}
\tag{4.70}
$$

subject to

$$
\begin{aligned}
\sigma_w \geq \sum_{y\in\mathcal{Y}}z^{(\nu)}_{y,w} \quad & + \sum_{y\in\mathcal{Y}}\sum_{k\in\mathcal{K}_E}\lambda^{\theta^{-}(\nu)}_{k,y,w}\left(\theta^{-}_{k,y}-\theta^{-\,(\nu)}_{k,y}\right)\\
& + \sum_{y\in\mathcal{Y}}\sum_{k\in\mathcal{K}_C}\lambda^{\theta^{+}(\nu)}_{k,y,w}\left(\theta^{+}_{k,y}-\theta^{+\,(\nu)}_{k,y}\right)
\end{aligned}
$$

$$+\sum_{y\in\mathcal{Y}}\sum_{z\in\mathcal{Z}}\lambda_{z,y,w}^{sol^{(v)}}\left(sol_{z,y}-sol_{z,y}^{(v)}\right)$$
$$+\sum_{y\in\mathcal{Y}}\sum_{z\in\mathcal{Z}}\lambda_{z,y,w}^{wind^{(v)}}\left(wind_{z,y}-wind_{z,y}^{(v)}\right)$$
$$+\sum_{y\in\mathcal{Y}}\sum_{h\in\mathcal{H}_C}\lambda_{h,y,w}^{\theta_h{}^{(v)}}\left(\theta_{h,y}-\theta_{h,y}^{(v)}\right)$$
$$+\sum_{y\in\mathcal{Y}}\sum_{b\in\mathcal{B}}\lambda_{b,y,w}^{cap^{\mathrm{Batt}(v)}}\left(cap_{b,y}^{\mathrm{Batt}}-cap_{b,y}^{\mathrm{Batt}(v)}\right)$$
$$+\sum_{y\in\mathcal{Y}}\sum_{l\in\mathcal{L}_C}\lambda_{l,y,w}^{\theta_l{}^{(v)}}\left(\theta_{l,y}-\theta_{l,y}^{(v)}\right)$$

$$w\in\mathcal{W}, v=1,\dots,j-1 \qquad (4.71)$$

$$\sigma_w \geq \sigma^{down} \qquad w\in\mathcal{W} \qquad (4.72)$$

(4.2)–(4.23) (4.73)

$$\sum_{z|ma(z)=m}\sum_{c\in\mathcal{C}^y} wg_c \sum_{t=1}^{24}\left[\mu_{z,t}^c\left(sol_{z,0}+\sum_{i=1}^{y}sol_{z,i}\right)+\rho_{z,t}^c\left(wind_{z,0}+\sum_{i=1}^{y}wind_{z,i}\right)\right]$$
$$\geq \varphi'_{m,y}\left(\sum_{z|ma(z)=m}\sum_{c\in\mathcal{C}^y} wg_c\sum_{t=1}^{24}D_{z,t}^c\right)$$
$$m\in\mathcal{M}, y\in\mathcal{Y} \qquad (4.74)$$

(4.56)–(4.61) (4.75)

The objective function (4.70) includes investment, decommissioning and fixed costs and the auxiliary variables σ_w approximating the operation cost under scenario w. The solution of the master problem represents a lower bound for the optimal objective function value of the original problem, since the master problem is a relaxation of the original problem. Indeed, auxiliary variables σ_w lower approximate the second-stage costs as shown in [74], chapter 3.3. At each iteration j, once the master problem is solved, the optimal values of the objective function and auxiliary variables σ_w are stored in the vectors $z_{down}^{(j)}$ and $\sigma_w^{(j)}$, respectively. Besides, the optimal values of the first-stage variables are stored in the parameters $\theta_{k,y}^{-(j)}$, $\theta_{k,y}^{+(j)}$, $sol_{z,y}^{(j)}$, $wind_{z,y}^{(j)}$, $\theta_{h,y}^{(j)}$, $cap_{b,y}^{\mathrm{Batt}(j)}$, and $\theta_{l,y}^{(j)}$. These parameters will be used to build constraints (4.71) in subsequent iterations,

along with the optimal value of subproblems objective function $z_{y,w}^{(j)}$ (4.76) and dual variable vectors $\lambda_{y,w}^{(j)}$, obtained from fixing constraints (4.79)–(4.85) in the subproblems at iteration j, as described in the next paragraph. Specifically, the constraints (4.71), referred to as Benders optimality cuts, tighten the feasible region of the master problem over iterations. While in the original Benders decomposition algorithm, a single cut is generated at each iteration [73], in our approach we implement a multi-cut strategy, generating at each iteration one cut per scenario. As observed in [23], [83] and [84], also in our application, the multi-cut Benders decomposition showed a faster convergence than the mono-cut algorithm.

Lower bound constraints (4.72) on variables σ_w avoid the master problem being unbounded in the first iteration, while constraints (4.73) control investment and decommissioning decisions as in the original problem. To both accelerate the convergence speed to the optimal solution and ensure the feasibility of subproblems, new constraints on the production from wind and solar power plants (4.74) are imposed. Indeed, the achievement of the challenging renewable penetration targets mainly requires installing large shares of new solar and wind power capacity, while the possibility to expand the hydropower production is generally limited. For instance, in the Italian power system the hydropower generation is considered as a mature technology that has already reached its technological limit: the production from hydropower plants could only marginally increase in the future. Thus, we introduce a new parameter $\varphi'_{m,y}$, representing the lower bound for solar and wind power penetration in macro-area m in year y, and we impose constraints (4.74) that can be included in the master problem, since wind and solar power production do not depend on the scenario realization for prices, as opposed to the hydro generation, which is a second-stage variable. In this way, we force the model to install new wind and solar power capacity from the first iterations, increasing the convergence speed. Finally, constraints (4.75) define optimization variables for the master problem.

It is worth mentioning that in our analysis the master problem contains only optimality cuts, while Benders feasibility cuts are not included. Indeed, due to the second-stage variables $ENP_{z,t,w}^{c}$, $OG_{z,t,w}^{c}$ and $RNP_{z,t,w}^{c}$ that model energy not provided, over-generation (i.e., energy in excess) and reserve not provided, respectively, the subproblems are always feasible.

4.4.2 Subproblems

At each iteration j, for given values of the first-stage variables $\theta_{k,y}^{+\,(j)}$, $\theta_{k,y}^{-\,(j)}$, $\theta_{h,y}^{(j)}$, $sol_{z,y}^{(j)}$, $wind_{z,y}^{(j)}$, $cap_{b,y}^{\mathrm{Batt}(j)}$, and $\theta_{l,y}^{(j)}$ the subproblem associated with year y and scenario w is formulated as follows.

$$\begin{aligned}\min\ z_{y,w} \quad = \quad & \sum_{c\in\mathcal{C}^y} wg_c \sum_{t=1}^{24}\Bigg[\sum_{k\in\mathcal{K}}\left(CM_{k,y,w}\left(\underline{P}_k\,\gamma_{k,t,w}^{c}+p_{k,t,w}^{c}\right)+SUC_k\alpha_{k,t,w}^{c}\right) \\ & \qquad +\sum_{h\in\mathcal{H}} Cvar_h\,E_{h,t,w}^{\mathrm{OUT},c}+\sum_{b\in\mathcal{B}} Cvar_b\,E_{b,t,w}^{\mathrm{OUT},c} \\ & \qquad +\sum_{z\in Z}\left(c_{ENP}ENP_{z,t,w}^{c}+c_{OG}OG_{z,t,w}^{c}+c_{RNP}RNP_{z,t,w}^{c}\right)\Bigg] \end{aligned} \tag{4.76}$$

subject to

$$(4.24)\text{–}(4.50),\ (4.52)\text{–}(4.55) \qquad 1\le t\le 24, c\in\mathcal{C}^y \tag{4.77}$$

$$\sum_{k\in\Omega_z}\left[\left(\overline{P}_k-\underline{P}_k\right)\gamma_{k,t,w}^{c}-p_{k,t,w}^{c}\right]+RNP_{z,t,w}^{c}\ge R_{z,t}^{c} \qquad z\in\mathcal{Z}, 1\le t\le 24, c\in\mathcal{C}^y \tag{4.78}$$

$$\theta_{k,y}^{+}=\theta_{k,y}^{+\,(j)} \qquad :\lambda_{k,y,w}^{\theta^+} \qquad k\in\mathcal{K}_C \tag{4.79}$$

$$\theta_{k,y}^{-}=\theta_{k,y}^{-\,(j)} \qquad :\lambda_{k,y,w}^{\theta^-} \qquad k\in\mathcal{K}_E \tag{4.80}$$

$$\theta_{l,y}=\theta_{l,y}^{(j)} \qquad :\lambda_{l,y,w}^{\theta_l} \qquad l\in\mathcal{L}_C \tag{4.81}$$

$$wind_{z,y}=wind_{z,y}^{(j)} \qquad :\lambda_{z,y,w}^{wind} \qquad z\in\mathcal{Z} \tag{4.82}$$

$$sol_{z,y}=sol_{z,y}^{(j)} \qquad :\lambda_{z,y,w}^{sol} \qquad z\in\mathcal{Z} \tag{4.83}$$

$$cap_{b,y}=cap_{b,y}^{\mathrm{Batt}(j)} \qquad :\lambda_{b,y,w}^{cap^{\mathrm{Batt}}} \qquad b\in\mathcal{B} \tag{4.84}$$

$$\theta_{h,y}=\theta_{h,y}^{(j)} \qquad :\lambda_{h,y,w}^{\theta_h} \qquad h\in\mathcal{H}_C \tag{4.85}$$

$$0\le\gamma_{k,t,w}^{c},\alpha_{k,t,w}^{c},\beta_{k,t,w}^{c}\le 1 \qquad k\in\mathcal{K}, 1\le t\le 24, c\in\mathcal{C}^y \tag{4.86}$$

$$(4.63)\text{–}(4.68) \tag{4.87}$$

$$RNP_{z,t,w}^{c}\ge 0 \qquad z\in\mathcal{Z}, 1\le t\le 24, c\in\mathcal{C}^y \tag{4.88}$$

Objective function (4.76) minimizes the operating cost of the system in year y under scenario w by considering thermal production costs, start-up costs, hydro and batteries operational costs, and penalties for energy not provided, over-generation and reserve not provided. Constraints (4.77) include all operating constraints in the original problem but the reserve constraint (4.51), which is replaced by equation (4.78) that ensures the feasibility of each subproblem by introducing the reserve not provided, which is penalized in the objective function similarly to the energy not provided. Constraints (4.77) and (4.78) are imposed for every hour t of every representative day c belonging to the considered year y. Equations (4.79)–(4.85) fix the complicating variables to values determined by the master problem, while variables $\lambda^{\theta^+}_{k,y,w}$, $\lambda^{\theta^-}_{k,y,w}$, $\lambda^{\theta_l}_{l,y,w}$, $\lambda^{wind}_{z,y,w}$, $\lambda^{sol}_{z,y,w}$, $\lambda^{cap^{\text{Batt}}}_{b,y,w}$ and $\lambda^{\theta_h}_{h,y,w}$ are the dual variables of fixing constraints. At each iteration j, the values of the objective function and dual variables of fixing constraints in the subproblem associated to year y and scenario w are stored in the parameters $z^{(j)}_{y,w}$ and $\lambda^{(j)}_{y,w}$, respectively. Both these parameters are needed to add Benders optimality cuts (4.71) to the master problem. Finally, constraints (4.86)–(4.88) define the optimization variables.

It is worth mentioning that constraints (4.86) replace the original definition (4.62) imposed on thermal commitment variables, i.e., in the subproblems the second-stage binary variables that describe thermal power plants activation patterns (i.e., $\gamma^c_{k,t,w}$, $\alpha^c_{k,t,w}$ and $\beta^c_{k,t,w}$) are relaxed to be continuous variables in the interval $[0; 1]$. Indeed, one requirement for Benders algorithm convergence is the convexity of subproblems. Thus, to obtain convex subproblems, binary unit commitment variables are relaxed to be continuous. In this way, the Benders algorithm will converge to a solution z^*_{LP}, which is optimal for the relaxed problem (with continuous and binary investment decisions and continuous operation decisions), but that not necessarily feasible for the original investment problem. For this reason, once the algorithm reaches convergence, the investment decisions are fixed and the subproblems are solved as MILP models (i.e., by considering binary unit commitment variables) so as to obtain a "quasi-optimal" solution for the original problem z_{MILP}. The solution obtained with this procedure is therefore feasible, but not necessarily optimal (i.e., $z^*_{LP} \leq z^*_{MILP} \leq z_{MILP}$). However, in this application, as common in the literature when solving real-scale power systems, it is impossible to solve the expansion planning problem up to optimality [23, 30, 44]. We consider a optimality gap tolerance of 0.1% (i.e., $\frac{z_{MILP}-z^*_{LP}}{z^*_{LP}} \leq 0.001$) as done in analogous studies, see [30, 44]. Empirical results show how the two solutions z^*_{LP} and z_{MILP} are very

close, being the relative distance in the tests lower than 0.05% and, thus, much smaller than the reasonable optimality gap tolerance. Indeed, in our approach, we are modelling thermal unit commitment decisions by using the equations described in [29] and [30], which tighten the original problem's feasible region by reducing the distance between relaxed and integer solutions.

The solution of all subproblems allows computing the following upper bound to the optimal objective function value of the relaxed problem (with continuous and binary investment decisions and continuous operation decisions) at iteration j

$$z_{up}^{(j)} = z_{down}^{(j)} - \sum_{w \in \mathcal{W}} prob_w \cdot \sigma_w^{(j)} + \sum_{w \in \mathcal{W}} prob_w \sum_{y \in \mathcal{Y}} z_{y,w}^{(j)}. \tag{4.89}$$

Indeed, parameter $z_{up}^{(j)}$ in equation (4.89) is equal to the sum of investment, decommissioning and fixed costs associated with the master problem's solution at iteration j (i.e., $z_{down}^{(j)} - \sum_{w \in \mathcal{W}} prob_w \cdot \sigma_w^{(j)}$) and the expected value of operational costs associated with the subproblems solution at iteration j (i.e., $\sum_{w \in \mathcal{W}} prob_w \sum_{y \in \mathcal{Y}} z_{y,w}^{(j)}$). Since at each iteration the values of the first-stage and the second-stage variables are determined by solving two independent problems, $z_{up}^{(j)}$ represents an upper bound to the optimal objective function value of the relaxed problem.

4.4.3 Steps of the solution algorithm

Given a small tolerance value ε to control convergence, the Benders decomposition works as follows:

0. **Initialization.** Initialize the iteration counter, set $j = 1$. Set $z_{up}^{(j)} = \infty$ and $z_{down}^{(j)} = -\infty$.
1. **Master problem solution**. Solve the master problem (4.70)–(4.75). Update $z_{down}^{(j)}$ and the values of first-stage variables.
2. **First year**. Consider the first year of the planning horizon, i.e., $y = 1$.
3. **First scenario**. Consider the first scenario, i.e., $w = 1$.
4. **Subproblem solution**. Solve subproblem (4.76)–(4.88) for year y and scenario w. Compute $z_{y,w}^{(j)}$ and store the dual variables of the fixing constraints (4.79)–(4.85).
5. **Scenario update**. Consider the next scenario and repeat step 4. If all scenarios have been considered go to step 6.

6. **Year update**. Consider the next year of the planning horizon and repeat steps from 3 to 5. If all years have been considered go to step 7.
7. **Convergence checking**. Compute $z_{up}^{(j)}$. If $\frac{\left|z_{up}^{(j)} - z_{down}^{(j)}\right|}{z_{up}^{(j)}} < \varepsilon$, the optimal solution has been obtained, go to step 8. Otherwise, update the iteration counter, set $j = j + 1$ and go back to step 1.
8. **Subproblems final integer solution**. For each year y and each scenario w, solve subproblems (4.76)–(4.88) replacing constraints (4.86) with (4.62), i.e., considering MILP problems. The solution obtained is now feasible for the original problem.

4.4.4 Numerical tests

To evaluate the performances of the proposed algorithm, several tests have been conducted on a small scenario consisting in three market zones interconnected in a tree network by two transmission lines. The scenario includes 66 thermal power plants, divided into 48 existing facilities and 18 candidate thermal units, four candidate transmission lines, three equivalent hydropower plants (one per zone), and a planning horizon of nine years, from 2017 up to 2025. It is worth mentioning that such a scenario represents a very simplified version of the Italian power system. Indeed, the objective of the tests described in this paragraph was to evaluate the performances of the proposed algorithm by comparing computational times between the monolithic stochastic model and the decomposed model for different number of scenarios. Due to computational restrictions, to keep the monolithic stochastic model computationally tractable only a simplified representation of power system can be considered.

As regards to the long-term uncertainty modeling, we considered different scenarios for fuel and CO_2 prices by randomly modifying values observed in the last year before the beginning of the planning horizon. To assess the performances of the proposed algorithm, we solved the monolithic stochastic model and the decomposed model for different number of scenarios. Specifically, we applied the algorithm summarized in Section 4.4.3 by considering $\varepsilon = 10^{-4}$ as tolerance for the convergence, while in each iteration we solved the master problem up to optimality.

Table 4.2 reports the results of our numerical experiments that we obtained on a computer with two 2.10 GHz Intel® Xeon® Platinum 8160 CPU Processors and 128 GB of RAM, using language extension GUSS [85] integrated with solver Gurobi under GAMS 24.7.4.

Table 4.2 Performances of the proposed algorithm for an increasing number of scenarios

	Number of scenarios					
	2	5	10	15	20	30
Monolithic problem solution time [min]	2.52	12.31	109.20	376.15	1341.74	Out of memory
Decomposed problem solution time [min]	6.21	8.51	9.88	10.93	12.53	15.78
Number of iterations	19	17	15	14	14	13
z_{LP}^{*} [10^9 €]	84.103	83.691	83.496	83.294	83.120	82.930
z_{MILP} [10^9 €]	84.126	83.728	83.533	83.299	83.159	82.969
Difference between relaxed and integer solution [%]	0.03%	0.04%	0.05%	0.01%	0.05%	0.05%

As can be observed, the application of the multi-cut Benders decomposition strategy dramatically reduces solution times. Specifically, the computational savings increase as the number of scenario increases. For instance, while the monolithic and the decomposed problem present similar solution times when five scenarios are considered, for 20 scenarios the computational times with Benders decomposition are reduced of a factor 100. Moreover, the monolithic problem becomes computationally intractable when 30 scenarios are generated, while the decomposed model provides the solution to the problem in about 15 minutes.

Table 4.2 also provides the number of iterations needed to reach convergence. As can be noticed, as the number of scenarios increases the number of iterations required by the algorithm to reach convergence decreases. For instance, as shown in Fig. 4.1, when 10 scenarios are considered the algorithm reaches convergence in 15 iterations, being the upper and the lower bound equal to respectively 83.496 billions and 83.488 billions, hence satisfying the predefined tolerance $\varepsilon = 10^{-4}$.

Finally, Table 4.2 reports the objective function values z_{LP}^{*} for the relaxed problem (with continuous and binary investment decisions and continuous operation decisions) and the "quasi-optimal" solution for the original problem z_{MILP} obtained once the algorithm reaches convergence by fixing the investment decisions and solving the subproblems as MILP models. As previously mentioned, empirical results show how the two solutions z_{LP}^{*} and z_{MILP} are very close, being the relative distance in the tests lower than 0.05% and, thus, much smaller than the optimality gap tolerance of 0.1% usually chosen [29, 43] when working with large-scale optimization models.

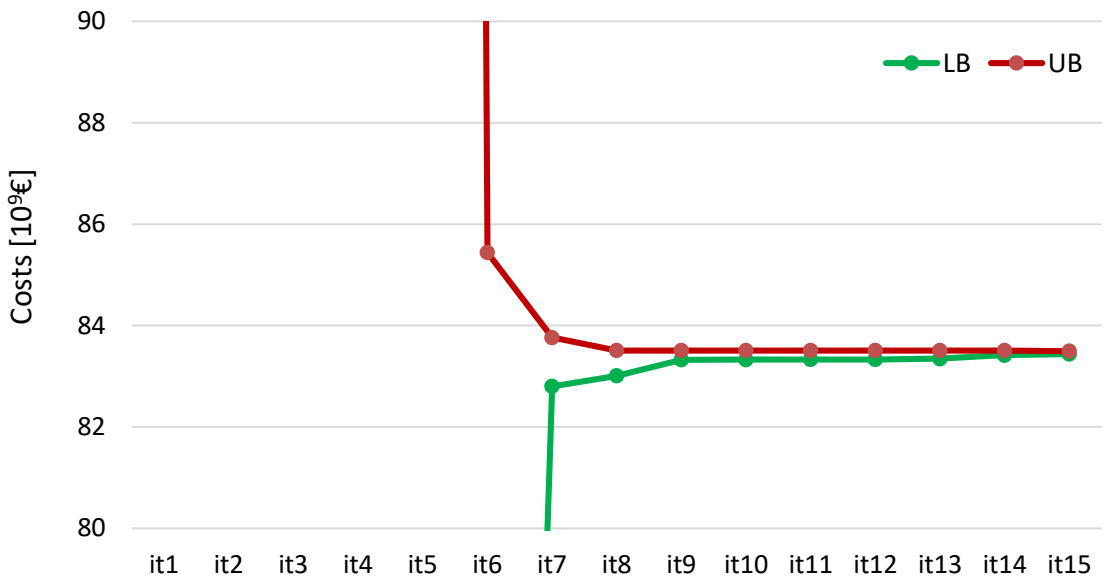


Fig. 4.1 Upper and lower bounds values over iterations in the 10 scenarios numerical test

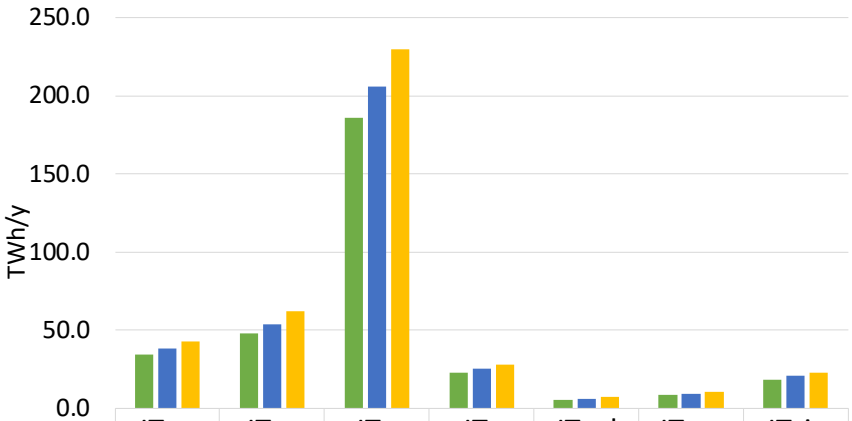
4.5 Case study

As a case study for planning the joint expansion of generation and transmission facilities under uncertainty, we chose the Italian power system. Specifically, we modified the scenario introduced in Section 3.3 by including the long-term uncertainty on prices and by considering a longer planning horizon, from 2020 up to 2040.

4.5.1 Scenario construction

In this paragraph, we describe the activities performed to adapt the Italian scenario introduced in the previous chapter to the stochastic analysis, by considering different scenarios for prices and a longer planning horizon in order to obtain more reliable expansion plans. Only the differences with respect to the scenario for deterministic GTEP analysis are discussed in this section. We refer the reader to Section 3.3.1 for a complete description of the scenario.

As described in the previous chapter, we selected five representative days to model the operations in each year of the planning horizon. Similarly to the other representative days, also representative days for the period 2030-2040 have been obtained by applying an annual average demand growth of 1%, as shown in Fig. 4.2.



	ITcn	ITcs	ITn	ITs	ITcal	ITsar	ITsic
2020	34.4	48.2	185.8	22.8	5.8	8.5	18.6
2030	38.3	54.1	206.1	25.2	6.4	9.4	20.6
2040	42.8	62.0	229.9	28.2	7.1	10.7	23.1

Fig. 4.2 Zonal load [TWh/year] in the stochastic analysis

As far as renewable installed capacity growth is concerned, in the stochastic analysis the challenging 55% target for renewable penetration has been imposed in every year of the planning horizon from 2030 to 2040. Moreover, a lower bound of 36% has been imposed for the penetration of wind and solar power generation in 2030.

Regarding the thermal fleet, given the longer planning horizon, we included in this scenario more candidate Combined Cycle Power Plants and Open Cycle Power Plants (both fuelled by natural gas) starting operation from 2025 as shown in Table 4.3.

Regarding fuel consumption, similarly to the deterministic analysis, we have considered a CO_2 emission cap of 70 Mtons for each year of the planning horizon, according to CESI elaborations of long-term targets set for the Italian power system [86, 87], aimed at reducing the impact of the electricity sector on greenhouse gas emissions. In terms of the storage system, we have considered the same candidate storage projects of the previous scenario, whose technical data are summarized in Table 3.3. Investment cost for batteries is assumed to decrease in the period 2020-2040, as stated in [54], with an exponentially decreasing trend as shown in Fig. 4.3.

Table 4.3 Installed, outgoing and incoming capacity [GW] in the stochastic analysis

	CCGT	COAL	OIL	GT	**TOT**
Installed in 2020	37.75	8.15	1.42	2.57	**49.89**
Outgoing capacity 2020/2030	3.57	8.15	1.42	1.45	**14.59**
Incoming capacity 2025/2030	11.20	-	-	15.45	**26.65**

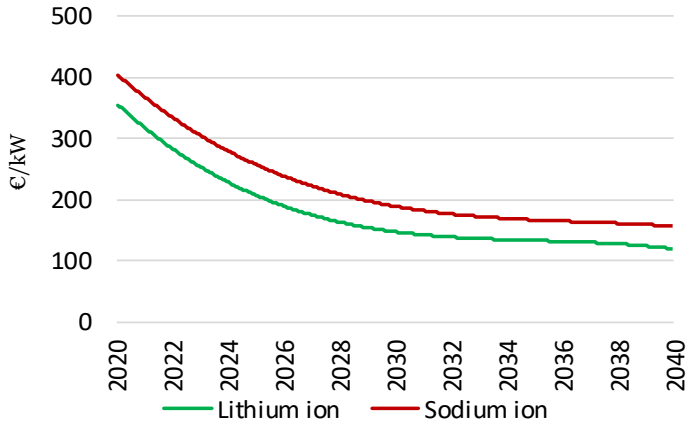


Fig. 4.3 Batteries investment cost trend

Also PV and wind investment costs are supposed to decrease along the planning horizon due to technology development. Specifically, while PV investment cost in the current analysis goes from 1000 €/kW in 2020 to 600 €/kW in 2030 to 450 €/kW in 2040, the wind investment cost is supposed to decrease from 1300 €/kW in 2020 to 900 €/kW in 2030 to 690 €/kW in 2040.

Fuel prices, together with CO_2 price, play an important role in the generation expansion plan optimization because they affect the merit order of thermal power plants and the economic viability of renewable generation. Moreover, fuel and CO_2 prices are quite volatile, depending on market fluctuations, the government's policy, and the political situation. For this reason, the stochastic analysis has been performed on prices, focusing on both the CO_2 and the fuel price. Indeed, we have used prices scenarios described in [51], which provides two scenarios for fuel prices (i.e., A and B scenario for fuel prices) and three different scenarios for emission costs (i.e., low, medium and high CO_2 prices). These scenarios are summarized in Table 4.4 and Table 4.5.

Table 4.4 CO_2 prices [€/ton] in different scenarios

Scenario	2020	2030	2040
Low	18	27	75
Medium	19	35	80
High	20	53	100

Table 4.5 Fuel prices [€/Gcal] in different scenarios

Fuel	Scenario	2020	2030	2040
Gas [€/Gcal]	A	23.43	28.881	27.63
	B	23.43	36.84	30.56
Gasoil [€/Gcal]	A	90.00	91.25	71.58
	B	90.00	85.81	102.14
Coal [€/Gcal]	A	9.50	10.47	10.47
	B	9.50	13.40	10.47
EUmix [€/Gcal]	A	19.68	29.72	39.76
	B	19.68	38.04	39.76

4.5.2 Results and discussion

We solved the proposed model on a computer with two 2.10 GHz Intel® Xeon® Platinum 8160 CPU Processors and 128 GB of RAM, using language extension GUSS integrated with solver Gurobi under GAMS 24.7.4. We considered $\varepsilon = 10^{-4}$ as tolerance for Benders decomposition convergence, while in each iteration we solved the master problem up to optimality. Fig. 4.4 illustrates the evolution over iterations of the multi-cut Benders algorithm, which converges in 41 iterations. Indeed, at iteration 41 the relative distance between upper and lower bound equals $0.95 \cdot 10^{-4}$, satisfying the predefined tolerance.

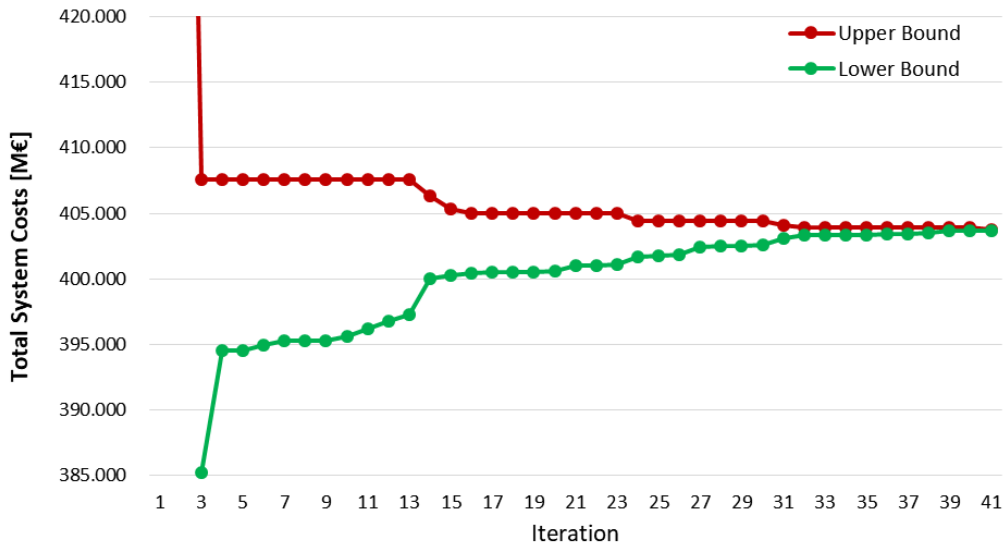


Fig. 4.4 Upper and lower bounds values over iterations in multi-cut Benders algorithm

Table 4.6 Size and solution time of master problem and subproblems at the last iteration of Benders algorithm

	# Constraints	# Decision Variables	# Discrete Variables	CPU Time [seconds]
Master Problem	9,665	12,859	5,670	27.408
First Subproblem (Base Case)	182,179	149,408	0	112.492
Updated Subproblem	182,179	149,408	0	2.622

The total time needed to solve the problem is 19,032 seconds, corresponding to 5 hours 17 minutes and 12 seconds. Table 4.6 provides more information about computational times, specifying the size and the solution time for the master problem and the subproblems at the last iteration of Benders decomposition algorithm. As can be noticed, two different types of subproblems are considered, namely the base case and the updated subproblem. Indeed, language extension GUSS works as follows. First, the base case, i.e., the model instance related to the first subproblem, is considered. After solving the base case, the update data for each subproblem is applied to the model. Then, GUSS communicates the changes from the previous model instance to the solver. This procedure not only reduces the amount of data communicated to the solver, but also, in the case of an LP model, allows the solver to restart from an advanced basis and its factorization, dramatically reducing computational times. Indeed, as can be observed in Table 4.6, while the base case solution requires almost two minutes, each of the updated subproblems is solved in only 2.622 seconds on average. Since the number of subproblems equals 126, at each iteration, the average time required to solve all the subproblems is 440.242 seconds.

Moreover, it is worth mentioning that, although in this analysis we considered a small number of scenarios, thanks to Benders decomposition and language extension GUSS, the proposed model is scalable. Specifically, Table 4.7 provides the solution time observed by considering the same tolerance for Benders convergence (i.e., $\varepsilon = 10^{-4}$) and a different number of scenarios, randomly built. As can be noticed, the stochastic model with 30 scenarios is solved in about 20 hours, a result compatible with time requirements in investment studies.

Table 4.7 Solution time for different number of scenarios

Number of scenarios	3	6	10	20	30
Solution time [h:min]	3:07	5:18	7:48	14:04	20:21

As explained in the previous section, since in each iteration subproblems are defined as linear problems, second-stage decisions $\gamma^c_{k,t,w}$, $\alpha^c_{k,t,w}$ and $\beta^c_{k,t,w}$, related to the commitment of thermal power plants, could be infeasible for the original problem. For this reason, once convergence is reached, subproblems are solved as mixed-integer linear problems, obtaining the final solution, characterized by total system costs higher than the convergence value of the Benders decomposition algorithm. However, thanks to the tight formulation of thermal unit commitment constraints, this difference is very small. Indeed, in this problem instance, the total system costs in the relaxed solution are 403,729 M€, while the integer solution is only 0.02% more expensive, with the objective function value being 403,810 M€.

Table 4.8 Breakdown of system costs in the stochastic analysis

Costs	M€	%
Thermal Capacity Expansion	2,988	0.74%
Wind Capacity Expansion	14,128	3.50%
Solar Capacity Expansion	33,832	8.38%
Transmission Capacity Expansion	1,627	0.40%
Pump Units Capacity Expansion	3,480	0.86%
Batteries Capacity Expansion	4,554	1.13%
Decommissioning Costs	211	0.05%
Thermal Fixed Costs	8	0.002%
Expected Thermal Production Cost	341,619	84.60%
Expected Start-Up Costs	399	0.10%
Expected Hydro Operation Costs	170	0.04%
Expected Batteries Operation Costs	766	0.19%
Expected Penalties for Overgeneration	28	0.01%
Expected Penalties for Energy Not Provided	0	0%
Total System Expected Costs	**403,810**	**100%**

Table 4.9 Breakdown of system operation costs for different scenarios [M€]

	Fuel Price, Scenario A			**Fuel Price, Scenario B**		
	Scenarios for CO_2 prices			**Scenarios for CO_2 prices**		
	Low	Medium	High	Low	Medium	High
Thermal Production	307,024	318,085	346,978	342,262	353,317	382,048
Start-Up	551	508	415	313	293	317
Penalties for Overgeneration	28	28	28	28	28	28
Penalties for Energy Not Provided	0	0	0	0	0	0
Batteries Operation	696	726	789	711	741	933
Hydro Operation	164	168	166	174	173	175
Total Operation	308,463	319,154	348,376	343,487	354,551	383,501
Investment	60,828	60,828	60,828	60,828	60,828	60,828
Total Cost	**369,291**	**380,342**	**409,204**	**404,315**	**415,380**	**444,329**

The system expected costs for the whole expansion planning period are shown in Table 4.8. As can be observed, there is a remarkable difference between the first-stage and second-stage costs: while the sum of investment, decommissioning and fixed costs represents 15% of total costs, operating costs account for 85% of total costs. Specifically, the most relevant cost for the system is related to the production costs of thermoelectric power plants, representing 99.6% of second-stage costs and 84.6% of total costs. On the contrary, start-up costs of thermoelectric power plants have a small impact, being 0.1% of total costs.

While investment, decommissioning and fixed costs are independent of the scenario realization, operation costs, being second-stage costs, depend on the considered scenario for CO_2 and fuel prices. Specifically, Table 4.9 describes how system costs vary depending on stochastic prices. As can be noticed, for each fuel price scenario A and B, the three CO_2 prices scenarios present similar values of start-up costs and operation costs of hydro plants, while batteries operational costs slightly differ between scenarios. As far as thermal production costs are concerned, they significantly differ in the six scenarios. As expected, since CO_2 and fuel prices affect the thermal plants marginal production costs $CM_{k,y,w}$, the higher these parameters, the greater the production costs, which vary from

307,024 M€ in the fuel price scenario A with low CO_2 prices to 382,048 M€ in the fuel price scenario B with high CO_2 prices.

Table 4.10 shows the additional capacity of wind and PV installed to reach the RES penetration target in 2040. The RES installed capacity consists of 59.2 GW of PV and 17.97 GW of wind power: this unbalance may be explained by the lower costs of the PV technology with respect to the wind technology.

Table 4.10 Renewable generation capacity expansion [GW] divided by source and implementation year in the stochastic analysis

Year	Wind	PV
2020	3.76	7.67
2021	0.76	8.31
2022	0.79	2.46
2023	0.78	2.46
2024	0.78	2.46
2025	0.78	2.46
2026	0.78	2.46
2027	0.78	2.46
2028	0.78	2.46
2029	0.78	2.46
2030	0.78	2.46
2031	0.78	2.46
2032	0.78	2.46
2033	0.78	2.46
2034	0.78	2.46
2035	0.78	2.46
2036	0.78	2.46
2037	0.78	2.46
2038	0.78	1.93
2039	0.26	1.93
Total	**17.97**	**59.20**

Moreover, as can be noticed, annual capacity increases are coincident in most of the years of the planning horizon. Indeed, the total yearly solar and wind power installed capacity is constrained by equations (4.10) and (4.11), whose right-hand sides in our numerical experiments linearly increase over time with a slope of 2.46 GW/year and 0.78 GW/year for solar and wind power technology, respectively. As regards to the solar technology, in the analysed scenario the optimal solution consists of installing all the possible capacity from year 2021 to year 2037. Since the total solar capacity installed in the Italian power system reaches its upper bound first in year 2021, yearly capacity increases from year 2022 to year 2037 are equal to 2.46 GW/year, as can be noticed from Table 4.10. Similar results can be observed also for the wind technology, with the upper bound to the total installed capacity first reached in year 2022.

As far as interconnection projects are concerned, new national and international cross border lines must be implemented in years 2025, 2029 and 2040 to better exploit the stochastic renewable energy sources and compensate for the decommissioning of some Italian thermoelectric power plants. The selected interconnections are listed in Table 4.11.

Moreover, the tool couples the installed RES capacity with energy storage systems, installing throughout the planning period 5.9 GW of batteries and 3.98 GW of pumping units. A list that summarizes the installed capacity according to technology and zone is reported in Table 4.12. As can be noticed, as regards to batteries capacity, the model suggests installing both Lithium-Ion batteries and Sodium-Ion batteries in all market zones, diversifying capacity.

Table 4.11 Candidate interconnections selected by the model in the stochastic analysis

From	To	Transmission Limits	Year of intervention
Montenegro	Central-South	[–600 MW ; 600 MW]	2025
Tunisia	Sicily	[–600 MW ; 600 MW]	2025
Central-South	Central-North	[–150 MW ; 150 MW]	2025
Central-South	Central-North	[–1000 MW ; 1000 MW]	2029
South	Central-South	[0 MW ; 900 MW]	2040
South	Central-South	[0 MW ; 200 MW]	2040

Table 4.12 Installed capacity of energy storage systems [MW] in the stochastic analysis

Zone	Technology	Installed Capacity [MW]
North	Lithium-Ion Batteries	355
North	Sodium-Ion Batteries	252
Central-North	Lithium-Ion Batteries	530
Central-North	Sodium-Ion Batteries	200
Central-South	Lithium-Ion Batteries	301
Central-South	Sodium-Ion Batteries	370
Central-South	Pumping Unit	1000
South	Lithium-Ion Batteries	600
South	Sodium-Ion Batteries	584
South	Pumping Unit	450
Calabria	Lithium-Ion Batteries	600
Calabria	Sodium-Ion Batteries	600
Calabria	Pumping Unit	1250
Sicily	Lithium-Ion Batteries	600
Sicily	Sodium-Ion Batteries	231
Sicily	Pumping Unit	480
Sardinia	Lithium-Ion Batteries	79
Sardinia	Sodium-Ion Batteries	600
Sardinia	Pumping Unit	800

In the list of thermoelectric candidate projects, ten CCGT power plants have been selected as thermoelectric expansion capacity, starting operation in 2025. On the contrary, the decommissioning of one oil and three old CCGT power plants has been planned for 2027 in Sardinia, as a result of the forecasted increase of the natural gas price and of the high RES penetration in this zone. The new thermal power plants introduced in the system are located in the North, Central-South and South zones. These new thermal power plants ensure the availability of energy reserve margins.

Fig. 4.5 reports the expected energy generation in 2040 divided by energy source, for each Italian zone.

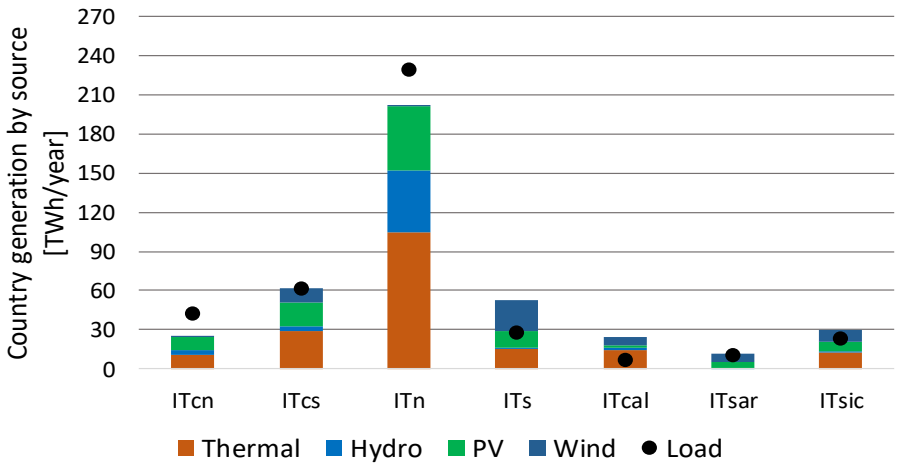
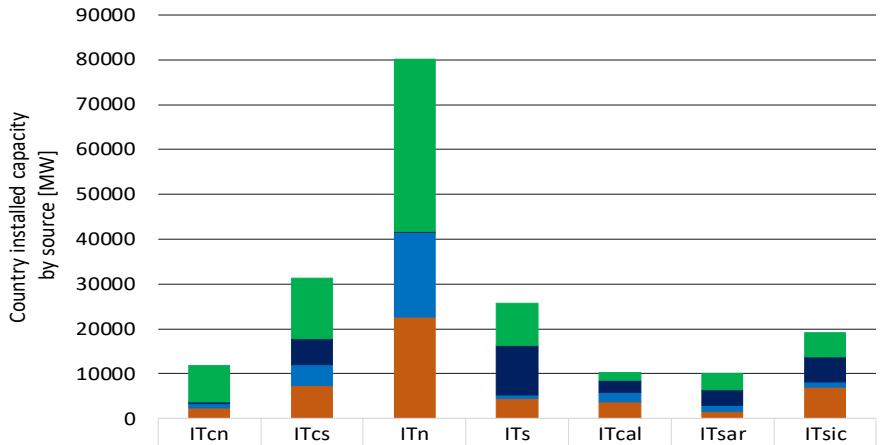


Fig. 4.5 Expected energy generation by source and domestic demand for each Italian market zone in year 2040

As can be noticed, since the RES production (i.e., from wind, solar and hydro sources) is 222 TWh while the load is 404 TWh, the target of reaching 55% of renewable penetration by 2040 has been fully achieved. Half of the total solar expansion capacity, as well as most of the thermoelectric expansion capacity, has been installed in the North, due to its high electricity demand. Even though generation exceeds demand in some market zones, in 2040 Italy will import 17 TWh from neighboring countries.

Fig. 4.6 shows the installed capacity at the end of the planning period, grouped by source and zone. As can be noticed, the model suggests installing large shares of PV capacity in all Italian market zones and especially in the North, while the wind expansion is mainly located in southern regions, which are characterized by the highest wind capacity factors.



	ITcn	ITcs	ITn	ITs	ITcal	ITsar	ITsic
PV	8156	13536	38484	9517	1577	3625	5383
Wind	384	5812	188	10839	2792	3518	5639
Hydro	996	4597	18911	795	2057	1310	1170
Thermal	2363	7433	22586	4542	3794	1600	6958

Fig. 4.6 Installed capacity by source for each market zone in year 2040

4.6 Chapter conclusions

In this chapter, the problem of planning the joint generation and transmission expansion under uncertainty has been discussed. First, a comprehensive review of the different approaches developed in the literature to deal with long-term uncertainties has been provided, comparing the different works according to some relevant features for expansion planning models.

Then, our approach to the stochastic GTEP analysis has been introduced. Specifically, in our work, we decided to focus on the uncertainty of thermal power plants production costs, because of the important role they play in the generation and transmission expansion planning by affecting the merit order of thermal power plants and the economic viability of renewable generation. We consider several scenarios for fuel prices and CO_2 price and we define capacity expansion decisions using a two-stage stochastic MILP model that aims at minimizing the sum of investment, decommissioning and fixed costs and the expected value of operational costs.

Given the long-term planning horizon and the high dimensionality of GTEP problems, the proposed stochastic model results computationally intractable even for a small number of scenarios. Thus, also a solution algorithm based on the multi-cut Benders decomposition strategy has been introduced in this chapter. Numerical tests performed on several case studies showed the huge computational savings obtained by applying the proposed algorithm, especially when a high number of scenarios is considered.

Finally, an example of stochastic GTEP analysis has been presented by planning the capacity expansion of the Italian power system in the planning horizon 2020-2040 to reach decarbonisation targets set by the European Commission. Empirical results show how solar PV technology could play a key role in achieving these policy targets, being the main technology installed in the solution provided by the model. The huge solar penetration would require installing new storage systems (pumping units, mainly in southern Italian regions, and batteries) and reinforcing the transmission network, by building new national and international cross border transmission lines, so as to better exploit the intermittent renewable energy sources and compensate for the decommissioning of some Italian thermoelectric plants.

In this chapter, the expansion planning problem has been addressed by focusing only on the power system. By exogenously modeling all sectors but the power sector, the proposed approach does not allow considering the feedback effects and the interactions

between power sector and other sectors, such as the natural gas sector. However, power systems and natural gas systems are highly interrelated: the deployment of gas-fueled thermal power plants and Power-to-Gas units increases the interdependencies between these two systems and requires the development of an integrated expansion planning tool. The inclusion of the natural gas system in the expansion planning framework and the modeling of the bi-directional energy conversion between electricity and gas is addressed in the following chapter.

Chapter 5

Integrated electricity and gas systems with bi-directional energy conversion

5.1 Research motivation

The achievement of the challenging decarbonisation targets and climate change policies requires installing large shares of intermittent renewable energy sources in power systems. In the presence of high levels of renewable penetration, power systems face great challenges to meet the demand, because of the unpredictable daily and seasonal nature of renewable generation. Thus, in modern power systems the need is growing for flexible resources that could help balancing the demand and the supply for electricity to ensure power network stability and reliability. Beside storage technologies, natural gas represents an existing option to cope with the variability of renewable power sources. Specifically, the energy conversion between electricity and natural gas is bi-directional.

First, natural gas can be converted into electricity by gas-fired power plants that can be fired up in just a few minutes, therefore being suited to compensate for the variability of renewable generation, which makes the dispatch of conventional units quite volatile over time along with short but steep ramps [23]. Given the fast-response capability and the lower emission coefficients with respect to coal thermal plants, gas-fired power plants are expected to proliferate in the next years [88]. Secondly, electricity can be transformed into gas by the Power-to-Gas (PtG) technology, which provides an opportunity to integrate large shares of renewables by converting surplus renewable power generation to gas fuel that can be stored locally to be used later or injected in the natural gas network. Researches such as [89] demonstrate how PtG could solve the challenge of long-term electricity storage and ease situations of grid congestion in scenarios with very high shares of renewables, predicting a large development of this technology. The deployment of both gas-fired power plants and PtG increases the interconnection between electricity and gas systems and requires an integrated planning framework that could accurately consider this coupling.

This chapter describes the research activities performed to integrate the gas system in the expansion planning framework to design a tool for the analysis of the real-scale integrated electricity and gas systems. The structure of the chapter is as follows. Section 5.2 reviews the existing literature for integrated electricity and gas systems. Section 5.3 describes the proposed method to plan the expansion of integrated systems. A case study concerning the Italian energy system is introduced in Section 5.4. Finally, Section 5.5 concludes the chapter.

5.2 Literature review

Several studies have been conducted in recent years to deal with integrated electricity and gas systems. They can be divided into two groups: short-term and long-term. While short-term works investigate how to optimally operate integrated power systems, long-term studies deal with the coordinated expansion of gas networks and power systems considering long-term policy goals, such as decarbonisation. Due to computational restrictions, long-term models usually present a lower level of spatial aggregation, considering power systems consisting of market zones connected by arcs representing transmission capacity between regions and therefore ignoring grid networks within each zone [3]. In this chapter, we deal with the long-term expansion planning problem. We refer the reader to [90, 91] for a detailed description of the short-term operational problem.

In the literature, most of the works concerning the long-term expansion planning of integrated systems only consider the gas dependency of the electricity system, ignoring the role of the PtG technology. For instance, reference [92] formulates the expansion co-planning problem as a mixed-integer nonlinear model whose objective is to minimize the total cost of the gas and power sectors, including operational costs and investment costs in new transmission lines, new pipelines and new gas-fired power plants, without considering the installation of new renewable capacity. Instead, authors in [93] propose a mixed-integer linear programming model considering the natural gas value chain, i.e., from the supply to end-consumers through pipelines, and the electrical systems value chain, i.e., power generation and transmission, in an integrated way. The objective of the proposed model is to determine the optimal location, technology, and installation time of new candidate facilities, as well as the optimal dispatch over a long-term planning horizon. Reference [94] determines expansion plans for an integrated electricity and gas system by computing for several candidate plans different attributes, such as the

operational and investment cost of gas system, the operational and investment cost of power system, the minimum of maximum regret and the β-robustness. Expansion plans are then determined by ranking the candidate plans according to different weights assigned to the attributes. Also in [95] a combined gas and electricity expansion planning model is developed by considering gas-fired generation plants as linkages between the two systems. The proposed model simultaneously minimizes gas and electricity operational cost and network expansion costs.

While the studies mentioned above deal with deterministic models, some works in the literature consider the long-term uncertainty in the expansion planning framework. For instance, authors in [96] present a chance constrained programing approach to minimize the investment cost of integrating new natural gas-fired generators, natural gas pipelines, compressors, and storage required to ensure desired confidence levels of meeting future stochastic power and natural gas demands. Also reference [97] considers the uncertainty related to electricity and natural gas demands and proposes a two-stage stochastic programming model to address the coordinated expansion planning of natural gas and power systems by defining new pipelines, new transmission lines, and new generation units to build so as to minimize the sum of investment cost and expected operational cost. A static stochastic programming model for integrated planning of electricity and natural gas system is also presented in [98]. Both short-term uncertainties of renewable generation and long-term uncertainties, such as load growth rates and gas prices, are included in the proposed model. Finally, authors in [99] consider the uncertainties in the net load demand and propose a multi-stage stochastic programming model for the coordinated expansion of gas and power networks. Compared with the two-stage approach, the proposed multi-stage stochastic programming model provides a better representation of long-term dynamics, including sequential investment decisions with the uncertainties gradually revealed over time. However, the increasing computational burden in multi-stage models often requires reducing the complexity of the analysis. For instance, compared with references [97] and [98], the multi-stage model proposed in [99] does not determine investment in new generating facilities, despite its relevance in the long-term planning of an integrated system.

Both the deterministic and the stochastic models so far introduced are based on a centralized approach, determining the expansion plan that is optimal for the integrated gas and electricity system as a whole. Other works in the literature propose to independently consider the two systems, while having some form of coordination between them [100]. For instance, in [101] a leader-follower approach is adopted to perform the expansion planning of the joint gas and electricity networks. Specifically,

electric system operator is the leader deciding about generation and transmission capacity expansion, while gas operator is the follower defining investment in new gas pipelines. In [102] expansion planning for gas and power systems are determined by applying an iterative process to simulate the physical and economic interactions between the two systems. In this work, we adopt the centralized approach by assuming that a single central entity determines the joint expansion plans. Different coordination mechanisms between electricity and gas systems will be investigated in a future work.

Furthermore, in all studies mentioned above, gas-fired power plants have been considered as the only linkage between electricity and natural gas systems. The bi-directional energy conversion allowed by the PtG technology has only recently appeared in the literature, with several works focused on the short-term, investigating the impacts and benefits of employing PtG in the operation of integrated systems [103, 104, 105]. Only a few papers include the PtG technology into the expansion planning framework. For instance, reference [106] proposes a long-term robust co-optimization planning model for integrated systems, including a joint N-1 and probabilistic reliability criterion. Instead, authors in [107] formulate the co-expansion planning as a bi-level programming problem with the upper-level determining expansion decisions and the lower-level optimizing the economical dispatch under the operational constraints given by the upper-level decision. Although the proposed model provides an accurate representation of gas network operation, no storage facilities are considered in the planning framework, which, in our opinion, represents a very strong limitation. Indeed, storage technologies, as well as natural gas, could provide power systems with the flexibility needed to compensate for the variability of renewable power sources [108]. Thus, both storage facilities and natural gas system should be considered in the long-term planning to accommodate large shares of renewables.

5.3 The performed analysis

The inclusion of the gas system in the decision making framework further increases the computational burden of the expansion planning model. This section describes our approach to plan the expansion of integrated electricity and gas systems, including a high level of temporal and technical detail in the decision making framework, while keeping the optimization model computationally tractable.

5.3.1 Selection of representative days

As explained in the previous chapters, when addressing the expansion planning problem, a high level of temporal detail is needed to catch the fluctuation of intermittent renewable power sources and to model the dynamics of energy storage systems. Representative days may be selected to evaluate short-term operations with an hourly resolution, while maintaining the problem computationally tractable.

When considering integrated electricity and gas systems, the iterative procedure described in Section 2.3.2 has to be modified to include also the gas demand values. Specifically, in our approach, representative days for integrated systems are selected from a set

$$D_{z,t}^{\mathrm{ELEC},d},\ \mu_{z,t}^{d},\ \rho_{z,t}^{d},\ D_{n,t}^{\mathrm{GAS},d} \qquad z \in \mathcal{Z}, n \in \mathcal{N}, 1 \le d \le 365, 1 \le t \le 24$$

where the index d refers to the set of days of the first year of the planning horizon, the load data $D_{z,t}^{\mathrm{ELEC},d}$ are forecast values for the first year of the planning horizon, $\mu_{z,t}^{d}$ and $\rho_{z,t}^{d}$ are technical production/capacity ratios for solar power production and wind power production, respectively, and $D_{n,t}^{\mathrm{GAS},d}$ are gas demand values. By performing the clustering analysis on this data set, correlations among renewable production and electricity and gas demands, as well as spatial correlations, can be taken into account.

Load values $D_{z,t}^{\mathrm{ELEC},d}$ and gas demands $D_{n,t}^{\mathrm{GAS},d}$ are then normalized by subtracting the respective minimum demands and dividing by the difference between maximum and minimum demands. In this way, all time series assume values in the range $[0; 1]$ and variations for the renewable profiles and electricity and gas demands can be properly compared. For every zone z the load duration curve $LDC_{z,\tau}$, $1 \le \tau \le 8760$ (i.e., the curve in which the original hourly load data $D_{z,t}^{d}$, $1 \le d \le 365$ and $1 \le t \le 24$, are in order of decreasing magnitude) is determined so as to be compared in the termination test with the zonal load duration curves corresponding to the representative days and their associated weights. The following iterative procedure is then applied:

1. set $k = 2$;
2. the days of the data set are partitioned in k clusters by the k-medoids algorithm;
3. the representative day c_ξ, for ξ, $1 \le \xi \le k$, is selected from the original data set as the day corresponding to the centroid of cluster ξ; the weight associated to representative day c_ξ is the number of days in cluster ξ;

4. for each system zone, determine the load duration curve corresponding to the k representative days and their associated weights and compute the mean absolute percentage error (MAPE) between the original load duration curve and the one corresponding to the current set of representative days;
5. if the system average MAPE (i.e., the average between the MAPEs of zonal load duration curves) is below a given threshold (e.g., 5%), stop, otherwise increase k by 1 and go to step 2.

Once the representative days for the first year of the planning horizon are determined, the representative days of the following years are derived by applying annual growth factors to load profiles and gas demands in the representative days.

5.3.2 Clustering of thermal power plants

When addressing the expansion planning problem for integrated systems, incorporating traditional unit commitment constraints for individual power plants would be computationally infeasible. To consider unit commitment constraints in the optimization model while maintaining the problem computationally tractable, in this work the so-called clustered unit commitment (CUC) formulation is implemented. In this formulation, similar power plants are grouped into a cluster and a single integer variable (rather than multiple binary variables) is introduced to represent the number of online units within each cluster in each time step [109]. Thus, compared to the traditional unit commitment formulation, in the CUC approach the number of variables describing commitment statuses is dramatically reduced. Similarly, in CUC formulation also the number of constraints and continuous variables representing dispatching decisions is reduced, since they are now defined for only a small number of clusters rather than the complete set of thermal power plants.

Even when working with clusters of identical power plants, the CUC formulation can provide results different from those obtained by applying traditional unit commitment constraints on a plant-by-plant level. However, as discussed in [110], empirical results for the Central Western European electricity system show that the clustering approach induces negligible errors, being the difference in the operational costs between the traditional unit commitment and the CUC formulation lower than 0.06%, while reducing computational times of a factor 80-800. Given the huge computational savings and the small deviation from the optimal results provided by the traditional unit commitment formulation, in this work the CUC formulation is adopted to model the operation of thermal power plants.

Similar thermal units are therefore grouped into clusters determined by applying the following procedure, which is repeated for every system zone. Specifically, for each zone z, we first consider a data set including all the thermal power plants located in zone z described by the following features: minimum power output, maximum power output, minimum up time, minimum down time, operative and maintenance cost, start-up cost, heat rate, and fixed cost. In this data set, all features are normalized by subtracting the minimum values and dividing by the difference between maximum and minimum values. We then apply an iterative procedure to group similar power plants in clusters with the following steps:

1. set $k = 2$;
2. thermal power plants of the data set are grouped in k clusters by the k-means algorithm;
3. the total sum of squares (TSS), i.e., the measure of global data dispersion, and the between cluster sum of squares (BSS), i.e., the measure of the separation between clusters, are computed;
4. compute the ratio between BSS and TSS to measure the clustering suitability; the higher the ratio, the higher both the similarity of objects in the same cluster and the distance between different groups, indicating a good data partition;
5. if the ratio is above a given threshold (e.g., 90%), stop, otherwise increase k by 1 and go to step 2.

For each cluster ζ, $1 \leq \zeta \leq k$, minimum power output, maximum power output, operative and maintenance cost, start-up cost, heat rate, and fixed cost are determined by averaging the features values of all the power plants grouped in cluster ζ, while minimum up time and minimum down time are computed by taking the medians, since these two features have to assume integer values.

5.3.3 Modeling framework

Investments in new facilities and operating decisions for integrated electricity and gas system are determined by solving a two-stage stochastic MILP model. This section presents the modeling assumptions as well as the notation and the mathematical formulation of the optimization model.

5.3.3.1 Modeling assumptions

In this work, expansion plans are obtained by considering a centralized approach: by assuming the perspective of a single central entity, such as the Ministry of Economic Development, we plan the joint expansion of electricity and natural gas system so as to minimize the total system costs. As regards to the electricity component, we adopt the same assumptions introduced in the previous chapters. Thus, we consider the power system consisting of a set $\mathcal{Z}$ of zones, grouped into macro-areas, with the set of macro-areas denoted by $\mathcal{M}$. The structure of the power system at the beginning of the planning horizon is described by set $\mathcal{L}_E$ of transmission lines connecting zones, set $\mathcal{H}_E$ of hydropower plants and parameters sol_{z_0} and $wind_{z_0}$ representing the solar power capacity and the wind power capacity, respectively, installed in each system zone $z \in \mathcal{Z}$. As explained in the previous section, thermal power plants are grouped into clusters, with $\mathcal{K}$ being the set of clusters of thermal plants. For each $k \in \mathcal{K}$, parameter $N0_k$ indicates the number of existing thermal power plants at the beginning of the planning horizon. The decisions to be taken concern decommissioning of existing thermal power plants as well as building of new thermal units, new transmission lines, new hydropower plants, new wind and solar power capacity, and new batteries capacity, which are supposed to have a negligible installed capacity at the beginning of the planning horizon.

Capacity decisions are defined for every year of the planning horizon, with the set of years denoted by $\mathcal{Y}$. Investments and decommissioning of thermal power plants are modeled trough two sets of integer variables describing for each cluster of thermal power plants the number of new units built and existing facilities decommissioned in every year of the planning horizon. Investments in new hydropower plants and new transmission lines are modeled by introducing sets of discrete facilities $\mathcal{H}_C$ and $\mathcal{L}_C$ and binary variables. The investments in new RES power generation and new batteries capacity are instead represented by continuous variables, since usually any capacity for batteries and wind and solar power plants can be installed, as opposed to thermal units, hydropower plants and transmission facilities that usually present specified size.

As regards to the gas component, we consider the natural gas system consisting of a set $\mathcal{N}$ of nodes. The structure of the gas system at the beginning of the planning horizon is described by set $\mathcal{PL}_E$ of pipelines connecting nodes and parameters $\overline{GAS}_n^{\mathrm{IN}}$ representing the storage capacity for natural gas in each node $n \in \mathcal{N}$. Decisions to be taken concern building of new pipelines, modeled through binary variables, and PtG capacity, modeled through continuous variables.

As regards to short-term operation modeling, we consider in every year y of the planning horizon a small set $\mathcal{C}^y$ of representative days. Similarly to the hydroelectric dispatch, inter-day balance constraints described in Section 2.4.2 are introduced also to model the long-term operation of gas storages.

Given the long-term horizon, a transportation approach is introduced to model the operation of both the transmission and the gas networks, imposing only transmission limits on power and gas flows, without including in the model voltage variables and nodal pressures. Although transportation models do not provide a perfect representation of load flows and gas flows, this choice is justified by the long-term scope and the computational burden of the expansion planning problem.

Finally, given the important role that fuel and CO_2 prices play in the expansion planning of integrated systems, different scenarios for these parameters are introduced, with the set of scenarios denoted by $\mathcal{W}$.

5.3.3.2 Notation

To describe the two-stage stochastic MILP model for the integrated system expansion planning problem, the following notation is introduced.

Sets

$\mathcal{Y}$	Set of years, indexed by y and i
$\mathcal{Z}$	Set of zones, indexed by z
$\mathcal{M}$	Set of macro-areas, indexed by m
$\mathcal{K}$	Set of clusters of thermal power plants, indexed by k
$\mathcal{K}^{\text{GAS}} \subset \mathcal{K}$	Set of clusters of thermal power plants fuelled with natural gas
$\Omega_z^k \subset \mathcal{K}$	Set of clusters of thermal power plants located in zone z
$\mathcal{L}$	Set of transmission lines, indexed by l
$\mathcal{L}_E \subset \mathcal{L}$	Set of existing transmission lines
$\mathcal{L}_C \subset \mathcal{L}$	Set of candidate transmission lines
$\mathcal{H}$	Set of hydropower plants, indexed by h
$\mathcal{H}_E \subset \mathcal{H}$	Set of existing hydropower plants
$\mathcal{H}_C \subset \mathcal{H}$	Set of candidate hydropower plants
$\Omega_z^h \subset \mathcal{H}$	Set of hydropower plants located in zone z
$\mathcal{B}$	Set of batteries, indexed by b
$\Omega_z^b \subset \mathcal{B}$	Set of batteries located in zone z
$\mathcal{N}$	Set of nodes of the gas network, indexed by n

$\mathcal{PL}$	Set of pipelines, indexed by pl
$\mathcal{PL}_E \subset \mathcal{PL}$	Set of existing pipelines
$\mathcal{PL}_C \subset \mathcal{PL}$	Set of candidate pipelines
$\mathcal{PtG}$	Set of Power-to-Gas plants, indexed by ptg
$\Omega_z^{\mathrm{PtG}} \subset \mathcal{PtG}$	Set of Power-to-Gas plants located in zone z
$\mathcal{C}^y$	Set of representative days of year y, indexed by c
$\mathcal{D}^y$	Set of all days of year y, indexed by d and d'
$\mathcal{D}_M^y \subset \mathcal{D}^y$	Set of days of year y in which the level of the long-term storage is checked
$\mathcal{T}$	Set of hours, from 1 to 24, indexed by t and τ
$\mathcal{F}$	Set of fuels, indexed by f
$\Phi_{z,f} \subset \Omega_z^k$	Set of clusters of thermal power plants located in zone z using fuel f
$\mathcal{W}$	Set of scenarios, indexed by w
$ma(z)$	Macro-area that contains zone z
$rz(l)$	Receiving-end zone of transmission line l
$sz(l)$	Sending-end zone of transmission line l
$rn(pl)$	Receiving-end node of pipeline pl
$sn(pl)$	Sending-end node of pipeline pl
$fuel(k)$	Fuel used in thermal power plants of cluster k
$\mathrm{Map}_{z,n}^{\mathrm{GAS}}$	Injective map of each zone z to a node of the gas network n
$\mathrm{Map}_{d,c}^{\mathrm{TIME}}$	Cluster index, i.e., injective map of each day d to a representative day c

Parameters

y_0	[−]	Reference year to which all investment costs are discounted
r	[−]	Annual discount rate
c_{ENP}	[€/MWh]	Penalty for energy not provided
c_{OG}	[€/MWh]	Penalty for over-generation
c_{RNP}	[€/MWh]	Penalty for reserve not provided
$prob_w$	[−]	Probability of scenario w
wg_c	[−]	Weight of cluster c
DC_k	[€]	Decommissioning cost of existing thermal power plants of cluster k

IC_k^{The}	[€]	Investment cost of candidate thermal power plants of cluster k
FC_k	[€]	Annual fixed costs of thermal power plants of cluster k
$CM_{k,y,w}$	[€/MWh]	Marginal production cost of thermal power plants of cluster k in year y in scenario w
$\underline{N}_{k,y}^{\mathrm{Dec}}$	[−]	Minimum number of thermal power plants of cluster k to be decommissioned within year y
$\overline{N}_{k,y}^{\mathrm{Dec}}$	[−]	Maximum number of thermal power plants of cluster k to be decommissioned within year y
$\underline{N}_{k,y}^{\mathrm{Inv}}$	[−]	Minimum number of thermal power plants of cluster k to be built within year y
$\overline{N}_{k,y}^{\mathrm{Inv}}$	[−]	Maximum number of thermal power plants of cluster k to be built within year y
$N0_k$	[−]	Number of existing thermal power plants of cluster k at the beginning of the planning horizon
$\underline{P}_k$	[MW]	Minimum power output of thermal power plants of cluster k
$\overline{P}_k$	[MW]	Maximum power produced by thermal plants of cluster k
SUC_k	[€]	Start-up cost of thermal power plants of cluster k
MUT_k	[h]	Minimum up time of thermal power plants of cluster k
MDT_k	[h]	Minimum down time of thermal power plants of cluster k
γ_{k,w_0}^c	[−]	Number of thermal power plants of cluster k ON at the beginning of representative day c in scenario w
gas_k	[$\mathrm{MW_{th}}$/MW]	Gas consumption of thermal power plants of cluster k
HR_k	[Gcal/MWh]	Heat rate of thermal power plants of cluster k
OM_k	[€/MWh]	Operative and maintenance cost of thermal power plants of cluster k
$IC_{z,y}^{\mathrm{Sol}}$	[€/MW]	Investment cost of new solar power capacity in zone z in year y
$sol_{z,0}$	[MW]	Solar power capacity installed in zone z at the beginning of the planning horizon
$\underline{PV}_{z,y}$	[MW]	Lower bound for solar power capacity in zone z in year y
$\overline{PV}_{z,y}$	[MW]	Upper bound for solar power capacity in zone z in year y
$IC_{z,y}^{\mathrm{Wind}}$	[€/MW]	Investment cost of new wind power capacity in zone z in year y

$wind_{z,0}$	[MW]	Wind power capacity installed in zone z at the beginning of the planning horizon
$\underline{W}_{z,y}$	[MW]	Lower bound for wind power capacity in zone z in year y
$\overline{W}_{z,y}$	[MW]	Upper bound for wind power capacity in zone z in year y
$D_{z,t}^{\mathrm{ELEC},c}$	[MW]	Load in zone z in hour t of representative day c
$R_{z,t}^{c}$	[MW]	Reserve requirement for zone z in hour t of representative day c
$\mu_{z,t}^{c}$	[MWh/MW]	Solar power capacity factor for zone z in hour t of representative day c
$\rho_{z,t}^{c}$	[MWh/MW]	Wind power capacity factor for zone z in hour t of representative day c
$IC_{h,y}^{\mathrm{Hyd}}$	[€]	Investment cost of candidate hydropower plant $h \in \mathcal{H}_C$ in year y
$\underline{\tau}_h$	[−]	Earliest date for construction of hydropower plant h
$\overline{\tau}_h$	[−]	Latest date for construction of hydropower plant h
$Cvar_h$	[€/MWh]	Operating cost of hydropower plant h
$\bar{E}_h^{\mathrm{IN}}$	[MW]	Upper bound on hydropower plant h pumping power
$\bar{E}_h^{\mathrm{OUT}}$	[MW]	Upper bound on hydropower plant h power output
$\overline{sl}_h$	[MWh]	Upper bound on energy spillage from hydropower plant h
$F_{h,t}^{c}$	[MWh]	Hourly energy inflows for hydropower plant h at time t of representative day c
λ_h^{IN}	[−]	Loss coefficient for hydro plant h pumping $\left(0 \le \lambda_h^{\mathrm{IN}} \le 1\right)$
λ_h^{OUT}	[−]	Loss coefficient for hydro plant h power generation $\left(\lambda_h^{\mathrm{OUT}} \ge 1\right)$
E_{h_0}	[MWh]	Energy content of hydropower plant h at the beginning of planning horizon
EPR_h	[h]	Maximum energy to power ratio (in hours) for hydropower plant h
M	[h]	Size of the temporal window in long-term storage constraints, set to 7 days
$IC_{b,y}^{\mathrm{Batt}}$	[€/MW]	Investment cost for battery b in year y
$Cvar_b$	[€/MWh]	Operating cost of battery b
$\overline{CAP}_b^{\mathrm{Batt}}$	[MW]	Upper bound on battery b installed capacity

$E^{c}_{b\,0}$	[MWh]	Initial energy content of battery b in cluster c
λ_b	[−]	Loss coefficient for energy stored by battery b ($0 \leq \lambda_b \leq 1$)
λ_b^{IN}	[−]	Loss coefficient for battery b charge $\left(0 \leq \lambda_b^{in} \leq 1\right)$
λ_b^{OUT}	[−]	Loss coefficient for battery b discharge ($\lambda_b^{out} \geq 1$)
$\bar{E}_b^{\mathrm{IN}}$	[MW]	Upper bound on battery b charge
$\bar{E}_b^{\mathrm{OUT}}$	[MW]	Upper bound on battery b discharge
EPR_b	[h]	Maximum energy to power ratio (in hours) for battery b
IC_l^{Line}	[€]	Investment cost of candidate transmission line $l \subset \mathcal{L}_C$
$\underline{\tau}_l$	[−]	Earliest date for construction of candidate transmission line l
$\overline{\tau}_l$	[−]	Latest date for construction of candidate transmission line l
$\underline{F}_l$	[MW]	Minimum capacity of transmission line l
$\overline{F}_l$	[MW]	Maximum capacity of transmission line l
IC_{pl}^{Pipe}	[€]	Investment cost of candidate pipeline $pl \subset \mathcal{PL}_C$
$\underline{\tau}_{pl}$	[−]	Earliest date for construction of candidate pipeline pl
$\overline{\tau}_{pl}$	[−]	Latest date for construction of candidate pipeline pl
$\underline{F}_{pl}$	[MW]	Minimum capacity of pipeline pl
$\overline{F}_{pl}$	[MW]	Maximum capacity of pipeline pl
IC_{ptg}^{PtG}	[€/$\mathrm{MW_{th}}$]	Investment cost of PtG plant ptg
$Cvar_{ptg}$	[€/$\mathrm{MWh_{th}}$]	Operating cost of PtG plant ptg
$\overline{CAP}_{ptg,y}^{\mathrm{PtG}}$	[$\mathrm{MW_{th}}$]	Upper bound on PtG plant ptg installed capacity
η_{ptg}	[$\mathrm{MW_{th}}$/MW]	Efficiency of PtG plant ptg
c_{GC}	[€/$\mathrm{MWh_{th}}$]	Penalty for gas curtailment
$C_{n,y,w}^{\mathrm{PROD}}$	[€/$\mathrm{MW_{th}}$]	Production cost of gas in node n in year y in scenario w
$\underline{GAS}_n^{\mathrm{PROD}}$	[$\mathrm{MW_{th}}$]	Lower bound on gas production in node n
$\overline{GAS}_n^{\mathrm{PROD}}$	[$\mathrm{MW_{th}}$]	Upper bound on gas production in node n
$\overline{GAS}_n^{\mathrm{IN}}$	[$\mathrm{MW_{th}}$]	Upper bound on gas storage charge in node n
$\overline{GAS}_n^{\mathrm{OUT}}$	[$\mathrm{MW_{th}}$]	Upper bound on gas storage discharge in node n
$\overline{GAS}_n^{\mathrm{LEV}}$	[$\mathrm{MWh_{th}}$]	Upper bound on gas stored in node n
G_{n_0}	[$\mathrm{MWh_{th}}$]	Energy content of gas storage in node n at the beginning of planning horizon

$D_{n,t}^{\mathrm{GAS},c}$	[$\mathrm{MW_{th}}$]	Gas demand not related to power generation in node n in hour t
co_{2_f}	[ton/Gcal]	CO_2 emission factor of fuel f
$Pr_{y,w}^{f}$	[€/Gcal]	Price of fuel f in year y in scenario w
$\overline{CO_2}_{m,y}$	[ton]	CO_2 emission limit for macro-area m in year y
$\varphi_{m,y}$	[−]	Lower bound for renewables penetration in macro-area m in year y
$Pr_{y,w}^{\mathrm{CO2}}$	[€/ton]	Emission price in year y in scenario w

Variables

1) First-stage variables

$N_{k,y}^{-}$	[$\in \mathbb{N}$]	Number of thermal power plants of cluster k decommissioned in year y
$N_{k,y}^{+}$	[$\in \mathbb{N}$]	Number of thermal power plants of cluster k built in year y
$\delta_{h,y}$	[0/1]	1: hydro power plant $h \in \mathcal{H}_C$ is built in year y; 0: otherwise
$\delta_{l,y}$	[0/1]	1: transmission line $l \in \mathcal{L}_C$ is built in year y; 0: otherwise
$\delta_{pl,y}$	[0/1]	1: pipeline $pl \in \mathcal{PL}_C$ is built in year y; 0: otherwise
$N_{k,y}^{\mathrm{TOT}}$	[$\in \mathbb{N}$]	Number of thermal power plants in cluster k available in year y
$\theta_{h,y}$	[0/1]	1: hydro power plant $h \in \mathcal{H}_C$ is built within year y; 0: otherwise
$\theta_{l,y}$	[0/1]	1: transmission line $l \in \mathcal{L}_C$ is built within year y; 0: otherwise
$\theta_{pl,y}$	[0/1]	1: pipeline $pl \in \mathcal{PL}_C$ is built within year y; 0: otherwise
$sol_{z,y}$	[MW]	New solar capacity installed in zone z in year y
$wind_{z,y}$	[MW]	New wind capacity installed in zone z in year y
$cap_{b,y}^{\mathrm{Batt}}$	[MW]	Storage capacity of battery b installed in year y
$cap_{PtG,y}^{\mathrm{PtG}}$	[$\mathrm{MW_{th}}$]	Capacity of PtG plant ptg installed in year y
$RES_{z,t}^{c}$	[MWh]	Generation from wind and solar power plants in zone z in hour t of representative day c

2) Second-stage variables

$\gamma_{k,t,w}^{c}$	[$\in \mathbb{N}$]	Number of thermal power plants of cluster k ON in hour t of representative day c in scenario w
$\alpha_{k,t,w}^{c}$	[$\in \mathbb{N}$]	Number of thermal power plants of cluster k started up in hour t of representative day c in scenario w
$\beta_{k,t,w}^{c}$	[$\in \mathbb{N}$]	Number of thermal power plants of cluster k shut down in hour t of representative day c in scenario w
$p_{k,t,w}^{c}$	[MW]	Power production of thermal power plants of cluster k in hour t above its minimum output $\underline{P}_k$ in scenario w
$E_{h,t,w}^{\mathrm{IN},c}$	[MW]	Pumping power of hydro reservoir h in hour t of representative day c in scenario w
$E_{h,t,w}^{\mathrm{OUT},c}$	[MW]	Power output of hydro reservoir h in hour t of representative day c in scenario w
$sl_{h,t,w}^{c}$	[MWh]	Energy spillage from hydro reservoir h in hour t of representative day c in scenario w
$\hat{E}_{h,w}^{\mathrm{LT},d}$	[MWh]	Energy level of hydro reservoir h at the end of day d in scenario w
$E_{b,t,w}^{\mathrm{IN},c}$	[MW]	Charge of battery b in hour t of representative day c in scenario w
$E_{b,t,w}^{\mathrm{OUT},c}$	[MW]	Discharge of battery b in hour t of representative day c in scenario w
$E_{b,t,w}^{c}$	[MWh]	Energy level of battery b in hour t of representative day c in scenario w
$x_{l,t,w}^{c}$	[MW]	Power flow on transmission line l in hour t of representative day c in scenario w
$ENP_{z,t,w}^{c}$	[MWh]	Energy not provided in zone z in hour t of representative day c in scenario w
$OG_{z,t,w}^{c}$	[MWh]	Over-generation in zone z in hour t of representative day c in scenario w
$RNP_{z,t,w}^{c}$	[MWh]	Reserve not provided in zone z in hour t of representative day c in scenario w
$GAS_{n,t,w}^{\mathrm{PROD},c}$	[$\mathrm{MW_{th}}$]	Gas production in node n in hour t of representative day c in scenario w

$GAS_{n,t,w}^{\mathrm{CURT},c}$	[$\mathrm{MW_{th}}$]	Gas curtailment in node n in hour t of representative day c in scenario w
$GAS_{n,t,w}^{\mathrm{IN},c}$	[$\mathrm{MW_{th}}$]	Gas injected in storage facilities in node n in hour t of representative day c in scenario w
$GAS_{n,t,w}^{\mathrm{OUT},c}$	[$\mathrm{MW_{th}}$]	Gas released from storage facilities in node n in hour t of representative day c in scenario w
$\widehat{GAS}_{n,w}^{\mathrm{LT},d}$	[$\mathrm{MWh_{th}}$]	Level of the gas stored in node n at the end of day d in scenario w
$GAS_{pl,t,w}^{\mathrm{FLOW},c}$	[$\mathrm{MW_{th}}$]	Gas flow on pipeline pl in hour t of representative day c in scenario w
$GAS_{ptg,t,w}^{\mathrm{PtG},c}$	[$\mathrm{MW_{th}}$]	Gas produced by PtG plant ptg in hour t of representative day c in scenario w

5.3.3.3 Mathematical formulation

The expansion planning model for integrated systems can be formulated as the following two-stage stochastic MILP problem:

$$
\begin{aligned}
\min z \;=\; & \sum_{y\in\mathcal{Y}}\left(\sum_{k\in\mathcal{K}}\frac{DC_k\,N_{k,y}^{-}}{(1+r)^{y-y_0}}+\sum_{k\in\mathcal{K}}\frac{IC_k^{\mathrm{The}}\,N_{k,y}^{+}}{(1+r)^{y-y_0}}\right)+\\
+\; & \sum_{y\in\mathcal{Y}}\left(\sum_{z\in\mathcal{Z}}\frac{IC_{z,y}^{\mathrm{Sol}}\,sol_{z,y}}{(1+r)^{y-y_0}}+\sum_{z\in\mathcal{Z}}\frac{IC_{z,y}^{\mathrm{Wind}}\,wind_{z,y}}{(1+r)^{y-y_0}}\right)+\\
+\; & \sum_{y\in\mathcal{Y}}\sum_{h\in\mathcal{H}_C}\frac{IC_{h,y}^{\mathrm{Hyd}}\,\delta_{h,y}}{(1+r)^{y-y_0}}+\\
+\; & \sum_{y\in\mathcal{Y}}\sum_{b\in\mathcal{B}}\frac{IC_{b,y}^{\mathrm{Batt}}\,cap_{b,y}^{\mathrm{Batt}}}{(1+r)^{y-y_0}}+\\
+\; & \sum_{y\in\mathcal{Y}}\sum_{l\in\mathcal{L}_C}\frac{IC_{l}^{\mathrm{Line}}\,\delta_{l,y}}{(1+r)^{y-y_0}}+\\
+\; & \sum_{y\in\mathcal{Y}}\sum_{k\in\mathcal{K}}FC_k\,N_{k,y}^{\mathrm{TOT}}+\\
+\; & \sum_{y\in\mathcal{Y}}\sum_{pl\in\mathcal{PL}_C}\frac{IC_{pl}^{\mathrm{Pipe}}\,\delta_{pl,y}}{(1+r)^{y-y_0}}+
\end{aligned}
$$

$$+ \sum_{y\in\mathcal{Y}} \sum_{ptg\in\mathcal{P}t\mathcal{G}} \frac{IC_{ptg}^{\mathrm{PtG}}\, cap_{ptg,y}^{\mathrm{PtG}}}{(1+r)^{y-y_0}} +$$

$$+ \sum_{w\in\mathcal{W}} prob_w \left[\sum_{y\in\mathcal{Y}} \sum_{c\in\mathcal{C}^y} wg_c \sum_{t=1}^{24} \left(\sum_{k\in\mathcal{K}} CM_{k,y,w} \left(\underline{P}_k\, \gamma_{k,t,w}^c + p_{k,t,w}^c \right) + \right.\right.$$

$$+ \sum_{k\in\mathcal{K}} SUC_k \alpha_{k,t,w}^c + \sum_{h\in\mathcal{H}} Cvar_h\, E_{h,t,w}^{\mathrm{OUT},c} + \sum_{b\in\mathcal{B}} Cvar_b\, E_{b,t,w}^{\mathrm{OUT},c} +$$

$$+ \sum_{n\in\mathcal{N}} C_{n,y,w}^{\mathrm{PROD}} GAS_{n,t,w}^{\mathrm{PROD},c} + \sum_{ptg\in\mathcal{P}t\mathcal{G}} Cvar_{ptg}\, GAS_{ptg,t,w}^{\mathrm{PtG},c} +$$

$$+ c_{ENP} \sum_{z\in\mathcal{Z}} ENP_{z,t,w}^c + c_{OG} \sum_{z\in\mathcal{Z}} OG_{z,t,w}^c +$$

$$\left. \left. + c_{RNP} \sum_{z\in\mathcal{Z}} RNP_{z,t,w}^c + c_{GC} \sum_{n\in\mathcal{N}} GAS_{n,t,w}^{\mathrm{CURT},c} \right) \right] \qquad (5.1)$$

subject to

$$\underline{N}_{k,y}^{\mathrm{Dec}} \le \sum_{i=1}^{y} N_{k,i}^- \le \overline{N}_{k,y}^{\mathrm{Dec}} \qquad k\in\mathcal{K}, y\in\mathcal{Y} \qquad (5.2)$$

$$\underline{N}_{k,y}^{Inv} \le \sum_{i=1}^{y} N_{k,i}^+ \le \overline{N}_{k,y}^{Inv} \qquad k\in\mathcal{K}, y\in\mathcal{Y} \qquad (5.3)$$

$$\delta_{h,y} = 0 \qquad h\in\mathcal{H}_C, y\notin\left[\underline{\tau}_h, \overline{\tau}_h\right] \qquad (5.4)$$

$$\delta_{l,y} = 0 \qquad l\in\mathcal{L}_C, y\notin\left[\underline{\tau}_l, \overline{\tau}_l\right] \qquad (5.5)$$

$$\delta_{pl,y} = 0 \qquad pl\in\mathcal{PL}_C, y\notin\left[\underline{\tau}_{pl}, \overline{\tau}_{pl}\right] \qquad (5.6)$$

$$\underline{PV}_{z,y} \le sol_{z,0} + \sum_{i=1}^{y} sol_{z,i} \le \overline{PV}_{z,y} \qquad z\in\mathcal{Z}, y\in\mathcal{Y} \qquad (5.7)$$

$$\underline{W}_{z,y} \le wind_{z,0} + \sum_{i=1}^{y} wind_{z,i} \le \overline{W}_{z,y} \qquad z\in\mathcal{Z}, y\in\mathcal{Y} \qquad (5.8)$$

$$\sum_{y\in\mathcal{Y}} cap_{b,y}^{\mathrm{Batt}} \le \overline{CAP}_b^{\mathrm{Batt}} \qquad b\in\mathcal{B} \qquad (5.9)$$

$$\sum_{y\in\mathcal{Y}} cap_{ptg,y}^{\mathrm{PtG}} \le \overline{CAP}_{ptg}^{\mathrm{PtG}} \qquad ptg\in\mathcal{P}t\mathcal{G} \qquad (5.10)$$

$$N_{k,y}^{\mathrm{TOT}} = N0_k + \sum_{i=1}^{y} \left(N_{k,i}^+ - N_{k,i}^- \right) \qquad k\in\mathcal{K}, y\in\mathcal{Y} \qquad (5.11)$$

$$\theta_{h,y} = \sum_{i=1}^{y} \delta_{h,i} \qquad h \in \mathcal{H}_C, y \in \mathcal{Y} \qquad (5.12)$$

$$\theta_{l,y} = \sum_{i=1}^{y} \delta_{l,i} \qquad l \in \mathcal{L}_C, y \in \mathcal{Y} \qquad (5.13)$$

$$\theta_{pl,y} = \sum_{i=1}^{y} \delta_{pl,i} \qquad pl \in \mathcal{PL}_C, y \in \mathcal{Y} \qquad (5.14)$$

$$\gamma_{k,t,w}^{c} \leq N_{k,y}^{\mathrm{TOT}} \qquad k \in \mathcal{K}, 1 \leq t \leq 24, c \in \mathcal{C}^{y}, y \in \mathcal{Y}, w \in \mathcal{W} \qquad (5.15)$$

$$\underline{P}_k \gamma_{k,t,w}^{c} + p_{k,t,w}^{c} \leq \overline{P}_k \gamma_{k,t,w}^{c} \qquad k \in \mathcal{K}, 1 \leq t \leq 24, c \in \mathcal{C}^{y}, y \in \mathcal{Y}, w \in \mathcal{W} \qquad (5.16)$$

$$\gamma_{k,t,w}^{c} = \gamma_{k,t-1,w}^{c} + \alpha_{k,t,w}^{c} - \beta_{k,t,w}^{c} \qquad k \in \mathcal{K}, t = 1, c \in \mathcal{C}^{y}, y \in \mathcal{Y}, w \in \mathcal{W} \qquad (5.17)$$

$$\gamma_{k,t,w}^{c} = \gamma_{k,w_0}^{c} + \alpha_{k,t,w}^{c} - \beta_{k,t,w}^{c} \qquad k \in \mathcal{K}, 2 \leq t \leq 24, c \in \mathcal{C}^{y}, y \in \mathcal{Y}, w \in \mathcal{W} \qquad (5.18)$$

$$\sum_{\tau=t-MUT_k+1}^{t} \alpha_{k,\tau,w}^{c} \leq \gamma_{k,t,w}^{c} \qquad k \in \mathcal{K}, MUT_k \leq t \leq 24, c \in \mathcal{C}^{y}, y \in \mathcal{Y}, w \in \mathcal{W} \qquad (5.19)$$

$$\sum_{\tau=t-MDT_k+1}^{t} \beta_{k,\tau,w}^{c} \leq N_{k,y}^{TOT} - \gamma_{k,t,w}^{c} \qquad k \in \mathcal{K}, MDT_k \leq t \leq 24, c \in \mathcal{C}^{y}, y \in \mathcal{Y}, w \in \mathcal{W} \qquad (5.20)$$

$$E_{h,t,w}^{\mathrm{IN},c} \leq \bar{E}_h^{\mathrm{IN}} \qquad h \in \mathcal{H}_E, 1 \leq t \leq 24, c \in \mathcal{C}^{y}, y \in \mathcal{Y}, w \in \mathcal{W} \qquad (5.21)$$

$$E_{h,t,w}^{\mathrm{OUT},c} \leq \bar{E}_h^{\mathrm{OUT}} \qquad h \in \mathcal{H}_E, 1 \leq t \leq 24, c \in \mathcal{C}^{y}, y \in \mathcal{Y}, w \in \mathcal{W} \qquad (5.22)$$

$$E_{h,t,w}^{\mathrm{IN},c} \leq \bar{E}_h^{\mathrm{IN}}\, \theta_{h,y} \qquad h \in \mathcal{H}_C, 1 \leq t \leq 24, c \in \mathcal{C}^{y}, y \in \mathcal{Y}, w \in \mathcal{W} \qquad (5.23)$$

$$E_{h,t,w}^{\mathrm{OUT},c} \leq \bar{E}_h^{\mathrm{OUT}}\, \theta_{h,y} \qquad h \in \mathcal{H}_C, 1 \leq t \leq 24, c \in \mathcal{C}^{y}, y \in \mathcal{Y}, w \in \mathcal{W} \qquad (5.24)$$

$$\frac{E_{h,t,w}^{\mathrm{IN},c}}{\bar{E}_h^{\mathrm{IN}}} + \frac{E_{h,t,w}^{\mathrm{OUT},c}}{\bar{E}_h^{\mathrm{OUT}}} \leq 1 \qquad h \in \mathcal{H}, 1 \leq t \leq 24, c \in \mathcal{C}^{y}, y \in \mathcal{Y}, w \in \mathcal{W} \qquad (5.25)$$

$$sl_{h,t,w}^{c} \leq \bar{sl}_h \qquad h \in \mathcal{H}, 1 \leq t \leq 24, c \in \mathcal{C}^{y}, y \in \mathcal{Y}, w \in \mathcal{W} \qquad (5.26)$$

$$\hat{E}_{h,w}^{\mathrm{LT},d} = E_{h_0} + \sum_{d'=d-M+1}^{d} \sum_{c \in \mathrm{Map}_{d',c}^{\mathrm{TIME}}} \sum_{t=1}^{24} \left(F_{h,t}^{c} + \lambda_h^{\mathrm{IN}}\, E_{h,t,w}^{c,\mathrm{IN}} - \lambda_h^{\mathrm{OUT}}\, E_{h,t,w}^{c,\mathrm{OUT}} - sl_{h,t,w}^{c} \right) \qquad h \in \mathcal{H}, d = M, w \in \mathcal{W} \qquad (5.27)$$

$$\hat{E}_{h,w}^{\mathrm{LT},d} = \hat{E}_{h,w}^{\mathrm{LT},d-M} + \sum_{d'=d-M+1}^{d} \sum_{c \in \mathrm{Map}_{d',c}^{\mathrm{TIME}}} \sum_{t=1}^{24} \left(F_{h,t}^{c} + \lambda_h^{\mathrm{IN}} E_{h,t,w}^{c,\mathrm{IN}} - \lambda_h^{\mathrm{OUT}} E_{h,t,w}^{c,\mathrm{OUT}} - sl_{h,t,w}^{c} \right)$$
$$h \in \mathcal{H}, d \in \mathcal{D}_M^y, y \in \mathcal{Y}, d > M, w \in \mathcal{W} \qquad (5.28)$$

$$\hat{E}_{h,w}^{\mathrm{LT},d} \leq EPR_h \, \bar{E}_h^{\mathrm{IN}} \qquad h \in \mathcal{H}, d \in D_M^y, y \in \mathcal{Y}, w \in \mathcal{W} \qquad (5.29)$$

$$\hat{E}_{h,w}^{\mathrm{LT},d} = E_{h_0} \qquad h \in \mathcal{H}, d = |\mathcal{D}^y|, y \in \mathcal{Y}, w \in \mathcal{W} \qquad (5.30)$$

$$E_{b,t,w}^{\mathrm{IN},c} \leq \sum_{i=1}^{y} cap_{b,i}^{\mathrm{Batt}} \qquad b \in \mathcal{B}, 1 \leq t \leq 24, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W} \qquad (5.31)$$

$$E_{b,t,w}^{\mathrm{OUT},c} \leq \sum_{i=1}^{y} cap_{b,i}^{\mathrm{Batt}} \qquad b \in \mathcal{B}, 1 \leq t \leq 24, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W} \qquad (5.32)$$

$$E_{b,t,w}^{c} \leq EPR_b \sum_{i=1}^{y} cap_{b,i}^{\mathrm{Batt}} \qquad b \in \mathcal{B}, 1 \leq t \leq 24, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W} \qquad (5.33)$$

$$E_{b,t,w}^{c} = (1 - \lambda_b) \, E_{b,t-1,w}^{c} + \lambda_b^{\mathrm{IN}} E_{b,t,w}^{\mathrm{IN},c} - \lambda_b^{\mathrm{OUT}} E_{b,t,w}^{\mathrm{OUT},c}$$
$$b \in \mathcal{B}, 2 \leq t \leq 24, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W} \qquad (5.34)$$

$$E_{b,t,w}^{c} = (1 - \lambda_b) \, E_{b,c_0} + \lambda_b^{\mathrm{IN}} E_{b,t,w}^{\mathrm{IN},c} - \lambda_b^{\mathrm{OUT}} E_{b,t,w}^{\mathrm{OUT},c}$$
$$b \in \mathcal{B}, t = 1, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W} \qquad (5.35)$$

$$E_{b,t,w}^{c} = E_{b\,0}^{c} \qquad b \in \mathcal{B}, t = 24, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W} \qquad (5.36)$$

$$\underline{F_l} \leq x_{l,t,w}^{c} \leq \overline{F}_l \qquad l \in \mathcal{L}_E, 1 \leq t \leq 24, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W} \qquad (5.37)$$

$$\theta_{l,y} \, \underline{F_l} \leq x_{l,t,w}^{c} \leq \theta_{l,y} \, \overline{F}_l \qquad l \in \mathcal{L}_C, 1 \leq t \leq 24, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W} \qquad (5.38)$$

$$\sum_{k \epsilon \Omega_z^k} \left(\underline{P}_k \, \gamma_{k,t,w}^{c} + p_{k,t,w}^{c} \right) + \mu_{z,t}^{c} \left(sol_{z,0} + \sum_{i=1}^{y} sol_{z,i} \right) + \rho_{z,t}^{c} \left(wind_{z,0} + \sum_{i=1}^{y} wind_{z,i} \right) +$$
$$+ \sum_{l | rz(l) = z} x_{l,t,w}^{c} + \sum_{h \epsilon \Omega_z^h} E_{h,t,w}^{\mathrm{OUT},c} + \sum_{b \epsilon \Omega_z^b} E_{b,t,w}^{\mathrm{OUT},c} + ENP_{z,t,w}^{c} = D_{z,t}^{\mathrm{ELEC},c} +$$
$$+ \sum_{l | sz(l) = z} x_{l,t,w}^{c} + \sum_{h \epsilon \Omega_z^h} E_{h,t,w}^{\mathrm{IN},c} + \sum_{b \epsilon \Omega_z^b} E_{b,t,w}^{\mathrm{IN},c} + \sum_{ptg \epsilon \Omega_z^{\mathrm{PtG}}} \frac{GAS_{ptg,t,w}^{\mathrm{PtG},c}}{\eta_{ptg}} + OG_{z,t,w}^{c}$$
$$z \in \mathcal{Z}, 1 \leq t \leq 24, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W} \qquad (5.39)$$

$$\sum_{k\in\mathcal{K}}\left[\left(\overline{P}_k - \underline{P}_k\right)\gamma^c_{k,t,w} - p^c_{k,t,w}\right] + RNP^c_{z,t,w} \geq R^c_{z,t}$$
$$z \in \mathcal{Z}, 1 \leq t \leq 24, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W} \quad (5.40)$$

$$\underline{F}_{pl} \leq GAS^{\mathrm{FLOW},c}_{pl,t,w} \leq \overline{F}_{pl} \qquad pl \in \mathcal{PL}_E, 1 \leq t \leq 24, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W} \quad (5.41)$$

$$\theta_{pl,y}\,\underline{F}_{pl} \leq GAS^{\mathrm{FLOW},c}_{pl,t,w} \leq \theta_{pl,y}\,\overline{F}_{pl}$$
$$pl \in \mathcal{PL}_C, 1 \leq t \leq 24, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W} \quad (5.42)$$

$$GAS^{\mathrm{IN},c}_{n,t,w} \leq \overline{GAS}^{\mathrm{IN}}_n \qquad n \in \mathcal{N}, 1 \leq t \leq 24, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W} \quad (5.43)$$

$$GAS^{\mathrm{OUT},c}_{n,t,w} \leq \overline{GAS}^{\mathrm{OUT}}_n \qquad n \in \mathcal{N}, 1 \leq t \leq 24, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W} \quad (5.44)$$

$$\widehat{GAS}^{\mathrm{LT},d}_{n,w} = G_{n_0} + \sum_{d'=d-M+1}^{d}\sum_{c\in \mathrm{Map}^{\mathrm{TIME}}_{d',c}}\sum_{t=1}^{24}\left(GAS^{\mathrm{IN},c}_{n,t,w} - GAS^{\mathrm{OUT},c}_{n,t,w}\right)$$
$$n \in \mathcal{N}, d = M, y \in \mathcal{Y}, w \in \mathcal{W} \quad (5.45)$$

$$\widehat{GAS}^{\mathrm{LT},d}_{n,w} = \widehat{GAS}^{\mathrm{LT},d-M}_{n,w} + \sum_{d'=d-M+1}^{d}\sum_{c\in \mathrm{Map}^{\mathrm{TIME}}_{d',c}}\sum_{t=1}^{24}\left(GAS^{\mathrm{IN},c}_{n,t,w} - GAS^{\mathrm{OUT},c}_{n,t,w}\right)$$
$$n \in \mathcal{N}, d \in D^y_M, d > M, y \in \mathcal{Y}, w \in \mathcal{W} \quad (5.46)$$

$$\widehat{GAS}^{\mathrm{LT},d}_{n,w} \leq \overline{GAS}^{\mathrm{LEV}}_n \qquad n \in \mathcal{N}, d \in D^y_M, y \in \mathcal{Y}, w \in \mathcal{W} \quad (5.47)$$

$$\widehat{GAS}^{\mathrm{LT},d}_{n,w} = G_{n_0} \qquad n \in \mathcal{N}, d = |\mathcal{D}^y|, y \in \mathcal{Y}, w \in \mathcal{W} \quad (5.48)$$

$$\underline{GAS}^{\mathrm{PROD}}_n \leq GAS^{\mathrm{PROD},c}_{n,t,w} \leq \overline{GAS}^{\mathrm{PROD}}_n$$
$$n \in \mathcal{N}, 1 \leq t \leq 24, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W} \quad (5.49)$$

$$GAS^{\mathrm{CURT},c}_{n,t,w} \leq D^{\mathrm{GAS},c}_{n,t} \qquad n \in \mathcal{N}, 1 \leq t \leq 24, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W} \quad (5.50)$$

$$GAS^{\mathrm{PtG}}_{ptg,t,w} \leq \sum_{i=1}^{y} cap^{\mathrm{PtG}}_{ptg,i} \qquad ptg \in \mathcal{PtG}, 1 \leq t \leq 24, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W} \quad (5.51)$$

$$GAS^{\mathrm{PROD},c}_{n,t,w} + \sum_{z\in Map^{GAS}_{z,n}}\sum_{ptg\in\Omega^{\mathrm{PtG}}_z} GAS^{\mathrm{PtG}}_{ptg,t,w} + \sum_{pl|rn(pl)=n} GAS^{\mathrm{FLOW},c}_{pl,t,w} +$$

$$+ GAS_{n,t,w}^{\mathrm{OUT},c} + GAS_{n,t,w}^{\mathrm{CURT},c} = \sum_{z \in Map_{z,n}^{GAS}} \sum_{k \epsilon (\mathcal{K}^{GAS} \cap \Omega_z^k)} gas_k (\underline{P}_k \, \gamma_{k,t,w}^c + p_{k,t,w}^c) +$$

$$+ D_{n,t}^{\mathrm{GAS},c} + \sum_{pl|sn(pl)=n} GAS_{pl,t,w}^{\mathrm{FLOW},c} + GAS_{n,t,w}^{\mathrm{IN},c}$$

$$n \in \mathcal{N}, 1 \le t \le 24, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W} \qquad (5.52)$$

$$\sum_{c \in \mathcal{C}^y} wg_c \sum_{t=1}^{24} \sum_{z|ma(z)=m} \sum_{k \epsilon \Omega_z^k} HR_k (\underline{P}_k \, \gamma_{k,t,w}^c + p_{k,t,w}^c) \, co_{2_{fuel(k)}} \le \overline{CO}_{2_{m,y}}$$

$$m \in \mathcal{M}, y \in \mathcal{Y}, w \in \mathcal{W} \qquad (5.53)$$

$$RES_{z,t}^c = \mu_{z,t}^c \left(sol_{z,0} + \sum_{i=1}^{y} sol_{z,i} \right) + \rho_{z,t}^c \left(wind_{z,0} + \sum_{i=1}^{y} wind_{z,i} \right)$$

$$z \in \mathcal{Z}, 1 \le t \le 24, c \in \mathcal{C}^y, y \in \mathcal{Y} \qquad (5.54)$$

$$\sum_{z|ma(z)=m} \sum_{c \in \mathcal{C}^y} wg_c \sum_{t=1}^{24} RES_{z,t}^c \ge \varphi_{m,y} \left(\sum_{z|ma(z)=m} \sum_{c \in \mathcal{C}^y} wg_c \sum_{t=1}^{24} D_{z,t}^{\mathrm{ELEC},c} \right)$$

$$m \in \mathcal{M}, y \in \mathcal{Y} \qquad (5.55)$$

$$N_{k,y}^-, N_{k,y}^+, N_{k,y}^{\mathrm{TOT}} \in \mathbb{N} \qquad k \in \mathcal{K}, y \in \mathcal{Y} \qquad (5.56)$$

$$\delta_{h,y}, \theta_{h,y} \in \{0,1\} \qquad h \in \mathcal{H}_C, y \in \mathcal{Y} \qquad (5.57)$$

$$\delta_{l,y}, \theta_{l,y} \in \{0,1\} \qquad l \in \mathcal{L}_C, y \in \mathcal{Y} \qquad (5.58)$$

$$\delta_{pl,y}, \theta_{pl,y} \in \{0,1\} \qquad pl \in \mathcal{PL}_C, y \in \mathcal{Y} \qquad (5.59)$$

$$sol_{z,y}, wind_{z,y} \ge 0 \qquad z \in \mathcal{Z}, y \in \mathcal{Y} \qquad (5.60)$$

$$cap_{b,y}^{\mathrm{Batt}} \ge 0 \qquad b \in \mathcal{B}, y \in \mathcal{Y} \qquad (5.61)$$

$$cap_{PtG,y}^{\mathrm{PtG}} \ge 0 \qquad ptg \in \mathcal{PtG}, y \in \mathcal{Y} \qquad (5.62)$$

$$RES_{z,t}^c \ge 0 \qquad z \in \mathcal{Z}, 1 \le t \le 24, c \in \mathcal{C}^y, y \in \mathcal{Y} \qquad (5.63)$$

$$\alpha_{k,t,w}^c, \beta_{k,t,w}^c, \gamma_{k,t,w}^c \in \mathbb{N} \qquad k \in \mathcal{K}, 1 \le t \le 24, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W} \qquad (5.64)$$

$$p_{k,t,w}^c \ge 0 \qquad k \in \mathcal{K}, 1 \le t \le 24, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W} \qquad (5.65)$$

$$E^{\mathrm{IN},c}_{h,t,w}, E^{\mathrm{OUT},c}_{h,t,w}, sl^{c}_{h,t,w} \geq 0 \qquad h \in \mathcal{H}, 1 \leq t \leq 24, c \in \mathcal{C}^{y}, y \in \mathcal{Y}, w \in \mathcal{W} \qquad (5.66)$$

$$\hat{E}^{\mathrm{LT},d}_{h,w} \geq 0 \qquad h \in \mathcal{H}, d \in \mathcal{D}^{y}, y \in \mathcal{Y}, w \in \mathcal{W} \qquad (5.67)$$

$$E^{c}_{b,t,w}, E^{\mathrm{IN},c}_{b,t,w}, E^{\mathrm{OUT},c}_{b,t,w} \geq 0 \qquad b \in \mathcal{B}, 1 \leq t \leq 24, c \in \mathcal{C}^{y}, y \in \mathcal{Y}, w \in \mathcal{W} \qquad (5.68)$$

$$ENP^{c}_{z,t,w}, OG^{c}_{z,t,w}, ENP^{c}_{z,t,w} \geq 0 \qquad z \in \mathcal{Z}, 1 \leq t \leq 24, c \in \mathcal{C}^{y}, y \in \mathcal{Y}, w \in \mathcal{W} \qquad (5.69)$$

$$x^{c}_{l,t,w} \text{ free variable} \qquad l \in \mathcal{L}, 1 \leq t \leq 24, c \in \mathcal{C}^{y}, y \in \mathcal{Y}, w \in \mathcal{W} \qquad (5.70)$$

$$GAS^{\mathrm{PROD},c}_{n,t,w}, GAS^{\mathrm{CURT},c}_{n,t,w}, GAS^{\mathrm{IN},c}_{n,t,w}, GAS^{\mathrm{OUT},c}_{n,t,w} \geq 0$$
$$n \in \mathcal{N}, 1 \leq t \leq 24, c \in \mathcal{C}^{y}, y \in \mathcal{Y}, w \in \mathcal{W} \qquad (5.71)$$

$$\widehat{GAS}^{\mathrm{LT},d}_{n,w} \geq 0 \qquad n \in \mathcal{N}, d \in \mathcal{D}^{y}, y \in \mathcal{Y}, w \in \mathcal{W} \qquad (5.72)$$

$$GAS^{\mathrm{PtG},c}_{ptg,t,w} \geq 0 \qquad ptg \in \mathcal{PG}, 1 \leq t \leq 24, c \in \mathcal{C}^{y}, y \in \mathcal{Y}, w \in \mathcal{W} \qquad (5.73)$$

$$GAS^{\mathrm{FLOW},c}_{pl,t,w} \text{ free variable} \qquad pl \in \mathcal{PL}, 1 \leq t \leq 24, c \in \mathcal{C}^{y}, y \in \mathcal{Y}, w \in \mathcal{W}. \qquad (5.74)$$

The proposed model co-optimizes strategic and operational decisions for integrated electricity and natural gas systems. Indeed, the objective function (5.1) comprises the nine terms below:

1. $\sum_{y\in\mathcal{Y}}\left(\sum_{k\in\mathcal{K}}\frac{DC_k\, N^{-}_{k,y}}{(1+r)^{y-y_0}}+\sum_{k\in\mathcal{K}}\frac{IC^{\mathrm{The}}_{k}\, N^{+}_{k,y}}{(1+r)^{y-y_0}}\right)$ are the annualized decommissioning costs of existing thermal power plants and investment costs in new thermal power generation;
2. $\sum_{y\in\mathcal{Y}}\left(\sum_{z\in\mathcal{Z}}\frac{IC^{\mathrm{Sol}}_{z,y}\, sol_{z,y}}{(1+r)^{y-y_0}}+\sum_{z\in\mathcal{Z}}\frac{IC^{\mathrm{Wind}}_{z,y}\, wind_{z,y}}{(1+r)^{y-y_0}}\right)$ are the annualized investment costs in new solar and wind capacity;
3. $\sum_{y\in\mathcal{Y}}\sum_{h\in\mathcal{H}_C}\frac{IC^{\mathrm{Hyd}}_{h,y}\,\delta_{h,y}}{(1+r)^{y-y_0}}$ are the annualized investment costs in new hydropower plants;
4. $\sum_{y\in\mathcal{Y}}\sum_{b\in\mathcal{B}}\frac{IC^{\mathrm{Batt}}_{b,y}\, cap^{\mathrm{Batt}}_{b,y}}{(1+r)^{y-y_0}}$ are the annualized investment costs in new batteries capacity;
5. $\sum_{y\in\mathcal{Y}}\sum_{l\in\mathcal{L}_C}\frac{IC^{\mathrm{Line}}_{l}\,\delta_{l,y}}{(1+r)^{y-y_0}}$ are the annualized investment costs in new transmission lines;
6. $\sum_{y\in\mathcal{Y}}\sum_{k\in\mathcal{K}} FC_k\, N^{\mathrm{TOT}}_{k,y}$ are the fixed costs for the available thermal power plants;

7. $\sum_{y\in\mathcal{Y}}\sum_{pl\in\mathcal{PL}_C}\frac{IC_{pl}^{\text{Pipe}}\,\delta_{pl,y}}{(1+r)^{y-y_0}}$ are the annualized investment costs in new pipelines;
8. $\sum_{y\in\mathcal{Y}}\sum_{ptg\in\mathcal{PtG}}\frac{IC_{ptg}^{\text{PtG}}\,cap_{ptg,y}^{\text{PtG}}}{(1+r)^{y-y_0}}$ are the annualized investment costs in new PtG capacity;
9. $\sum_{w\in\mathcal{W}} prob_w\left[\sum_{y\in\mathcal{Y}}\sum_{c\in\mathcal{C}^y} wg_c\ \sum_{t=1}^{24}\left(\sum_{k\in\mathcal{K}}\ CM_{k,y,w}\left(\underline{P}_k\,\gamma_{k,t,w}^c + p_{k,t,w}^c\right) + \sum_{k\in\mathcal{K}} SUC_k\cdot\right.\right.$ $\alpha_{k,t,w}^c\ +\ \sum_{h\in\mathcal{H}}\ Cvar_h\ E_{h,t,w}^{\text{OUT},c}\ +\ \sum_{b\in\mathcal{B}}\ Cvar_b\ E_{b,t,w}^{\text{OUT},c}\ +\ \sum_{n\in\mathcal{N}}\ C_{n,y,w}^{\text{PROD}}\ GAS_{n,t,w}^{\text{PROD},c}\ +$ $\sum_{ptg\,\in\,\mathcal{PtG}}\ Cvar_{ptg}\ GAS_{ptg,t,w}^{\text{PtG},c}\ +\ c_{ENP}\ \sum_{z\in\mathcal{Z}}\ ENP_{z,t,w}^c\ +\ c_{OG}\ \sum_{z\in\mathcal{Z}}\ OG_{z,t,w}^c\ +\ c_{RNP}\ \cdot$ $\left.\left.\sum_{z\in\mathcal{Z}} RNP_{z,t,w}^c + c_{GC}\sum_{n\in\mathcal{N}} GAS_{n,t,w}^{\text{CURT},c}\right)\right]$ are the second-stage costs, i.e., the operating costs, which are computed by considering for each representative day thermal power production costs, start-up costs, hydro and batteries operational costs, gas production costs, PtG operational costs, and penalties for energy not provided, over-generation and gas curtailment.

The proposed model is constrained by four groups of constraints, namely investment constraints (5.2)–(5.14), power system operational constraints (5.15)–(5.40), gas system operational constraints (5.41)–(5.52) and target constraints (5.53)–(5.55). Investment constraints regulate investment decisions in new facilities for both the power and the gas system. In particular, inequalities (5.2) impose for every cluster k lower and upper bounds to the total number of thermal power plants decommissioned within each year of the planning horizon. Constraints (5.3) work similarly by limiting the number of new built power plants. Equations (5.4) impose a temporal window for investments in new hydropower plants. Constraints (5.5) and (5.6) work similarly for the investment decisions in new transmission lines and new pipelines, respectively. Inequalities (5.7) and (5.8) bound the solar power capacity and the wind power capacity installed over the planning horizon. The total capacity installed at the end of the planning horizon for each battery is bounded by constraints (5.9), while inequalities (5.10) limit the installed capacity for PtG plants. Equations (5.11) compute for each cluster and for each year the number of thermal power plants available to supply load as the sum of the existing power plants at the beginning of the planning horizon, plus the newly built units, minus the decommissioned power plants. Equations (5.12)–(5.14) compute the values of the binary variables $\theta_{h,y}$, $\theta_{l,y}$ and $\theta_{pl,y}$ that express if investment decisions for new hydropower plants, new transmission lines, and new pipelines, respectively, have been made within every year y of the planning horizon.

Power system operational constraints (5.15)–(5.40) model the technical conditions for operating thermal and hydropower plants, power transmission and storages and consider the flexibility provided to the energy system by the hydro-thermal dispatch and

the storage units. In particular, the block of equations (5.15)–(5.20) models the thermal component of the energy system. Constraints (5.15) ensure consistency between operational and investment decisions, bounding the number of thermal power plants within each cluster that can be used to supply load to the number of available thermal power plants. Inequalities (5.16) restrict the power output from thermal units. Indeed, the total power output of thermal units of cluster k is expressed as $\underline{P}_k \gamma^c_{k,t,w} + p^c_{k,t,w}$, which is bounded below $\overline{P}_k \gamma^c_{k,t,w}$ by constraints (5.16). These constraints also set to zero the power output from clusters with no online units (i.e., $\gamma^c_{k,t,w} = 0$). Equations (5.17) and (5.18) compute the number of online thermal power plants in hour t within each cluster as the sum of online units in the previous hour, plus the number of power plants started-up in hour t, minus the number of units shut down in hour t. Parameters γ^c_{k,w_0} in equations (5.18) represent the number of units within cluster k online at the beginning of representative day c under scenario w. They are determined by summing the initial statuses assigned to all thermal power plants in cluster k according to the procedure described in Section 2.5. Inequalities (5.19) and (5.20) are the minimum up time and down time constraints.

Constraints (5.21)–(5.30) model the operation of hydropower plants. Specifically, the pumping power and the power output of the existing hydroelectric reservoirs are bounded below their respective upper limits by constraints (5.21) and (5.22). Inequalities (5.23) and (5.24) work similarly to candidate hydropower plants. Indeed, if hydro plant $h \in \mathcal{H}_C$ is built within year y, these constraints define the upper bounds to pumping power and power output, otherwise they set the corresponding variables to zero. Constraints (5.25) limit for each reservoir the power production and the pumping power in the same hour, while equations (5.26) bound the energy spillage from reservoirs. Constraints (5.27) and (5.28) model the long-term operations of hydropower plants, reconstructing the seasonal profile of reservoirs. Constraints (5.29) ensure that the long-term energy levels of reservoirs are within bounds, while equations (5.30) enforce the equality between energy levels of each reservoir h at the beginning and the end of each year of the planning horizon.

Constraints (5.31)–(5.36) model the operation of batteries. Specifically, Inequalities (5.31)–(5.33) impose upper bounds to charge, discharge and stored energy and enforce consistency between the values of investment and operational variables. Energy balances (5.34) and (5.35) are intra-day constraints, controlling the energy stored by batteries. Constraints (5.36) enforce the equality between energy levels at the beginning and the end of each representative day for each battery b.

Inequalities (5.37) and (5.38) restrict the energy flows on the existing and candidate transmission lines, respectively. The zonal balance equations (5.39) impose equality between energy sources and uses in every zone and every hour. The left-hand side of these equations represents the hourly energy sources of zone z, which are given by thermal, solar and wind generation, incoming energy flows, hydro generation, and energy released by batteries. Instead, the right-hand side describes the energy uses, which are represented by the load, outgoing energy flows, pumping power, energy absorbed by batteries, and power absorbed by PtG plant to produce synthetic gas. The variables $ENP^{c}_{z,t,w}$ and $OG^{c}_{z,t,w}$ allow detecting and evaluating problems in the simulated system that can cause a mismatch between supply and demand. Inequalities (5.40) ensure the fulfilment of zonal reserve requirements provided by online thermal power plants, with variable $RNP^{c}_{z,t,w}$ representing the reserve not provided.

Gas operational constraints (5.41)–(5.52) model the technical conditions for operating the gas system. Similarly to transmission lines, inequalities (5.41) restrict the gas flows on the existing pipelines, while constraints (5.42) impose lower and upper bounds to the gas flows on candidate pipelines. Inequalities (5.43) and (5.44) bound the gas charged and discharged from storage facilities in each node of the gas network to their respective limits. Similarly to constraints (5.27) and (5.28), equations (5.45) and (5.46) model the reconstruct the seasonal profile of the gas storage. Inequalities (5.47) limit the gas stored, while constraints (5.48) enforce the equality between storage levels at the beginning and the end of each year of the planning horizon. Inequalities (5.49) impose lower and upper bounds to the nodal gas production, while constraints (5.50) bound the gas curtailment to the gas demand. Inequalities (5.51) enforce consistency between values of operational and investment variables for PtG plants, limiting the hourly production of synthetic gas from PtG plants to the total installed capacity. The nodal balance equations (5.52) impose equality between gas sources and uses in every node and every hour. Indeed, the left-hand side of these equations represents the hourly gas sources of node n (given by nodal supply, production of synthetic gas from PtG plants, incoming flows, and storage discharge) and the right-hand side describes the gas uses (represented by gas demand related to power generation, gas demand not related to power generation, outgoing flows, and storage charge). The slack variable $GAS^{\mathrm{CURT},c}_{n,t,w}$ modeling the gas curtailment assumes positive values when the gas uses exceed the gas sources.

Target constraints (5.53)–(5.55) model conditions required to promote sustainable development of energy systems, which include a reduction of CO_2 emissions and an

increase of generation from renewable power sources. Specifically, inequalities (5.53) impose limits for thermal energy production due to CO_2 emissions. Equations (5.54) compute the zonal hourly renewable production, by considering solar and wind generation. Constraints (5.55) control the renewables penetration, forcing the total generation from solar and wind power plants in macro-area m in year y to cover at least ratio $\varphi_{m,y}$ of the total yearly load. Finally, constraints (5.56)–(5.74) define the optimization variables.

5.3.4 Solution algorithm

As explained in the previous chapters, when considering real-scale energy systems, providing a solution to the two-stage stochastic programming model is computationally infeasible. Thus, to obtain a solution, in this work, we apply the multi-cut Benders decomposition algorithm introduced in Section 4.4, which is here modified to include also the gas component.

5.3.4.1 Master problem

The master problem aims to provide values of the first-stage variables by solving at each iteration j the following MILP model, whose optimization variables are $N^{-}_{k,y}$, $N^{+}_{k,y}$, $N^{\mathrm{TOT}}_{k,y}$, $sol_{z,y}$, $wind_{z,y}$, $\delta_{h,y}$, $\theta_{h,y}$, $cap^{\mathrm{Batt}}_{b,y}$, $\delta_{l,y}$, $\theta_{l,y}$, $\delta_{pl,y}$, $\theta_{pl,y}$, $cap^{\mathrm{PtG}}_{ptg,y}$ and σ_w.

$$
\begin{aligned}
\min\ z_{down} \quad = \quad & \sum_{y\in\mathcal{Y}}\left(\sum_{k\in\mathcal{K}}\frac{DC_k N^{-}_{k,y}}{(1+r)^{y-y_0}}+\sum_{k\in\mathcal{K}}\frac{IC^{\mathrm{The}}_{k} N^{+}_{k,y}}{(1+r)^{y-y_0}}\right)+\\
+ \quad & \sum_{y\in\mathcal{Y}}\left(\sum_{z\in\mathcal{Z}}\frac{IC^{\mathrm{Sol}}_{z,y}\, sol_{z,y}}{(1+r)^{y-y_0}}+\sum_{z\in\mathcal{Z}}\frac{IC^{\mathrm{Wind}}_{z,y}\, wind_{z,y}}{(1+r)^{y-y_0}}\right)+\\
+ \quad & \sum_{y\in\mathcal{Y}}\sum_{h\in\mathcal{H}_C}\frac{IC^{\mathrm{Hyd}}_{h,y}\,\delta_{h,y}}{(1+r)^{y-y_0}}+\\
+ \quad & \sum_{y\in\mathcal{Y}}\sum_{b\in\mathcal{B}}\frac{IC^{\mathrm{Batt}}_{b,y}\, cap^{\mathrm{Batt}}_{b,y}}{(1+r)^{y-y_0}}+\\
+ \quad & \sum_{y\in\mathcal{Y}}\sum_{l\in\mathcal{L}_C}\frac{IC^{\mathrm{Line}}_{l}\,\delta_{l,y}}{(1+r)^{y-y_0}}+\\
+ \quad & \sum_{y\in\mathcal{Y}}\sum_{k\in\mathcal{K}} FC_k\, N^{\mathrm{TOT}}_{k,y}
\end{aligned}
$$

$$+ \sum_{y\in\mathcal{Y}} \sum_{pl\in\mathcal{PL}_C} \frac{IC_{pl}^{\text{Pipe}}\, \delta_{pl,y}}{(1+r)^{y-y_0}}$$
$$+ \sum_{y\in\mathcal{Y}} \sum_{ptg\in\mathcal{PtG}} \frac{IC_{ptg}^{\text{PtG}}\, cap_{ptg,y}^{\text{PtG}}}{(1+r)^{y-y_0}}$$
$$+ \sum_{w\in\mathcal{W}} prob_w\, \sigma_w \tag{5.75}$$

subject to

$$\sigma_w \geq \sum_{y\in\mathcal{Y}} z_{y,w}^{(\nu)} + \sum_{y\in\mathcal{Y}} \sum_{k\in\mathcal{K}} \lambda_{k,y,w}^{N^{\text{TOT}(\nu)}} \left(N_{k,y}^{\text{TOT}} - N_{k,y}^{\text{TOT}\,(\nu)}\right) + \sum_{y\in\mathcal{Y}} \sum_{z\in\mathcal{Z}} \lambda_{z,y,w}^{sol^{(\nu)}} \left(sol_{z,y} - sol_{z,y}^{(\nu)}\right) + \sum_{y\in\mathcal{Y}} \sum_{z\in\mathcal{Z}} \lambda_{z,y,w}^{wind^{(\nu)}} \left(wind_{z,y} - wind_{z,y}^{(\nu)}\right) + \sum_{y\in\mathcal{Y}} \sum_{h\in\mathcal{H}_C} \lambda_{h,y,w}^{\theta_h{}^{(\nu)}} \left(\theta_{h,y} - \theta_{h,y}^{(\nu)}\right) + \sum_{y\in\mathcal{Y}} \sum_{b\in\mathcal{B}} \lambda_{b,y,w}^{cap^{\text{Batt}(\nu)}} \left(cap_{b,y}^{\text{Batt}} - cap_{b,y}^{\text{Batt}(\nu)}\right) + \sum_{y\in\mathcal{Y}} \sum_{l\in\mathcal{L}_C} \lambda_{l,y,w}^{\theta_l{}^{(\nu)}} \left(\theta_{l,y} - \theta_{l,y}^{(\nu)}\right) + \sum_{y\in\mathcal{Y}} \sum_{l\in\mathcal{PL}_C} \lambda_{pl,y,w}^{\theta_{pl}{}^{(\nu)}} \left(\theta_{pl,y} - \theta_{pl,y}^{(\nu)}\right) + \sum_{y\in\mathcal{Y}} \sum_{ptg\in\mathcal{PtG}} \lambda_{ptg,y,w}^{cap^{\text{PtG}(\nu)}} \left(cap_{ptg,y}^{\text{PtG}} - cap_{ptg,y}^{\text{PtG}(\nu)}\right) \quad w\in\mathcal{W}, \nu = 1,\dots,j-1 \tag{5.76}$$

$$\sigma_w \geq \sigma^{down} \qquad w\in\mathcal{W} \tag{5.77}$$

(5.2)–(5.14) (5.78)

(5.54)–(5.55) (5.79)

(5.56)–(5.63) (5.80)

The objective function (5.75) includes investment, decommissioning and fixed costs and the auxiliary variables σ_w that lower-approximate the operation cost under scenario w. Constraints (5.76) are Benders optimality cuts, which tighten the feasible region of

the master problem over iterations. As previously explained, no feasibility cuts are needed, since the subproblems are always feasible due to the second-stage variables $ENP^{c}_{z,t,w}$, $OG^{c}_{z,t,w}$ and $RNP^{c}_{z,t,w}$ that model energy not provided, over-generation and reserve not provided, respectively. Lower bound constraints (5.77) on variables σ_w avoid the master problem being unbounded in the first iteration, while constraints (5.78) control investment and decommissioning decisions as in the original problem. Equations (5.79) control the renewable penetration, which is now a first-stage variable since it depends only on the solar and wind power installed capacity. Finally, constraints (5.80) define optimization variables for the master problem.

5.3.4.2 Subproblems

At each iteration j, for given values of the first-stage variables $N^{\mathrm{TOT}\,(j)}_{k,y}$, $\theta^{(j)}_{h,y}$, $sol^{(j)}_{z,y}$, $wind^{(j)}_{z,y}$, $cap^{\mathrm{Batt}(j)}_{b,y}$, $\theta^{(j)}_{l,y}$, $\theta^{(j)}_{pl,y}$, and $cap^{\mathrm{PtG}(j)}_{b,y}$ the subproblem associated with year y and scenario w is formulated as follows.

$$\begin{aligned}\min\ z_{y,w} \quad = \quad \sum_{c\in\mathcal{C}^y} wg_c \sum_{t=1}^{24}\Bigg[\sum_{k\in\mathcal{K}}\left(CM_{k,y,w}\left(\underline{P}_k\,\gamma^{c}_{k,t,w}+p^{c}_{k,t,w}\right)+SUC_k\alpha^{c}_{k,t,w}\right)\\ +\sum_{h\in\mathcal{H}} Cvar_h\,E^{\mathrm{OUT},c}_{h,t,w}+\sum_{b\in\mathcal{B}} Cvar_b\,E^{\mathrm{OUT},c}_{b,t,w}\\ +\sum_{n\in\mathcal{N}} C^{\mathrm{PROD}}_{n,y,w}GAS^{\mathrm{PROD},c}_{n,t,w}+\sum_{ptg\in\mathcal{PTG}} Cvar_{ptg}\,GAS^{\mathrm{PtG},c}_{ptg,t,w}\\ +c_{ENP}\sum_{z\in\mathcal{Z}} ENP^{c}_{z,t,w}+c_{OG}\sum_{z\in\mathcal{Z}} OG^{c}_{z,t,w}\\ +c_{RNP}\sum_{z\in\mathcal{Z}} RNP^{c}_{z,t,w}+c_{GC}\sum_{n\in\mathcal{N}} GAS^{\mathrm{CURT},c}_{n,t,w}\Bigg]\end{aligned} \tag{5.81}$$

subject to

$$(5.15)-(5.53) \qquad 1\le t\le 24, c\in\mathcal{C}^y \tag{5.82}$$

$$N^{\mathrm{TOT}}_{k,y}=N^{\mathrm{TOT}\,(j)}_{k,y} \qquad :\lambda^{N^{\mathrm{TOT}}}_{k,y,w} \qquad k\in\mathcal{K} \tag{5.83}$$

$$\theta_{l,y}=\theta^{(j)}_{l,y} \qquad :\lambda^{\theta_l}_{l,y,w} \qquad l\in\mathcal{L}_C \tag{5.84}$$

$$wind_{z,y}=wind^{(j)}_{z,y} \qquad :\lambda^{wind}_{z,y,w} \qquad z\in\mathcal{Z} \tag{5.85}$$

$$sol_{z,y} = sol_{z,y}^{(j)} \quad : \lambda_{z,y,w}^{sol} \quad z \in \mathcal{Z} \tag{5.86}$$

$$cap_{b,y} = cap_{b,y}^{\mathrm{Batt}(j)} \quad : \lambda_{b,y,w}^{cap^{\mathrm{Batt}}} \quad b \in \mathcal{B} \tag{5.87}$$

$$\theta_{h,y} = \theta_{h,y}^{(j)} \quad : \lambda_{h,y,w}^{\theta_h} \quad h \in \mathcal{H}_C \tag{5.88}$$

$$\theta_{pl,y} = \theta_{pl,y}^{(j)} \quad : \lambda_{pl,y,w}^{\theta_{pl}} \quad pl \in \mathcal{PL}_C \tag{5.89}$$

$$cap_{PtG,y}^{\mathrm{PtG}} = cap_{PtG,y}^{\mathrm{PtG}(j)} \quad : \lambda_{b,y,w}^{cap^{\mathrm{PtG}}} \quad ptg \in \mathcal{PtG} \tag{5.90}$$

$$\gamma_{k,t,w}^{c}, \alpha_{k,t,w}^{c}, \beta_{k,t,w}^{c} \geq 0 \quad k \in \mathcal{K}, 1 \leq t \leq 24, c \in \mathcal{C}^{y} \tag{5.91}$$

(5.65)–(5.74) (5.92)

Objective function (5.81) minimizes the operating cost of the integrated system in year y under scenario w by considering thermal production costs, start-up costs, hydro and batteries operational costs, gas production costs, PtG operational costs, and penalties for energy not provided, over-generation, reserve not provided and gas curtailment. Constraints (5.82) include all operating constraints in the original problem. Equations (5.83)–(5.90) fix the complicating variables to values determined by the master problem. At each iteration j, the values of the objective function and dual variables of fixing constraints in the subproblem associated to year y and scenario w are stored in the parameters $z_{y,w}^{(j)}$ and $\lambda_{y,w}^{(j)}$, respectively. Both these parameters are needed to add Benders optimality cuts (5.76) to the master problem. Finally, constraints (5.91) and (5.92) define the optimization variables, with thermal commitment variables relaxed to be continuous variables.

The solution of all subproblems allows computing the following upper bound to the optimal objective function value of the relaxed problem (with continuous and binary investment decisions and continuous operation decisions) at iteration j

$$z_{up}^{(j)} = z_{down}^{(j)} - \sum_{w \in \mathcal{W}} prob_w \cdot \sigma_w^{(j)} + \sum_{w \in \mathcal{W}} prob_w \sum_{y \in \mathcal{Y}} z_{y,w}^{(j)} \tag{5.93}$$

5.3.4.3 Steps of the solution algorithm

Given a small tolerance value ε to control convergence, the Benders decomposition works as follows:

0. **Initialization.** Initialize the iteration counter, set $j = 1$. Set $z_{up}^{(j)} = \infty$ and $z_{down}^{(j)} = -\infty$.
1. **Master problem solution**. Solve the master problem (5.75)–(5.80). Update $z_{down}^{(j)}$ and the values of first-stage variables.
2. **First year**. Consider the first year of the planning horizon, i.e., $y = 1$.
3. **First scenario**. Consider the first scenario, i.e., $w = 1$.
4. **Subproblem solution**. Solve subproblem (5.81)–(5.92) for year y and scenario w. Compute $z_{y,w}^{(j)}$ and store the dual variables of the fixing constraints (5.83)–(5.90).
5. **Scenario update**. Consider the next scenario and repeat step 4. If all scenarios have been considered go to step 6.
6. **Year update**. Consider the next year of the planning horizon and repeat steps from 3 to 5. If all years have been considered go to step 7.
7. **Convergence checking**. Compute $z_{up}^{(j)}$. If $\frac{\left|z_{up}^{(j)} - z_{down}^{(j)}\right|}{z_{up}^{(j)}} < \varepsilon$, the optimal solution has been obtained, go to step 8. Otherwise, update the iteration counter, set $j = j + 1$ and go back to step 1.
8. **Subproblems final integer solution**. For each year y and each scenario w, solve subproblems (5.81)–(5.92) replacing constraints (5.91) with (5.64), i.e., considering MILP problems. The solution obtained is now feasible for the original problem.

5.4 Case study

As a case study for planning the joint expansion of electricity and gas systems, we chose the Italian energy system. Specifically, we modified the scenario introduced in Section 4.5 by including the gas component, so as to determine the joint capacity expansion plans for the Italian integrated power system to reach by 2040 policy goals set by the European Commission.

5.4.1 The Italian natural gas system

In this paragraph, we describe the assumptions introduced to model the Italian natural gas system. We refer the reader to Section 4.5.1 for a detailed description of the assumptions introduced to model also the Italian power system.

Fig. 5.1 illustrates the structure of the Italian gas system considered in the analysis.

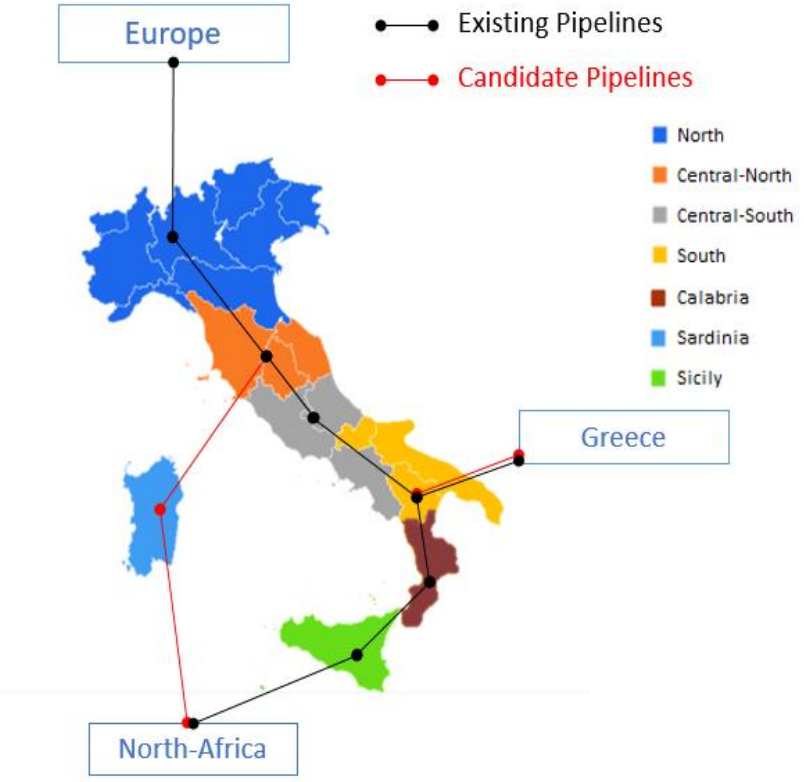


Fig. 5.1 Representation of the Italian gas system

Coherently with the electric power system part of the model, the market analysis for the natural gas sector considers the Italian system divided into seven interconnected market zones: North, Central-North, Central-South, South, Calabria, Sicily, and Sardinia. In terms of internal natural gas transit limits in-between the Italian market zones, the simplified and neutral assumption of (virtually) unlimited transfer capacity has been adopted due to the difficulties in finding detailed and reliable data. Three equivalent areas from which natural gas may be imported are included in the system:

1. The Europe area (EU) represents the natural gas flow at the northern Italian borders (natural gas supply from Russia and North Sea via Passo Gries, Tarvisio, Gorizia pipelines entry points).
2. The North Africa area (Nafr) represents the natural gas flow coming from North African countries (from Algeria and Tunisia via Mazara Del Vallo and Gela entry points in Sicily).
3. The Greece area (GR) represents the natural gas flow of the Trans Adriatic Pipeline – TAP (natural gas from Azerbaijan via Turkey, Greece and Albania to Melendugno, Puglia in South Italy). TAP has been assumed to be fully commissioned and in operation since 2020.

Table 5.1 Fuel prices: low and high prices scenario [€/Gcal]

Low price scenario [€/Gcal]	2020	2030	2040	High price scenario [€/Gcal]	2020	2030	2040
IT_Gas	23.43	28.88	27.63	IT_Gas	23.43	36.84	30.56
EU_Gas	21.43	26.88	25.63	EU_Gas	21.43	34.84	28.56
Nafr_Gas	19.43	24.88	23.63	Nafr_Gas	19.43	32.84	26.56
Gasoil	90.0	91.25	71.58	Gasoil	90.0	85.81	102.14
Coal	9.5	10.47	10.47	Coal	9.5	13.40	10.47
EUmix	42.86	53.77	51.26	EUmix	42.86	69.67	57.12

Along with other fuel prices and CO_2 price, the stochasticity has been applied also to gas fuel production cost, assuming two scenarios, as shown in Table 5.1. Specifically, gas fuel price depends on the specific gas hub: the Italian internal production cost (IT_gas) is more expensive than the imported gas from Greece and the Northern Europe (EU_gas); on the contrary, the gas coming from Algeria is the cheapest (Nafr_gas).

5.4.1.1 Supply side assumptions

There are three categories of natural gas supplies for Italy: (i) pipeline imports; (ii) LNG imports (with the related regasification plants for injecting the natural gas in the national network); and (iii) national production. All the three sources have been considered for modelling natural gas supply for Italy. For the pipeline imports, the flows coming from the three areas described in the previous section – Europe, North Africa and Greece – have been delineated by setting a maximum hourly import, based on the total expected annual import, as reported in the Snam-Terna[2] scenarios published in September 2019 [111]. Table 5.2 shows the data of minimum and maximum natural gas supply from the different borders resulting from Snam-Terna scenario elaborations. LNG imports and national production have been grouped together in terms of total expected supply volumes, and "virtually" fully located in the Central-North Italian market zone.

[2] Since 2018, the Italian natural gas TSO (Snam) and the electricity TSO (Terna) design joint energy scenarios used for the elaborations of their respective development plans, according to the provisions of the Italian regulatory authority (ARERA) and in line with the ENTSO-E and ENTSO-G process to develop the TYNDP.

Table 5.2 Minimum and maximum natural gas supply per pipe entry point in Snam-Terna scenarios [Billion m3/year]

Entry Point	2025		2030		2040	
	Min	Max	Min	Max	Min	Max
Passo Gries (UE)	0.0	1.0	0.0	1.0	0.0	1.0
Tarvisio (UE)	22.0	31.0	24.0	32.0	12.0	31.0
Mazara del Vallo (Nafr)	13.0	22.0	8.0	26.0	7.0	31.0
Gela (Nafr)	5.0	5.0	5.0	5.0	5.0	5.0
TAP (GR)	7.0	9.0	7.0	9.0	7.0	9.0
TOT	**47.0**	**68.0**	**44.0**	**73.0**	**31.0**	**77.0**

Similarly to the pipeline imports, Table 5.3 reports the data for natural gas supply from LNG imports and national production resulting from Snam-Terna scenario elaborations [111].

Natural gas storage has also been included in the simulations, with a total capacity of 17.6 billion m3/year, as reported in the Snam-Terna scenario document. Like LNG imports and national production, also storage has been "virtually" fully located in the Central-North Italian market zone.

Table 5.3 Natural gas supply via LNG and national production in Snam-Terna scenarios [Billion m3/year].

Supply source	2025		2030		2040	
	Min	Max	Min	Max	Min	Max
LNG	10.0	13.0	5.0	9.0	0.0	10.0
National production	5.4		4.2		1.0	

5.4.1.2 New investments

The considered scenario include the possibility of new investments on the supply side of the natural gas sector:

- A Power-to-Gas plant with a gas production capacity up to 10 GW and an efficiency of 68%.
- Two new pipelines:

- The doubling of TAP (*TAP2*), i.e., same route and same capacity of the first pipeline, landing in South Italy as the first pipe.
- The so-called "Gasdotto Algeria Sardegna Italia – GALSI", from Algeria to Sardinia (Italy) and to Italy mainland, with a total capacity of 8 billion m^3/year.

Table 5.4 resumes the investment costs of the candidate projects.

Table 5.4 Investment costs of gas supply new projects

	P2G	GALSI Nafr-Sar	GALSI Sar-C.North	TAP2
Capacity [MWth]	10000	9750	9750	10873
IC [M€/MWth]	0.350	0.154	0.1025	0.0934

5.4.1.3 Demand side assumptions

The natural gas demand is composed of two parts treated and modelled differently. The first part is the industrial and tertiary (residential and offices) sectors. The industry plus tertiary gas demand has been aggregated and modeled as a proper input to the model, and, in line with the power sector load, it has an hourly profile, with the clustering analysis described in Section 5.3.1 applied to identify representative days and keep the computation tractable.

The current (2020) level of industrial and tertiary gas demand and the hourly profiles are based on the operational data published by Snam on its website, while the demand trend from 2020 to 2040 is based on the outcomes of Snam-Terna scenarios for natural gas demand for final uses[3]. Snam-Terna elaborated two long-term (2040) scenarios, called "centralised" and "decentralised": while the centralised scenario shows a slight decrease in gas demand for final uses compared to the historical values, the decentralised scenario shows a more intense gas demand decline for final uses, as shown in Table 5.5.

The average percentage decrease from the latest historical data (2017) to 2040 that can be calculated from the data reported in Snam-Terna scenarios document has been used to calculate the target demand in 2040 (-12.2%). From 2020 to 2040 a linear constant percentage decrease (compound annual growth rate) has then been used (-0.6%/year).

[3] Final uses represent all uses apart from the thermal power plants consumptions.

Table 5.5 Natural gas demand for final uses in centralised (CEN) and decentralised (DEC) Snam-Terna scenarios [Billion m3/year]

	2017	2025			2030			2040		
		CEN	DEC	AVG	CEN	DEC	AVG	CEN	DEC	AVG
Gas Demand [Billion m³/year]	46.5	45.4	42.4	43.9	45.3	40.7	43.0	45.2	36.5	40.9
Change from 2017 [%]		-2.4%	-8.8%	-5.6%	-2.6%	-12.5%	-7.5%	-2.8%	-21.5%	-12.2%

The starting point (year 2020) is based on the latest available natural gas demand data, as published by the Ministry of Economic Development[4].

The second part of the natural gas demand is related to the thermal power generation, representing an output of the simulations.

5.4.2 Clusters of Italian thermal power plants

As explained in the previous sections, when addressing the expansion planning for real-scale integrated systems, incorporating traditional unit commitment constraints would be computationally infeasible. To keep the scenario computationally tractable, we applied the clustering procedure described in Section 5.3.2 on Italian thermal power plants. Specifically, we performed several tests for different values of the threshold to be used in the termination test.

Table 5.6 Number of clusters of thermal power plants for different threshold values

	Threshold					Number of power plants
	0.80	0.85	0.90	0.95	0.99	
North	4	5	7	14	25	74
Central-North	3	4	5	6	9	13
Central-South	5	5	7	10	16	34
South	3	3	4	5	9	12
Calabria	2	2	2	3	4	7
Sicily	4	5	6	9	14	19
Sardinia	4	4	5	6	9	12
TOT	**25**	**28**	**36**	**53**	**86**	**171**

[4] https://www.mise.gov.it/index.php/it/energia/gas-naturale-e-petrolio/gas-naturale

As can be observed in Table 5.6, a trade-off between computational costs and approximation accuracy has to be considered in the threshold choice. Indeed, the higher the threshold value, the greater the similarity of thermal power plants grouped in the same cluster. However, very high threshold values require considering a large number of clusters, reducing the computational saving induced by the CUC formulation. In our numerical experiments, we decided to fix the threshold to 90%, so as to group the 171 Italian thermal power plants included in the scenario into 36 clusters. Specifically, in our analysis seven clusters are introduced both in the North and in the Central-South zones. Thermal power plants in Sicily are grouped into six clusters. Five clusters are considered both in the Central-North and in Sardinia. The thermal power plants in the South of Italy are grouped into four clusters, while only two clusters are introduced to represent the thermal units in Calabria.

5.4.3 Results and discussion

We solved the proposed model on a computer with two 2.10 GHz Intel® Xeon® Platinum 8160 CPU Processors and 128 GB of RAM, using language extension GUSS integrated with solver Gurobi under GAMS 24.7.4. We considered $\varepsilon = 10^{-3}$ as tolerance for Benders decomposition convergence, while in each iteration we solved the master problem up to optimality. Fig. 5.2 illustrates the evolution over iterations of the multi-cut Benders algorithm, which converges in 63 iterations. Indeed, as better shown in Fig. 5.3 at iteration 63 the relative distance between upper and lower bound equals $0.97 \cdot 10^{-3}$, satisfying the predefined tolerance.

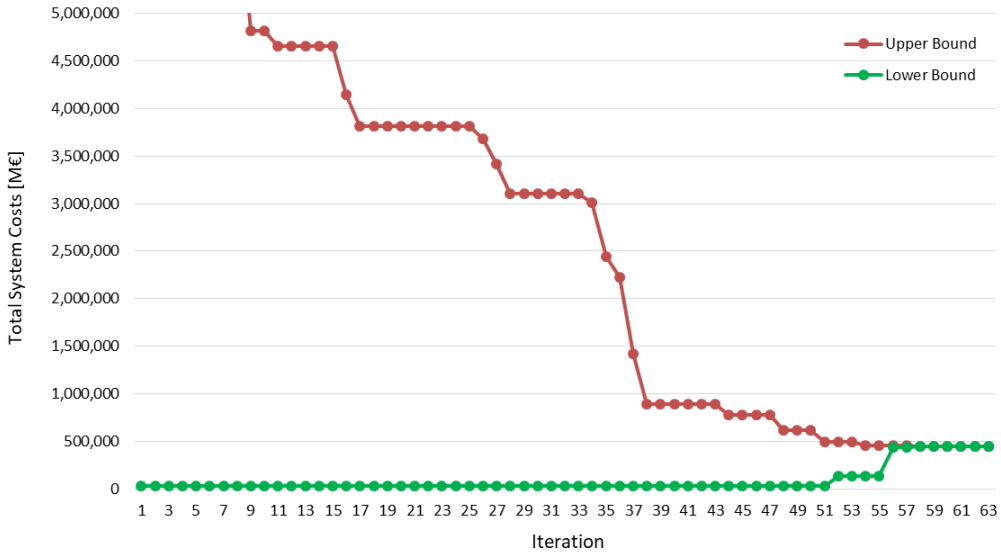


Fig. 5.2 Upper and lower bounds values over iterations in multi-cut Benders algorithm

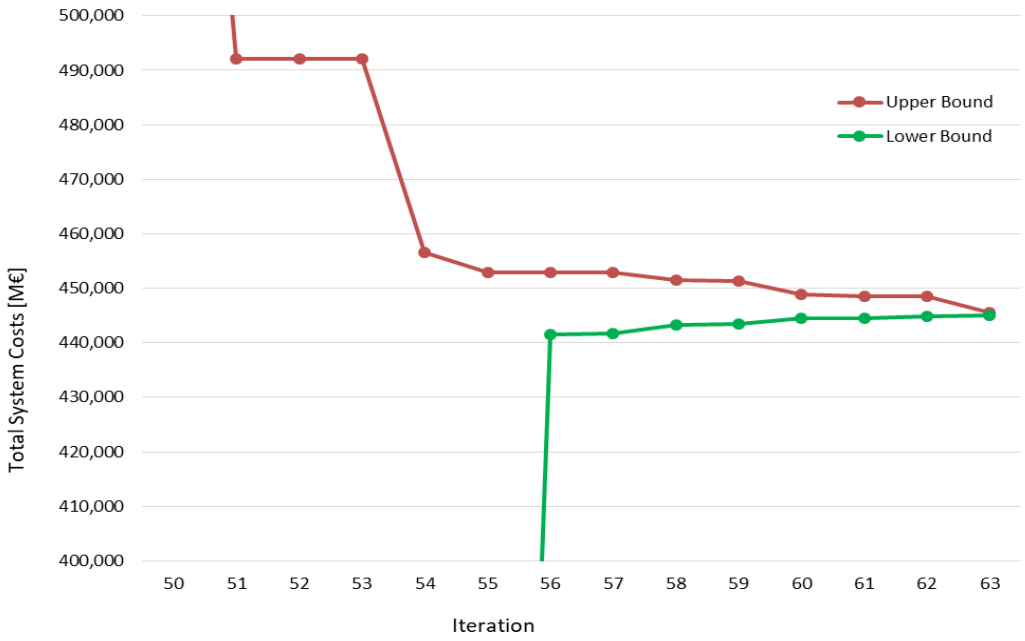


Fig. 5.3 Upper and lower bounds values at the last iterations of multi-cut Benders algorithm

The total time needed to solve the problem is 36,127 seconds, corresponding to 10 hours 2 minutes and 7 seconds. Table 5.7 provides more information about computational times, specifying the size and the solution time for the master problem and the subproblems at the last iteration of Benders decomposition algorithm. As regards to subproblems, while the base case solution requires about two minutes, each of the updated subproblems is solved in only 3.342 seconds on average. Since the number of subproblems equals 126, at each iteration, the average time required to solve all the subproblems is 538.041 seconds. As explained in the previous section, once convergence is reached, subproblems are solved as MILP models, obtaining the final solution, characterized by total system costs higher than the convergence value of the Benders decomposition algorithm. However, thanks to the tight formulation of thermal unit commitment constraints, this difference is very small.

Table 5.7 Size and solution time of master problem and subproblems at the last iteration of Benders algorithm

	# Constraints	# Decision Variables	# Discrete Variables	CPU Time [seconds]
Master Problem	5,441	4,795	2,016	10.018
First Subproblem (Base Case)	318,651	306,918	0	120.291
Updated Subproblem	318,651	306,918	0	3.342

Indeed, in this problem instance, the total system costs in the relaxed solution are 445,515 M€, while the integer solution is only 0.015% more expensive, with the objective function value being 445,582 M€.

The system expected costs for the whole expansion planning period are shown in Table 5.8. As can be observed, there is a huge difference between first-stage and second-stage costs: while the sum of investment, decommissioning and fixed costs represents 11.7% of total costs, operating costs account for 88.3% of total costs. Specifically, the most relevant cost for the system is related to the production costs of thermoelectric power plants, representing 96.3% of total operating costs: the variable cost includes O&M, CO_2 emissions costs and fuel consumption, except for gas-fired plants whose fuel consumption is included in the gas production cost.

On the contrary, start-up costs of thermoelectric power plants have a small impact, being 0.2% of total costs.

Table 5.8 Breakdown of integrated system costs

Costs	M€	%
Thermal Capacity Expansion	2,113	0.47%
Wind Capacity Expansion	16,775	3.76%
Solar Capacity Expansion	27,234	6.11%
Transmission Capacity Expansion	1,075	0.24%
Batteries Capacity Expansion	2,702	0.61%
Pipeline Capacity Expansion	2,222	0.50%
Decommissioning Costs	179	0.04%
Thermal Fixed Costs	9	0.002%
Expected Thermal Production Cost	378,875	85.03%
Expected Start-Up Costs	740	0.17%
Expected Hydro Operation Costs	131	0.03%
Expected Batteries Operation Costs	789	0.18%
Expected Gas Production Costs	413	0.09%
Expected Penalties for Overgeneration	10,135	2.27%
Expected Penalties for Reserve Not Provided	2,190	0.49%
Total System Expected Costs	**445,582**	**100%**

Table 5.9 Integrated system operation costs breakdown for different scenarios [M€]

	Low Fuel Prices			High Fuel Prices		
	Scenarios for CO_2 prices			Scenarios for CO_2 prices		
	Low	Medium	High	Low	Medium	High
Thermal Production	335,802	349,754	386,658	378,472	392,673	429,621
Start-Up	618	646	740	776	799	861
Penalties for Overgeneration	10,077	10,085	10,060	10,310	10,162	10,114
Penalties for Reserve Not Provided	4,552	2,308	3,038	305	2,609	329
Hydro Operation	144	139	131	128	124	122
Batteries Operation	777	777	789	794	793	803
Gas Production Costs	378	378	378	448	448	448
Total Operation	352,348	364,087	401,794	391,233	407,608	442,298
Investment	52,309	52,309	52,309	52,309	52,309	52,309
Total Cost	**404,657**	**416,396**	**454,103**	**443,542**	**459,917**	**494,607**

Table 5.9 describes how system costs vary depending on stochastic prices: while investment, decommissioning and fixed costs are independent of the scenario realization, operation costs, being second-stage costs, depend on the considered scenario. As can be noticed, all scenarios present similar values of operational costs for both hydropower plants and batteries, while thermal production costs and gas production costs significantly differ between scenarios. As expected, since CO_2 and fuel prices affect the thermal plants marginal cost $CM_{k,y,w}$, the higher these parameters, the greater the production costs, which vary from 336 M€ in the scenario with low emission costs and low fuel price to 430 M€ in the scenario with high CO_2 price and high fuel prices. Furthermore, gas production cost increases from 378 M€ in low fuel price scenarios to 448 in high fuel price scenarios.

Table 5.10 shows the additional capacity of wind and PV installed to reach the RES penetration target in 2040. The RES installed capacity consists of 44.8 GW of PV and 22.0 GW of wind power: this unbalance may be explained by the lower costs of the PV technology with respect to the wind technology.

Table 5.10 Cumulative renewable generation capacity expansion [GW] for the Italian integrated system divided by source and implementation year

Year	Wind	PV
2020	3.8	13.4
2025	8.7	21.6
2035	18.6	38.0
2040	22.0	44.8

As far as interconnection projects are concerned, one new national and one international cross border lines are expected to be implemented in years 2025 to better exploit the variable renewable energy sources and tackle the decommissioning of some Italian thermoelectric power plants. The selected interconnections are listed in Table 5.11.

Table 5.11 Candidate interconnections selected by the model for the Italian integrated system

From	To	Transmission Limits	Year of intervention
Tunisia	Sicily	[–600 MW ; 600 MW]	2025
Central-South	Central-North	[–1000 MW ; 1000 MW]	2025

As described in the previous sections, the total natural gas demand consists of two parts: (i) the industrial and tertiary sectors demand, which is an input to the model; and (ii) the thermal power plants demand, which is an output of the model. In the simulations performed, the total natural gas demand results equal to 845 TWh(th) in 2020, to 741 TWh(th) in 2030 and to 660 TWh(th) in 2040, as reported in Fig. 5.4, with very slight differences according to the selected fuel and CO2 costs scenario. The natural gas demand decrease over the planning horizon is due both to the assumed progressive decline of the industrial and tertiary sectors demand, as well as to increasing penetration of renewables, which decreases the gas consumption from thermal power plants.

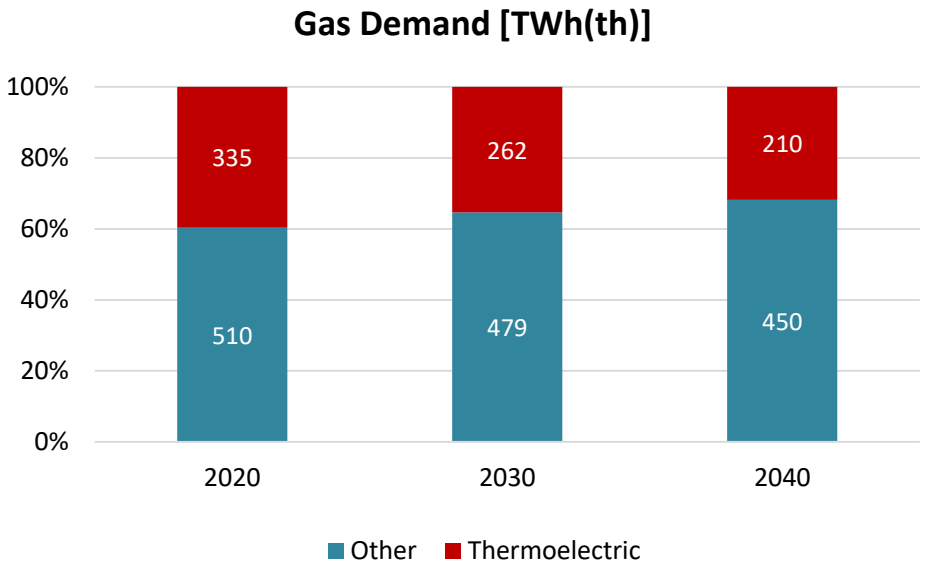


Fig. 5.4 Expected domestic gas demand divided by sector and year

In terms of new pipelines projects, the decommissioning of all coal and oil thermal power plants requires the availability of an additional amount of gas and therefore the development of the so called "Gasdotto Algeria Sardegna Italia – GALSI" projects starting from year 2022. In the current scenario, no capacity of PtG plants is supposed to be installed because of the high investment costs. Fig. 5.5 shows the gas suppliers distribution for Italy by year 2040. As can be observed, the majority of Italian gas demand is satisfied by importing gas from North-Africa, while the domestic production provides only a marginal contribution to the gas demand supply.

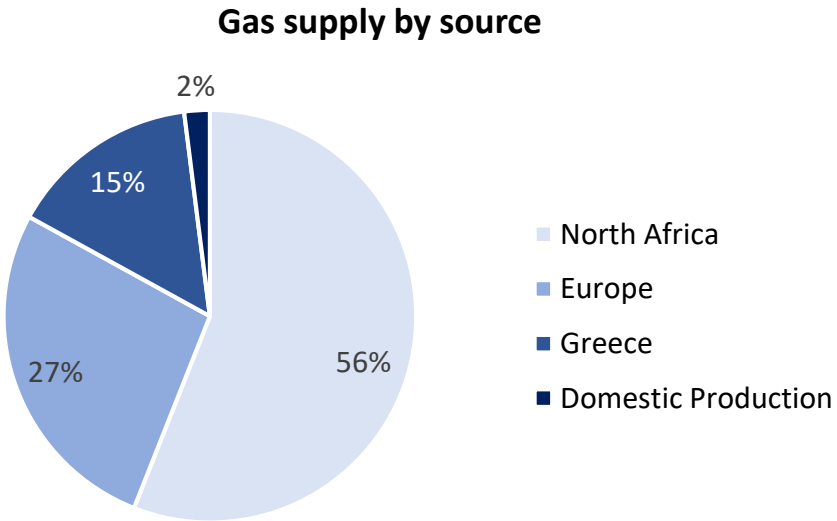


Fig. 5.5 Italian expected gas supply by source in year 2040

Table 5.12 Installed capacity of batteries [MW] in the Italian integrated system

Zone	Technology	Installed Capacity [MW]
North	Lithium-Ion Batteries	95.5
South	Lithium-Ion Batteries	600
South	Sodium-Ion Batteries	600
Sicily	Lithium-Ion Batteries	600
Sicily	Sodium-Ion Batteries	600
Sardinia	Lithium-Ion Batteries	600

Moreover, the tool couples the installed RES capacity with energy storage systems, installing throughout the planning period 3.1 GW of batteries. A list that summarizes the installed capacity according to technology and zone is reported in Table 5.12. As can be noticed, Lithium Ion batteries are preferred to Sodium ones due to their lower investment costs. Instead, no additional pumping power plants are built in the solution provided by the model.

The decommissioning of 14.7 GW oil and coal power plants has been planned for year 2024, according to the forecast scenarios of European TSO's.

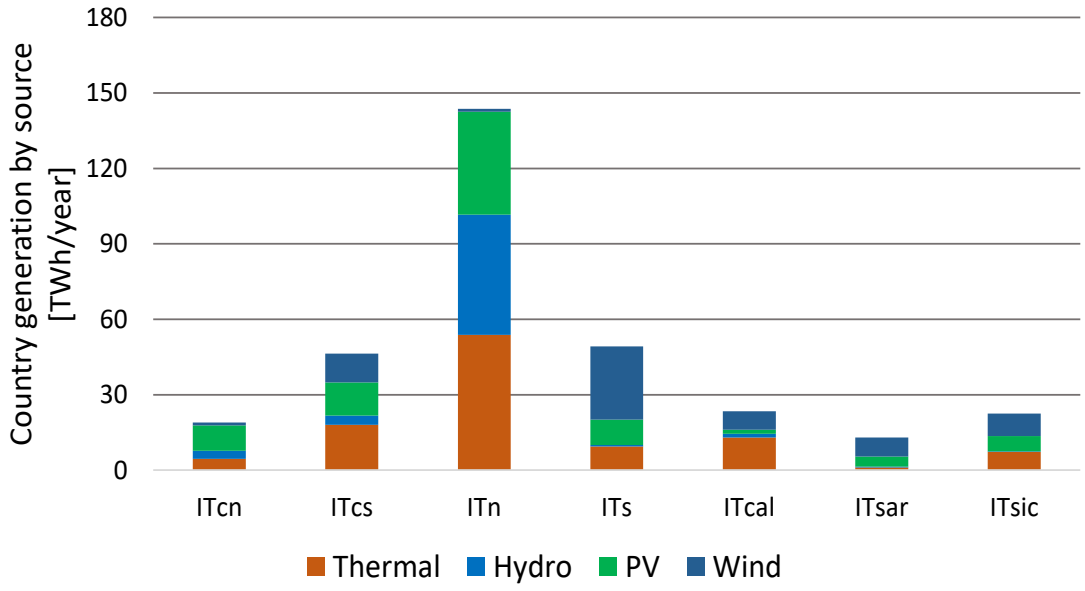
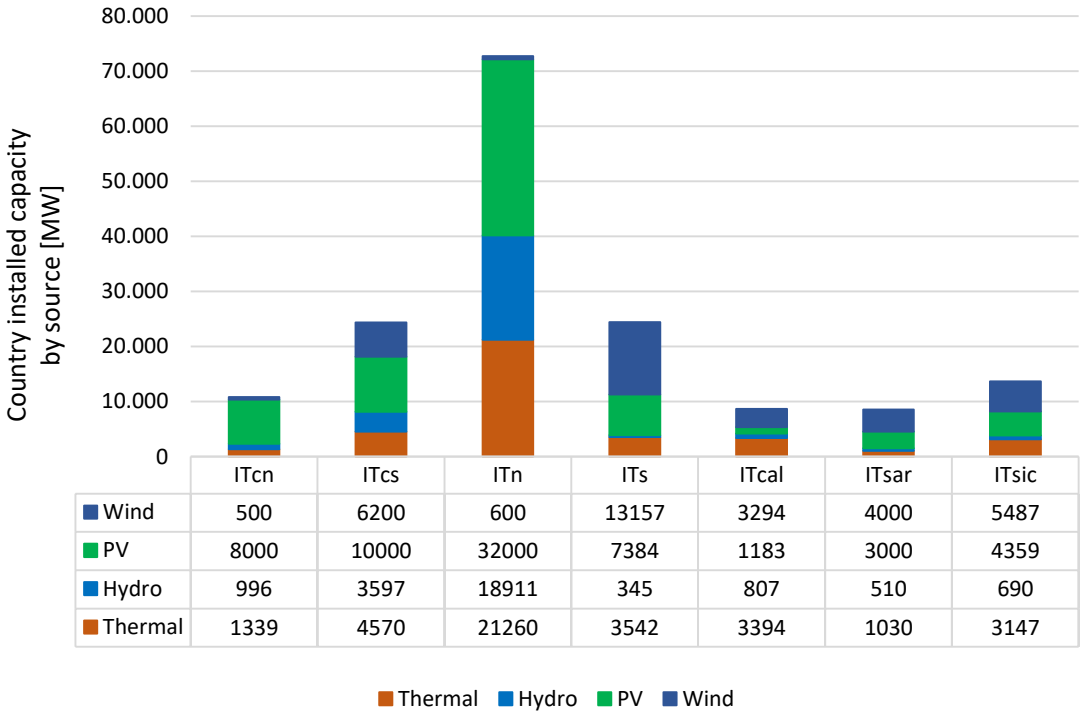


Fig. 5.6 Expected energy generation by source for each Italian market zone in year 2040



	ITcn	ITcs	ITn	ITs	ITcal	ITsar	ITsic
Wind	500	6200	600	13157	3294	4000	5487
PV	8000	10000	32000	7384	1183	3000	4359
Hydro	996	3597	18911	345	807	510	690
Thermal	1339	4570	21260	3542	3394	1030	3147

Fig. 5.7 Installed capacity by source for each market zone in year 2040

Fig. 5.6 reports the expected energy generation in year 2040 divided by energy source, for each Italian zone. As it can be noticed, since the RES production (i.e., from wind and solar) is 152.6 TWh/year while the load is 400 TWh/year, the target of reaching 36% of non-programmable renewable penetration by year 2040 has been fully achieved.

Finally, Fig. 5.7 shows the installed capacity at the end of the planning period, grouped by source and zone. As can be noticed, the model suggests installing large shares of PV capacity in all Italian market zones and especially in the North, while the wind expansion is mainly located in southern regions, which are characterized by the highest wind capacity factors.

5.5 Chapter conclusions

In this chapter, the problem of planning the joint expansion for electricity and gas systems has been addressed. First, a detailed review of the integrated planning frameworks developed in the literature to accurately consider the coupling between electricity and gas systems has been provided.

Then, our approach to the expansion co-planning of integrated systems with bi-directional energy conversion has been introduced. Specifically, in our work, we consider

different scenarios for future values of fuel prices and CO_2 price and we develop a two-stage stochastic programming model that provides an appropriate balance between accuracy and computational tractability, with the first stage representing the investment problem and the second stage being the operational problem. Such a model aims at minimizing the sum of investment, decommissioning and fixed costs and expected value of operational costs, which are computed by considering under each scenario thermal power production costs, start-up costs, hydro and batteries operational costs, gas production costs, PtG operational costs, and penalties for energy not provided, over-generation, reserve not provided, and gas curtailment.

The inclusion of the gas system in the decision making framework further increases the computational burden of the expansion planning model. To keep the problem computationally tractable we select representative days, we implement the CUC formulation, by grouping similar thermal power plants into clusters, and we solve the stochastic model by applying a solution algorithm based on the multi-cut Benders decomposition strategy.

Finally, a case study concerning the Italian energy system has been presented. In our numerical study, we simulate a planning horizon of 21 years, from 2020 up to 2040, and we plan the joint expansion of the Italian electricity and gas system under CO_2 and fuel prices uncertainty. Empirical results show the need to install large shares of renewables to achieve policy goals set by the European Commission, with the solar technology preferred to the wind power technology due to the lower investment costs. As regards to the natural gas system, the decommissioning of all coal and oil thermal power plants requires the availability of an additional amount of gas and therefore the development of the so called “GALSI”, from Algeria to Sardinia and to Italy mainland, starting from year 2022.

The main assumption of the analysis described in this chapter is the hypothesis of demand inelasticity. However, in reality, customers react to electricity prices, by changing their electricity consumption according to the price signals. The inclusion of the demand elasticity in the expansion planning framework is addressed in the following chapter.

Chapter 6

Modeling demand reactions to electricity price signals

6.1 Research motivation

The variability of intermittent renewable power sources has to be managed in the short-term to ensure that power systems are operated in an efficient and reliable way. Such a variability can be controlled by adjusting both the generation and the demand for electricity. As regards to the supply side, the volatility of renewable generation can be compensated by either committing fast-reacting thermal power plants or transferring surplus electricity generation from off-peak hours to peak hours by means of storage technologies. Both alternatives have already been included in the expansion planning framework in the previous chapters. However, the demand–supply balance can be reached also by acting on load profiles, by either reducing the electricity consumption during peak hours or shifting the demand from peak hours to off-peak hours if consumption is inevitable.

The change of the demand profile reacting to system conditions is referred to as demand response. Demand response is an emerging technology in Europe's Electricity markets that will introduce a new degree of flexibility [112]. Specifically, reference [113] discusses the benefits associated with demand response, which can be classified into four main categories, namely market-wide, reliability, market performance, and participant benefits. First, an overall electricity price reduction is expected due to the more efficient utilization of the available infrastructure, producing benefits for all electrical consumers. Second, by having a well-designed demand response program, participants have the opportunity to help in reducing the risk of outages. Third, demand response programs can improve market performance by reducing both the ability of main market players to exercise power in the market and the price volatility in the spot market. Indeed, a small reduction of demand could lead to a significant price reduction, since generation cost exponentially increases near maximum generation capacity. Firth, reference [113] argue that customers participating in demand response can expect bill savings by reducing

their electricity consumption during peak hours, even if it may be expected that savings for small consumers could be estimated around a few tens of euros per year. Finally, in reference [114] it is claimed that demand response could also help the optimal transition to a European low emission power system by decreasing electricity usage and improving energy efficiency, estimating between 0.23% and 3.3% the demand response contribution to the CO_2 emissions reduction target from the energy sector set by the European Commission.

Although the several benefits provided by demand response programs, most of these outcomes may also be obtained by the deployment of other technologies which may compete with demand response, such as batteries, hydro pumping units or combined cycle gas turbines power plants. Thus, an integrated planning framework to evaluate the competition between different resources providing flexibility to power systems is needed.

This chapter focuses on the inclusion of the demand response in the expansion planning problem by modeling consumers reactions to electricity prices. The structure of the chapter is as follows. Section 6.2 reviews the existing literature for demand response modeling. Section 6.3 describes the proposed method to include customers reactions to electricity prices in the GTEP problem. Specifically, such an approach is based on the introduction of elasticities and demand functions. Several numerical tests to assess the proposed method are introduced in Section 6.4. Finally, Section 6.5 concludes the chapter.

6.2 Literature review

Several studies have been conducted in the literature to investigate the use of demand response in power systems with large shares of renewables. Many works focus on power systems operations by evaluating the short-term cost-savings potential of changes in the load shape. For instance, the operation of an electric system with high wind penetration is modeled by means of a unit commitment model in reference [115]. Results obtained by analysing the isolated power system of Gran Canaria show how demand response could be useful to level out variations in wind power production, leading to a reduction of both wind power curtailment and system costs.

The potential of demand response for energy systems with large shares of wind power generation is also discussed in [116] and [117]. Specifically, reference [116] proposes different methodologies to model short-term responsiveness, by including cross-price elasticities that account for load shifts among hours. Comparison of model results for a

single year optimization with and without demand response shows peak reduction and valley filling effects, demonstrating that demand response could facilitate the integration of renewable energy. Similar results are obtained in [117] by means of a day-ahead unit commitment problem combined with a real-time dispatch model. Specifically, results for the Texas power system demonstrate the potential increases in wind generation from implementing demand response programs.

The day-ahead scheduling problem for power systems with high shares of renewables is also addressed in [118], which aims at providing a systematic approach to evaluate the level of flexibility of a power system by considering fast-ramping units, hourly demand response and energy storage. The proposed formulation is a MINLP, which is then converted into a MILP by introducing auxiliary binary variables. Numerical tests indicate that hourly demand response could play a very important role in system flexibility.

Instead, other works in the literature deal with the integration of demand response in long-term investment problems. For instance, reference [119] presents a long-term power generation expansion planning model that features a long planning horizon, an hourly time resolution, multiperiod investment and retirement decisions, transmission constraints, start-up restrictions, and short-term demand response, with demand reactions to electricity prices modeled by means of inter-hour demand elasticities. The impact of demand responsiveness on decision making in generation expansion planning is also investigated in [120]. In such a work, a portion of the demand is assumed to be elastic and responsive to the price and a linear functionality is introduced to model the relationship between elastic demand and market price. Simulation results show that reducing just 3% of the customers' demand due to price elasticity may result in a benefit of about 10% for customers in the long term.

While references [119] and [120] only consider the generation expansion, other studies in the literature address the integration of demand response programs in the network investment planning problem. For instance, in [121] a probabilistic approach using Monte Carlo simulation is proposed for transmission investment planning along with demand response schedule considering high penetration of wind energy resources. The paper focuses on direct load curtailment control program, in which system operator has a contractual authority to curtail the demand at any particular bus. The objective of the model is to find the most probable optimal transmission network expansion along with optimum load curtailment schedule so that the proposed network is able to securely accommodate a high integration of wind energy. Interactions between demand response and grid expansion planning are analysed also in reference [122], which evaluates the influence of different demand response penetration levels on network investment

decisions. The analysis is based on a stochastic programming approach while using a reasonably accurate representation of the physical characteristics of the network via a linearized AC network model. Results obtained by analysing the Irish transmission system under a twelve-year planning horizon reveal that having a 2.5% reduction in peak energy demand could lead to a nearly 30% reduction in the number of lines required to be built by the end of the planning horizon. Indeed, the system-wide peak demand is one of the major drivers of network expansion needs: the reduction of the peak energy demand by a small factor could lead to huge system costs savings. Similar conclusions are drawn in [123], which proposes a probabilistic multi-objective transmission expansion planning model to find an optimal trade-off between transmission investment and demand response expenses. The Monte Carlo simulation method is implemented to handle the uncertainty of the loads, demand response programs and distributed generation in the expansion planning framework.

The joint expansion planning of distributed generation and the distribution network considering the impact of energy storage systems and demand response programs is addressed in [124]. Such a problem is formulated as a stochastic-programming-based model driven by the maximization of the net social benefit to determine the location and size of new generation and storage units and the distribution assets to be installed, reinforced or replaced. Results for the insular power system of La Graciosa, Canary Islands, show how demand response can substitute generation and distribution network expansion, increasing the volume of renewable power generation optimal allocated in the system.

Reference [125] presents an evaluation of the potential impacts of demand response on the GTEP problem. Specifically, a multi-period multi-objective GTEP model is proposed and formulated as a mixed integer quadratic programming problem and several levels of demand response penetration are considered in order to assess its impact on the system performance. Similarly to the previous works, results indicate that an increase in penetration of demand response causes a decrease of power generation, emissions and system costs whilst increasing renewables utilization in power system.

While most of the works developed in the literature investigate the impact of different levels of demand response penetration on power system costs, there is a lack of articles that include investment decisions in demand response in the expansion planning framework. For instance, reference [112] integrates investment and operation decisions on demand response devices within a two-stage stochastic GTEP model that includes uncertainty at the operational level and energy economics dynamics at a strategic level. Several classes of shiftable and curtailable loads in residential, commercial and industrial

sectors are considered, including flexibility periods, operational costs and endogenous demand response investments. Results obtained on the European power system show that demand response capacity substitutes partially flexible supply side capacity from peak gas plants and battery storage, in addition to enabling more solar PV production.

Following the approach adopted in [112], in this chapter we propose an integrated planning framework that could jointly consider investments in power balancing technologies. However, our method differs from the procedure applied in [112] in the way demand responsiveness is modeled. Indeed, while in [112] load is remotely controlled by the System Operator, in our approach customers reactions to price signals are modeled.

6.3 Modeling framework

In this section, we propose our approach to include demand response programs into the GTEP analysis. Such an approach is based on the introduction of elasticities and demand functions to model demand reactions to electricity prices. First, the load shifting objective is modeled. Then, the peak shaving strategy is considered. Finally, an optimization model to plan investments in electrical demand response devices is presented.

6.3.1 Notation

In this paragraph, we introduce only the additional notation needed to integrate the demand side management programs in the expansion planning framework. We refer the reader to Section 5.3.3.2 for the complete problem notation.

Parameters

$\delta_{\text{up}}^{\text{LS}}$	[–]	% of maximum demand variation upward for the load shifting strategy
$\delta_{\text{dw}}^{\text{LS}}$	[–]	% of maximum demand variation downward for the load shifting strategy
$\delta_{\text{dw}}^{\text{PS}}$	[–]	% of maximum demand variation downward for the peak shaving strategy
$\varepsilon_{\text{up}}^{\text{LS}}$	[–]	Demand elasticity for upward variations in the load shifting strategy

$\varepsilon_{\mathrm{dw}}^{\mathrm{LS}}$	[−]	Demand elasticity for downward variations in the load shifting strategy
$\varepsilon_{\mathrm{dw}}^{\mathrm{PS}}$	[−]	Demand elasticity for downward variations in the peak shaving strategy
$PrRef_{z,t,w}^{c}$	[€/MWh]	Reference price in zone z in hour t of representative day c under scenario w
$Tran_{z}^{\mathrm{LS}}$	[€/MWh]	Transaction cost for moving the demand in the load shifting strategy in zone z
$Tran_{z}^{\mathrm{PS}}$	[€/MWh]	Transaction cost for reducing the demand in the peak shaving strategy in zone z
$PenPr$	[MWh]	Penalty for price
$v_{z,0}^{\mathrm{LS}}$	[MW]	Total installed capacity of electrical devices for load shifting in zone z at the beginning of the planning horizon
$v_{z,0}^{\mathrm{PS}}$	[MW]	Total installed capacity of electrical devices for peak shaving in zone z at the beginning of the planning horizon
$X_{z,y}^{\mathrm{LS}}$	[MW]	Upper bound on the capacity for load shifting devices that can be installed in zone z in year y
$V_{z,y}^{\mathrm{LS}}$	[MW]	Upper bound on the total capacity for load shifting devices installed in zone z in year y
$X_{z,y}^{\mathrm{PS}}$	[MW]	Upper bound on the capacity for peak shaving devices that can be installed in zone z in year y
$V_{z,y}^{\mathrm{PS}}$	[MW]	Upper bound on the total capacity for peak shaving devices installed in zone z in year y
IC_{z}^{LS}	[€/MW]	Investment cost of load shifting devices in zone z
IC_{z}^{PS}	[€/MW]	Investment cost of peak shaving devices in zone z

Variables

1) First-stage variables

$x_{z,y}^{\mathrm{LS}}$	[MW]	Capacity of load shifting devices installed in zone z in year y
$v_{z,y}^{\mathrm{LS}}$	[MW]	Total capacity of load shifting devices available in zone z in year y
$x_{z,y}^{\mathrm{PS}}$	[MW]	Capacity of peak shaving devices installed in zone z in year y

$v_{z,y}^{\mathrm{PS}}$	[MW]	Total capacity of peak shaving devices available in zone z in year y

2) Second-stage variables

$dFlex_{z,t,w}^{c}$	[MW]	Flexible demand for electricity in zone z in hour t of representative day c under scenario w
$dVar_{z,t,c,w}^{\mathrm{LS,up}}$	[MW]	Upward demand variation due to load shifting in zone z in hour t of representative day c under scenario w
$dVar_{z,t,c,w}^{\mathrm{LS,dw}}$	[MW]	Downward demand variation due to load shifting in zone z in hour t of representative day c under scenario w
$dVar_{z,t,c,w}^{\mathrm{PS,dw}}$	[MW]	Downward demand variation due to peak shaving in zone z in hour t of representative day c under scenario w
$pr_{z,t,w}^{c}$	[€/MWh]	Price with demand side management in zone z in hour t of representative day c under scenario w

6.3.2 Load shifting

Load shifting aims at transferring the demand for electricity from peak hours to off-peak hours in order to flatten the demand curve, obtaining consequently an economic saving due to the substitution of more expensive energy with cheaper energy. Two different approaches to model load shifting are usually adopted in the literature, namely the centralized approach and the method based on elasticities and demand functions. Both the approaches are introduced in this section. Specifically, in the first method, the decision to shift electricity consumptions is taken by a central entity according only to a cost criterion. In such an approach, decision variables $dFlex_{z,t,w}^{c}$ are introduced to represent the hourly zonal flexible demand for electricity, which is now a model variable rather than an input parameter. The flexible demand is then computed from the original load $D_{z,t}^{\mathrm{ELEC},c}$ by adding the upward demand variation $dVar_{z,t,c,w}^{\mathrm{LS,up}}$ and subtracting the downward demand variation $dVar_{z,t,c,w}^{\mathrm{LS,dw}}$, as stated by contraints (6.1).

$$dFlex_{z,t,w}^{c} = D_{z,t}^{\mathrm{ELEC},c} + dVar_{z,t,c,w}^{\mathrm{LS,up}} - dVar_{z,t,c,w}^{\mathrm{LS,dw}}$$

$$z \in \mathcal{Z}, 1 \leq t \leq 24, c \in \mathcal{C}^{y}, y \in \mathcal{Y}, w \in \mathcal{W} \quad (6.1)$$

New load balance equations (6.2) are then formulated by replacing parameter $D_{z,t}^{\mathrm{ELEC},c}$ with the flexible demand in constraints (5.39).

$$\sum_{k\epsilon\Omega_z^k}\left(\underline{P}_k\,\gamma_{k,t,w}^c + p_{k,t,w}^c\right) + \mu_{z,t}^c\left(sol_{z,0} + \sum_{i=1}^{y} sol_{z,i}\right) + \rho_{z,t}^c\left(wind_{z,0} + \sum_{i=1}^{y} wind_{z,i}\right) +$$

$$+ \sum_{l|rz(l)=z} x_{l,t,w}^c + \sum_{h\epsilon\Omega_z^h} E_{h,t,w}^{\mathrm{OUT},c} + \sum_{b\epsilon\Omega_z^b} E_{b,t,w}^{\mathrm{OUT},c} + ENP_{z,t,w}^c = dFlex_{z,t,w}^c +$$

$$+ \sum_{l|sz(l)=z} x_{l,t,w}^c + \sum_{h\epsilon\Omega_z^h} E_{h,t,w}^{\mathrm{IN},c} + \sum_{b\epsilon\Omega_z^b} E_{b,t,w}^{\mathrm{IN},c} + \sum_{ptg\epsilon\Omega_z^{\mathrm{PtG}}} \frac{GAS_{ptg,t,w}^{\mathrm{PtG},c}}{\eta_{ptg}} + OG_{z,t,w}^c$$

$$z \in \mathcal{Z}, 1 \le t \le 24, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W} \qquad (6.2)$$

Let us denote by z^{ante} the system cost function without demand management resulting from equation (5.1). In the presence of load shifting, the system cost to be minimized is z^{LS}, which is expressed as

$$\min\ z^{\mathrm{LS}} = z^{\mathrm{ante}} + \sum_{w\in\mathcal{W}} prob_w\left(\sum_{y\in\mathcal{Y}}\sum_{c\in\mathcal{C}^y} wg_c \sum_{t=1}^{24}\sum_{z\in\mathcal{Z}} Tran_z^{\mathrm{LS}} dVar_{z,t,c,w}^{\mathrm{LS,dw}}\right) \qquad (6.3)$$

with $Tran_z^{\mathrm{LS}}$ being the transaction cost of shifting the demand. Furthermore, due to the load shifting, demand in high price hours can be lowered, but must be consumed during other hours. The balance between demand variations during the day is imposed by the following constraints:

$$\sum_{t=1}^{24} dVar_{z,t,c,w}^{\mathrm{LS,up}} = \sum_{t=1}^{24} dVar_{z,t,c,w}^{\mathrm{LS,dw}} \qquad z \in \mathcal{Z}, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W}. \qquad (6.4)$$

Finally, parameters $\delta_{\mathrm{up}}^{\mathrm{LS}}$ and $\delta_{\mathrm{dw}}^{\mathrm{LS}}$ are introduced to represent percentages of maximum demand variations upward and downward, respectively. Load variations are limited in both directions by imposing constraints (6.5) and (6.6).

$$0 \le dVar_{z,t,c,w}^{\mathrm{LS,up}} \le \delta_{\mathrm{up}}^{\mathrm{LS}} D_{z,t}^{\mathrm{ELEC},c} \qquad z \in \mathcal{Z}, 1 \le t \le 24, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W} \qquad (6.5)$$

$$0 \le dVar_{z,t,c,w}^{\mathrm{LS,dw}} \le \delta_{\mathrm{dw}}^{\mathrm{LS}} D_{z,t}^{\mathrm{ELEC},c} \qquad z \in \mathcal{Z}, 1 \le t \le 24, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W} \qquad (6.6)$$

A second modeling option for the load shifting scheme is based on the introduction of demand functions and elasticities [115]. Specifically, let us denote by $\varepsilon_{\text{up}}^{\text{LS}}$ and $\varepsilon_{\text{dw}}^{\text{LS}}$ demand elasticities, which are negative parameters since price increases always lead to demand reductions. In such an approach, demand variations are computed according to the linear inverse demand functions expressed by equations (6.7) and (6.8).

$$dVar_{z,t,c,w}^{\text{LS,up}} \geq \varepsilon_{\text{up}}^{\text{LS}} D_{z,t}^{\text{ELEC},c} \left(\frac{pr_{z,t,w}^{c}}{PrRef_{z,t,w}^{c}} - 1 \right) \quad z \in \mathcal{Z}, 1 \leq t \leq 24, c \in \mathcal{C}^{y}, y \in \mathcal{Y}, w \in \mathcal{W} \qquad (6.7)$$

$$dVar_{z,t,c,w}^{\text{LS,dw}} \geq \varepsilon_{\text{dw}}^{\text{LS}} D_{z,t}^{\text{ELEC},c} \left(1 - \frac{pr_{z,t,w}^{c}}{PrRef_{z,t,w}^{c}} \right) \quad z \in \mathcal{Z}, 1 \leq t \leq 24, c \in \mathcal{C}^{y}, y \in \mathcal{Y}, w \in \mathcal{W} \qquad (6.8)$$

Specifically, demand functions (6.7) and (6.8) have a slope representing consumer elasticities and are derived from the reference point $\left(D_{z,t}^{\text{ELEC},c}, PrRef_{z,t,w}^{c}\right)$, representing the situation without demand response. Indeed, while parameter $D_{z,t}^{\text{ELEC},c}$ is the reference demand for electricity, $PrRef_{z,t,w}^{c}$ represents a reference price, which is computed by solving the model without demand response. Thus, the modeling approach based on elasticities requires solving twice the model: the first time with a fixed demand, so as to obtain estimates for reference prices, and the second time with flexible demand, so as to capture consumers reactions to electricity prices in the expansion planning framework.

Variables $pr_{z,t,w}^{c}$ in constraints (6.7) and (6.8) represent estimates for the zonal hourly electricity prices with demand response. When the computed price $pr_{z,t,w}^{c}$ is higher than the reference price $PrRef_{z,t,w}^{c}$, since elasticities are negative parameters, constraints (6.8) are activated and variables $dVar_{z,t,c,w}^{\text{LS,dw}}$ become positive to model the demand reduction. Instead, when the computed price is lower than the reference price, variables $dVar_{z,t,c,w}^{\text{LS,up}}$ become positive as constraints (6.7) are activated. Under the simplified assumption that prices for electricity are set only by thermal power plants, estimates for hourly prices with elastic demand $pr_{z,t,w}^{c}$ are computed as:

$$pr_{z,t,w}^{c} \geq CM_{k,y,w}\, \gamma_{k,t,w}^{c} + \frac{SUC_{k}}{24\overline{P}_{k}} \alpha_{k,t,w}^{c} \quad z \in \mathcal{Z}, k \in \Omega_{z}^{k}, 1 \leq t \leq 24, c \in \mathcal{C}^{y}, y \in \mathcal{Y}, w \in \mathcal{W}. \qquad (6.9)$$

Specifically, constraints (6.9) force the zonal price to be at least as high as the operating cost of the most expensive thermal power plant committed in the given market zone. Since the merit order of thermal power plants depends also on the start-up costs, the operating cost for each thermal power plant is computed by adding to its variable cost a term including the start-up costs, under the simplified assumption that start-up costs are recovered during one day.

Since the computed price could assume any value greater than the thermal power plants operating costs, a penalty term $PenPr$ is introduced to force the price to equal the operating cost of the most expensive thermal power plant committed. Therefore, objective function of the centralized approach (6.3) is modified as follows:

$$\begin{aligned} \min \; z^{\mathrm{LS}} \;=\; & z^{\mathrm{ante}} + \sum_{w\in\mathcal{W}} prob_w \left(\sum_{y\in\mathcal{Y}} \sum_{c\in\mathcal{C}^y} wg_c \sum_{t=1}^{24} \sum_{z\in\mathcal{Z}} Tran_z^{\mathrm{LS}} dVar_{z,t,c,w}^{\mathrm{LS,dw}} \right) \\ & + \; PenPr \sum_{w\in\mathcal{W}} prob_w \left(\sum_{y\in\mathcal{Y}} \sum_{c\in\mathcal{C}^y} wg_c \sum_{t=1}^{24} \sum_{z\in\mathcal{Z}} pr_{z,t,w}^{c} \right) \end{aligned} \tag{6.10}$$

Thus, the optimization model for the load shaving objective can be formulated as the set of constraints (5.2)–(5.38), (5.40)–(5.74), (6.1), (6.2) and (6.4)–(6.10).

6.3.3 Peak shaving

Peak shaving aims at reducing electricity consumption in peak load hours, when the most expensive generating plants have to be committed, consequently determining the highest cost for the energy system. Thus, as opposed to load shifting, only downward variations are considered in the peak shaving scheme by introducing decision variables $dVar_{z,t,c,w}^{\mathrm{PS,dw}}$. The flexible demand for electricity can be then computed as the difference between the reference demand and the downward variation as expressed by constraints (6.11).

$$dFlex_{z,t,w}^{c} = D_{z,t}^{\mathrm{ELEC},c} - dVar_{z,t,c,w}^{\mathrm{PS,dw}} \qquad z\in\mathcal{Z}, 1\le t\le 24, c\in\mathcal{C}^y, y\in\mathcal{Y}, w\in\mathcal{W} \tag{6.11}$$

Similarly to load shifting, also the customers reactions in the peak shaving scheme can be modeled using demand functions. However, since only demand reductions are allowed in the peak shaving strategy, only the following equation is needed:

$$dVar_{z,t,c,w}^{\mathrm{PS,dw}} \ge \varepsilon_{\mathrm{dw}}^{\mathrm{PS}} D_{z,t}^{\mathrm{ELEC},c} \left(1 - \frac{pr_{z,t,w}^{c}}{PrRef_{z,t,w}^{c}} \right)$$

$$z \in \mathcal{Z}, 1 \leq t \leq 24, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W} \qquad (6.12)$$

with $\varepsilon_{\mathrm{dw}}^{\mathrm{PS}}$ being the demand elasticity for the peak shaving strategy.

Similarly to the load shifting case, load decreases are limited by introducing parameter $\delta_{\mathrm{dw}}^{\mathrm{PS}}$, representing the percentage of maximum downward demand variations, an by imposing the following constraints:

$$0 \leq dVar_{z,t,c,w}^{\mathrm{PS,dw}} \leq \delta_{\mathrm{dw}}^{\mathrm{PS}} D_{z,t}^{\mathrm{ELEC},c} \qquad z \in \mathcal{Z}, 1 \leq t \leq 24, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W}. \qquad (6.13)$$

Finally, the penalty term $Tran_z^{\mathrm{PS}}$ is introduced to penalize the downward reduction of the demand. The new objective function is expressed by equation (6.14).

$$\min \; z^{\mathrm{PS}} \;=\; z^{\mathrm{ante}} + \sum_{w \in \mathcal{W}} prob_w \left(\sum_{y \in \mathcal{Y}} \sum_{c \in \mathcal{C}^y} wg_c \sum_{t=1}^{24} \sum_{z \in \mathcal{Z}} Tran_z^{\mathrm{PS}} dVar_{z,t,c,w}^{\mathrm{PS,dw}} \right) \qquad (6.14)$$

Thus, the optimization model for the peak shaving objective can be formulated as the set of constraints (5.2)–(5.38), (5.40)–(5.74), (6.2), (6.9) and (6.11)–(6.14).

6.3.4 Planning investments in demand response devices

Demand response programs require installing equipment in the point of consumption and a communication and control infrastructure [115]. Such an equipment includes smart meters used to receive and send price signals and electrical devices reacting to those price signals by reducing or shifting their electric consumptions. In this paragraph, we propose an optimization model to plan investments in demand response devices considering both the load shifting and the peak shaving objectives.

Let us denote with $v_{z,0}^{\mathrm{LS}}$ and $v_{z,0}^{\mathrm{PS}}$ the total installed capacity of electrical devices for load shifting and peak shaving, respectively, in zone z at the beginning of the planning horizon. Decisions to be taken concern the new installed capacity for electrical devices performing load shifting and peak shaving in each system zone and for every year of the planning horizon, represented through continuous variables $x_{z,y}^{\mathrm{LS}}$ and $x_{z,y}^{\mathrm{PS}}$. The total available capacity $v_{z,y}^{\mathrm{LS}}$ for load shifting in a given year y of the planning horizon can be computed as:

$$v_{z,y}^{\mathrm{LS}} = v_{z,0}^{\mathrm{LS}} + \sum_{i=1}^{y} x_{z,y}^{\mathrm{LS}} \qquad z \in \mathcal{Z}, y \in \mathcal{Y}. \tag{6.15}$$

Both the yearly installed capacity and the total available capacity are limited by maximum values $X_{z,y}^{\mathrm{LS}}$ and $V_{z,y}^{\mathrm{LS}}$, as expressed by constraints (6.16) and (6.17).

$$0 \leq x_{z,y}^{\mathrm{LS}} \leq X_{z,y}^{\mathrm{LS}} \qquad z \in \mathcal{Z}, y \in \mathcal{Y} \tag{6.16}$$

$$0 \leq v_{z,y}^{\mathrm{LS}} \leq V_{z,y}^{\mathrm{LS}} \qquad z \in \mathcal{Z}, y \in \mathcal{Y} \tag{6.17}$$

Similarly, the available capacity $v_{z,y}^{\mathrm{PS}}$ for peak shaving in a given year y of the planning horizon is computed as:

$$v_{z,y}^{\mathrm{PS}} = v_{z,0}^{\mathrm{PS}} + \sum_{i=1}^{y} x_{z,y}^{\mathrm{PS}} \qquad z \in \mathcal{Z}, y \in \mathcal{Y}. \tag{6.18}$$

Upper bounds $X_{z,y}^{\mathrm{PS}}$ and $V_{z,y}^{\mathrm{PS}}$ are then imposed to limit both the yearly installed capacity and the total available capacity, as expressed by the following constraints.

$$0 \leq x_{z,y}^{\mathrm{PS}} \leq X_{z,y}^{\mathrm{PS}} \qquad z \in \mathcal{Z}, y \in \mathcal{Y} \tag{6.19}$$

$$0 \leq v_{z,y}^{\mathrm{PS}} \leq V_{z,y}^{\mathrm{PS}} \qquad z \in \mathcal{Z}, y \in \mathcal{Y} \tag{6.20}$$

As regards to the load shifting, demand increases and decreases are determined according to demand functions (6.7) and (6.8), with computed prices $pr_{z,t,w}^{c}$ defined by equations (6.9) and reference prices $PrRef_{z,t,w}^{c}$ determined by solving the model without demand response. However, new constraints are needed to ensure consistency between the continuous variables representing the upward and downward load variations and those representing the investment decisions. Such a connection is modeled by replacing the upper bounds of constraints (6.5) and (6.6) with the total available capacity for load shifting, as stated by constraints.

$$0 \leq dVar_{z,t,c,w}^{\mathrm{LS,up}} \leq v_{z,y}^{\mathrm{LS}} \qquad z \in \mathcal{Z}, 1 \leq t \leq 24, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W} \quad (6.21)$$

$$0 \leq dVar_{z,t,c,w}^{\mathrm{LS,dw}} \leq v_{z,y}^{\mathrm{LS}} \qquad z \in \mathcal{Z}, 1 \leq t \leq 24, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W} \quad (6.22)$$

Finally, as explained in Section 6.3.2, equations (6.4) are imposed to ensure the daily balance between upward and downward load variations.

As regards to the peak shaving strategy, the decreases of the demand are computed by considering demand functions, according to equations (6.12). The maximum load reduction is then limited by the total available capacity for peak shaving devices, as expressed by constraints (6.23).

$$0 \leq dVar_{z,t,c,w}^{\mathrm{PS,up}} \leq v_{z,y}^{\mathrm{PS}} \qquad z \in \mathcal{Z}, 1 \leq t \leq 24, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W} \quad (6.23)$$

In the presence of both load shifting and peak shaving strategies, the flexible demand, which is a decision variable rather than an input parameter, is computed by adding to the electrical load the upward variations due to load shifting and by subtracting the downward variations due to load shifting and peak shaving, as stated by equation (6.24).

$$dFlex_{z,t,w}^{c} = D_{z,t}^{\mathrm{ELEC},c} + dVar_{z,t,c,w}^{\mathrm{LS,up}} - dVar_{z,t,c,w}^{\mathrm{LS,dw}} - dVar_{z,t,c,w}^{\mathrm{PS,dw}}$$
$$z \in \mathcal{Z}, 1 \leq t \leq 24, c \in \mathcal{C}^y, y \in \mathcal{Y}, w \in \mathcal{W} \quad (6.24)$$

Finally, the new objective function z^{DR} to be minimized can be formulated as:

$$\begin{aligned}
\min z^{\mathrm{DR}} \;=\;& z^{\mathrm{ante}} \\
&+ \sum_{y \in \mathcal{Y}} \sum_{z \in \mathcal{Z}} \left(\frac{IC_z^{\mathrm{LS}} x_{z,y}^{\mathrm{LS}}}{(1+r)^{y-y_0}} + \frac{IC_z^{\mathrm{PS}} x_{z,y}^{\mathrm{PS}}}{(1+r)^{y-y_0}} \right) \\
&+ \sum_{w \in \mathcal{W}} prob_w \left(\sum_{y \in \mathcal{Y}} \sum_{c \in \mathcal{C}^y} wg_c \sum_{t=1}^{24} \sum_{z \in \mathcal{Z}} Tran_z^{\mathrm{LS}} dVar_{z,t,c,w}^{\mathrm{LS,dw}} + Tran_z^{\mathrm{PS}} dVar_{z,t,c,w}^{\mathrm{PS,dw}} \right) \\
&+ PenPr \sum_{w \in \mathcal{W}} prob_w \left(\sum_{y \in \mathcal{Y}} \sum_{c \in \mathcal{C}^y} wg_c \sum_{t=1}^{24} \sum_{z \in \mathcal{Z}} pr_{z,t,w}^{c} \right) \qquad (6.25)
\end{aligned}$$

Specifically, the objective function (6.25) comprises four terms: (i) the system cost function without demand management resulting from equation (5.1); (ii) the annualized investment costs in new electrical devices to perform load shifting and peak shaving; (iii) the operational costs related to the downward variations in both the load shifting and the peak shaving; and (iv) the penalties for prices, which are imposed to force the computed prices to be equal to the operating cost of the most expensive thermal power plant committed, as explained in the previous sections.

Thus, our model to plan investments in electrical devices performing load shifting and peak shaving can be formulated as the set of constraints (5.2)–(5.38), (5.40)–(5.74), (6.2), (6.4), (6.7)–(6.9), (6.12) and (6.15)–(6.25).

6.4 Numerical tests

The proposed method has been applied to a scenario elaborated by CESI S.p.A. in order to evaluate the consistency of the equations describing the demand reactions to electricity signals. Numerical tests described in this section have been conducted for validation purposes only and do not represent a real analysis of any specific energy scenario. Indeed, the application to a real case study of the models introduced in this chapter requires performing extensive studies, which are mainly needed to compute suitable values for elasticities according to the type of consumer (e.g., domestic, commercial or residential) and to the direction of the demand variation. Peak shaving and load shifting potential as well as data about elasticities and costs for the Italian power system cannot be derived by other works in the literature, as the studies may be very specific for a certain region and therefore may not be directly applicable to other regions. Studies to compute suitable values for demand response potential and costs for the Italian power system will represent a natural evolution of the research activities described in this dissertation.

The scenario analysed in this chapter is based on the Italian power system, consisting of six market zones: North, Central-North, Central-South, South, Sicily, and Sardinia. 180 thermal power plants divided into 150 existing units and 30 candidate projects are considered in the analysis, along with 9 existing transmission lines, 15 candidate lines, 16 equivalent hydropower plants, and 13 candidate storage facilities. A planning horizon of 11 years, from 2020 up to 2030, is simulated in the tests, while a single scenario for fuel prices and emission costs is considered.

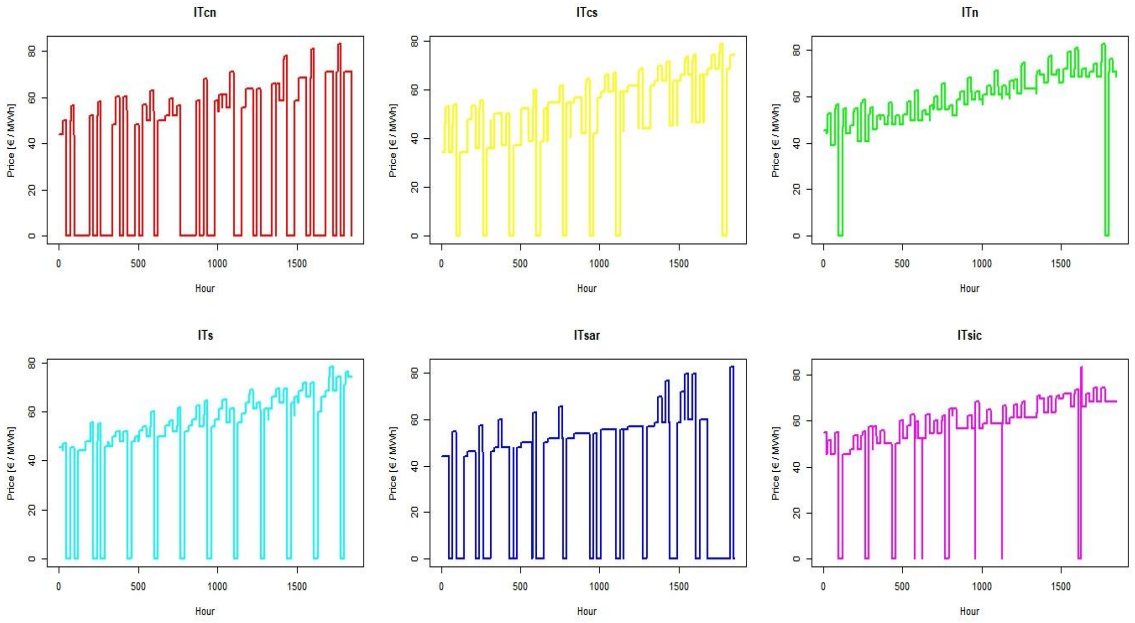


Fig. 6.1 Zonal electricity prices computed from the solution of the model without demand response

To model power system operations with an hourly resolution, the clustering analysis described in Section 2.3.2 has been applied by imposing an input threshold of 2.5%, obtaining seven representative days for each year, corresponding to a planning horizon of 1848 hours.

As explained in the previous section, when modeling demand variations by introducing elasticities and demand functions, the optimization model must be solved twice. The first time with a fixed demand to obtain estimates of electricity prices and the second time with elastic demand. Fig. 6.1 illustrates the prices for electricity in each system zone computed from the first solution of the model without considering demand side management programs. As can be observed, the zonal prices show an increasing trend over the planning horizon, which depends on the values assigned to fossil fuel prices, linearly increasing over the years. These prices have been then used as reference prices to drive demand variations by testing several models.

6.4.1 Load shifting

In the first numerical test we performed, we focused on the load shifting objective by executing the optimization model introduced in Section 6.3.2. Specifically, values of -1% and -1.5% have been assigned to elasticities $\varepsilon_{\mathrm{up}}^{\mathrm{LS}}$ and $\varepsilon_{\mathrm{dw}}^{\mathrm{LS}}$, respectively. Indeed, elasticities used to compute demand reductions are higher, since they refer to peak-load hours, which are characterized by much higher prices than those observed in hours with lower demand. An hourly limit of 7% for shiftable demand has been then imposed to bound the demand variations in both directions.

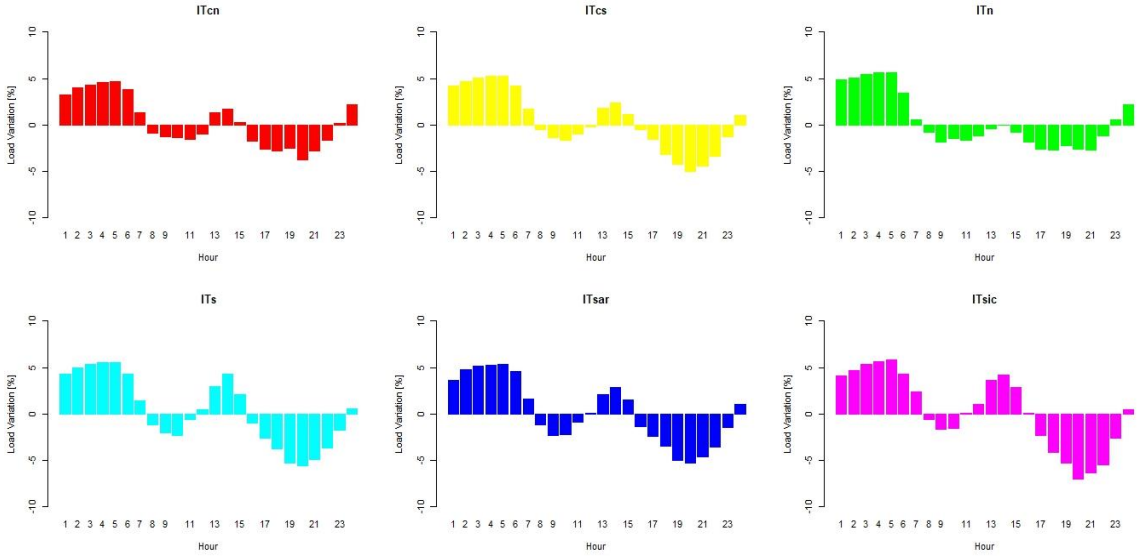


Fig. 6.2 Average relative demand shifting in each system zone

Fig. 6.2 illustrates for every system zone the average relative demand variations caused by the load shifting. As can be observed, in all system zones the load is mainly shifted from the two daily demand peaks to the night hours. Moreover, in southern regions more demand is shifted between the second load peak and the valley between the two peaks during the day, so as to flatten the load profile. For instance, Fig. 6.3 shows the total amount of demand shifted over the day in Sicily, by highlighting the original load $D_{z,t}^{\mathrm{ELEC},c}$, the flexible demand $dFlex_{z,t,w}^{c}$ and upward and downward load variations $dVar_{z,t,c,w}^{\mathrm{LS,up}}$ and $dVar_{z,t,c,w}^{\mathrm{LS,dw}}$.

Fig. 6.4 provides the demand shifting strategy for all the Italian market zones.

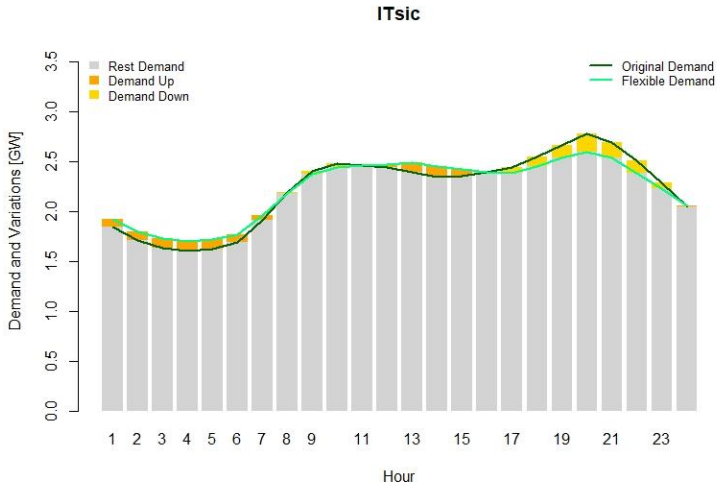


Fig. 6.3 Demand shifting in Sicily

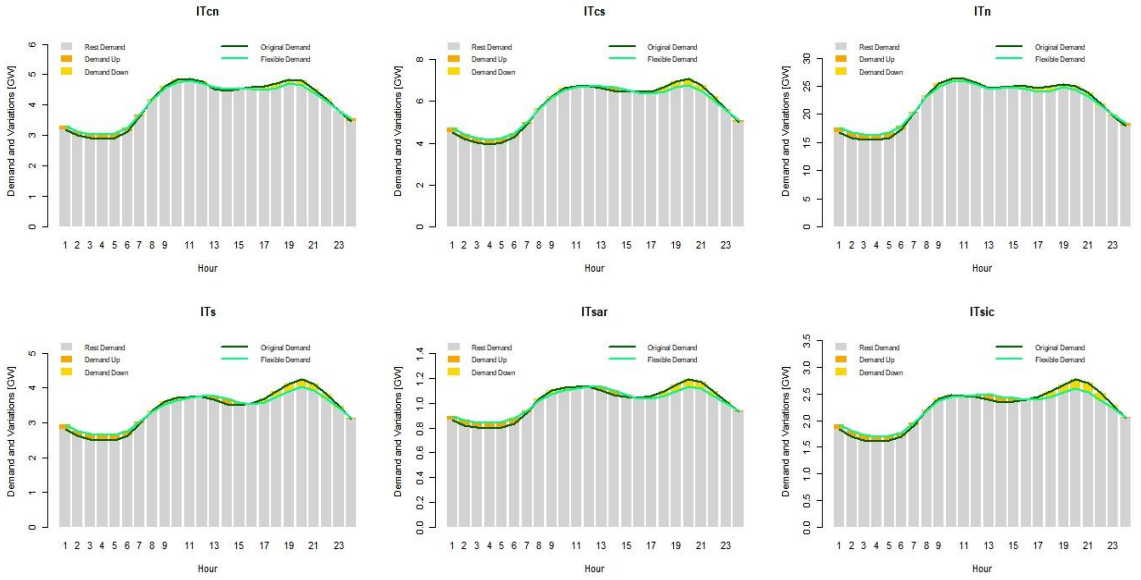


Fig. 6.4 Demand shifting in each Italian market zone

6.4.2 Peak shaving

In the second numerical test, we focused on the peak shaving objective by executing the optimization model introduced in Section 6.3.3. Elasticity ε_{dw}^{PS} for the peak shaving case have been set to -0.5%, which is a lower value than those related to the load shifting as in the peak shaving strategy the demand cannot be moved to another hour. Similarly to the load shifting test, an hourly limit of 7% for curtailable demand has been imposed to bound the downward demand variations.

Fig. 6.5 illustrates for every system zone the average relative demand variations caused by the peak shaving. As can be observed, as opposed to the previous case, only downward variations are allowed. Specifically, load reductions take place at all hours of the day, though to different degrees. In particular, the greatest reductions are mainly concentrated in the two peak hours of the day, while load decreases are smaller during both the night and the valley between the two peaks. Indeed, high load hours may be characterized by extremely high prices for electricity, discouraging the realization of some consumptions. On the other hand, during low load hours, prices are generally lower, resulting in few consumption reductions. Results with the peak shaving objective are similar to those obtained by considering the load shifting: a flattening of demand to reduce load ramps and obtain a smoother profile, as shown in Fig. 6.6 and Fig. 6.7.

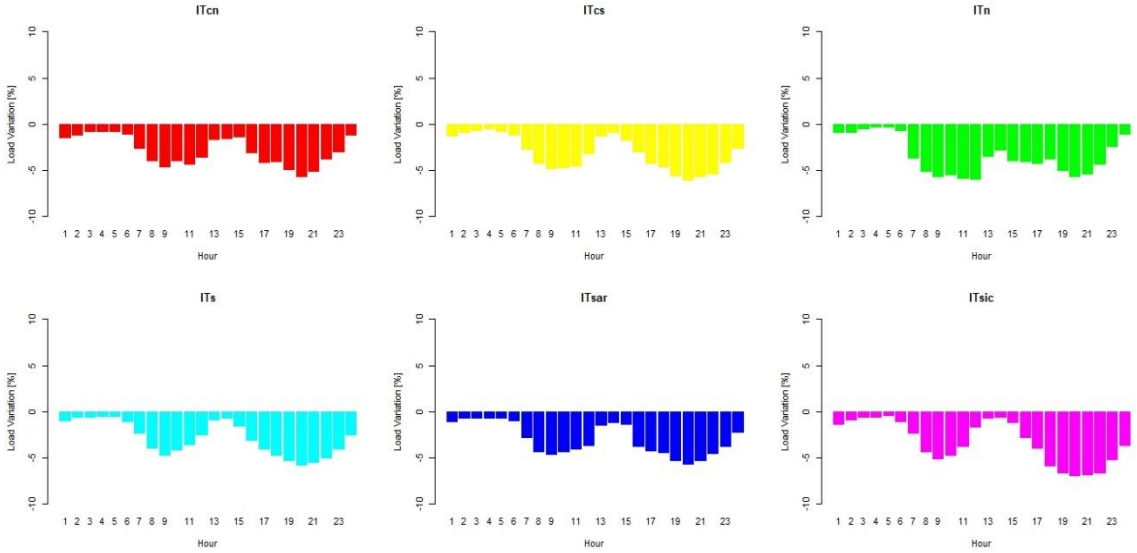


Fig. 6.5 Average relative peak shaving in each system zone

Specifically, while Fig. 6.6 shows the original load $D_{z,t}^{\mathrm{ELEC},c}$, the flexible demand $dFlex_{z,t,w}^{c}$ and the downward load variations $dVar_{z,t,c,w}^{\mathrm{PS,dw}}$ in Sicily, Fig. 6.7 provides the same information for all the Italian system zones.

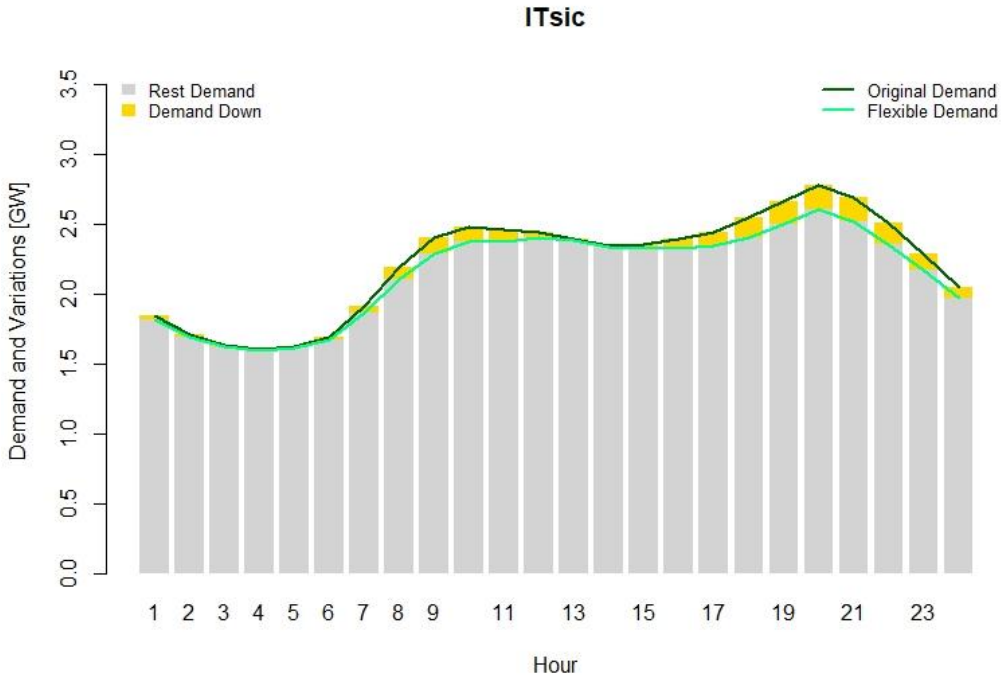


Fig. 6.6 Peak shaving in Sicily

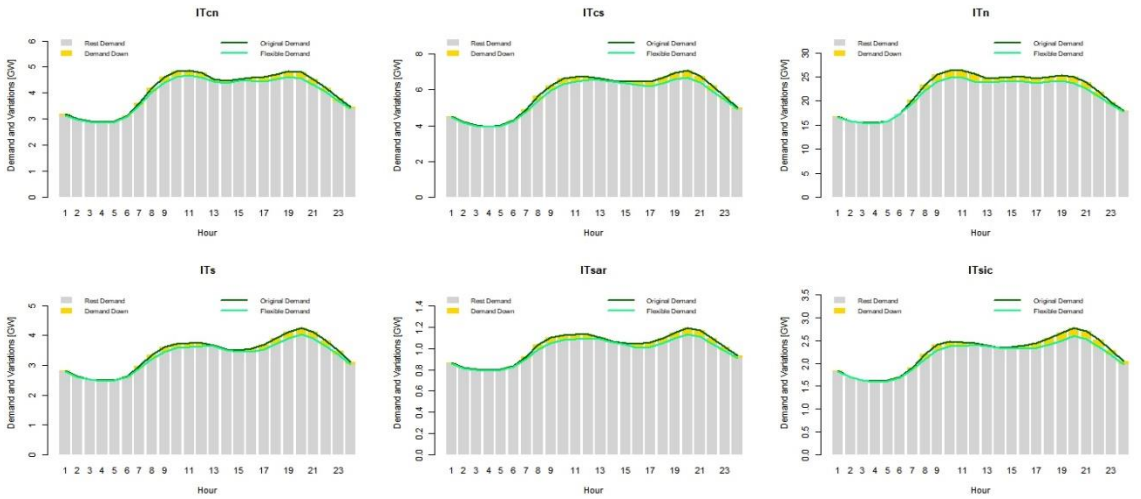


Fig. 6.7 Peak shaving in each Italian market zone

6.4.3 Investments in demand response devices

Finally, a numerical test has been performed also to validate the optimization model presented in Section 6.3.4. As explained in the previous section, we set the elasticity used to compute demand increase to -1%, while a value of -1.5% has been seleceted to compute demand reductions in the load shifting. Elasticity for load reduction in the peak shaving strategy has been fixed at -0.5%, as in this case the demand cannot be moved to other hours. Moreover, we considered an hourly upper bound of 5% for both the shiftable and the curtailable demand at the beginning of the planning horizon. However, also the possibility to invest in new electrical devices to increase both the shiftable and the curtailable demand has been included in such a test. Only for validation purposes, a very high potential for demand response technology has been considered in this analysis. Specifically, investments in electrical devices to shift up to 30% of the hourly demand have been considered in such a scenario, as well as investments to increase at 10% the hourly curtailable load.

Results provided by the application of the model in such a scenario are illustrated in Fig. 6.8. Specifically, the model suggests increasing the load shifting capacity to 5 GW in the North and to 0.6 GW in Sicily.

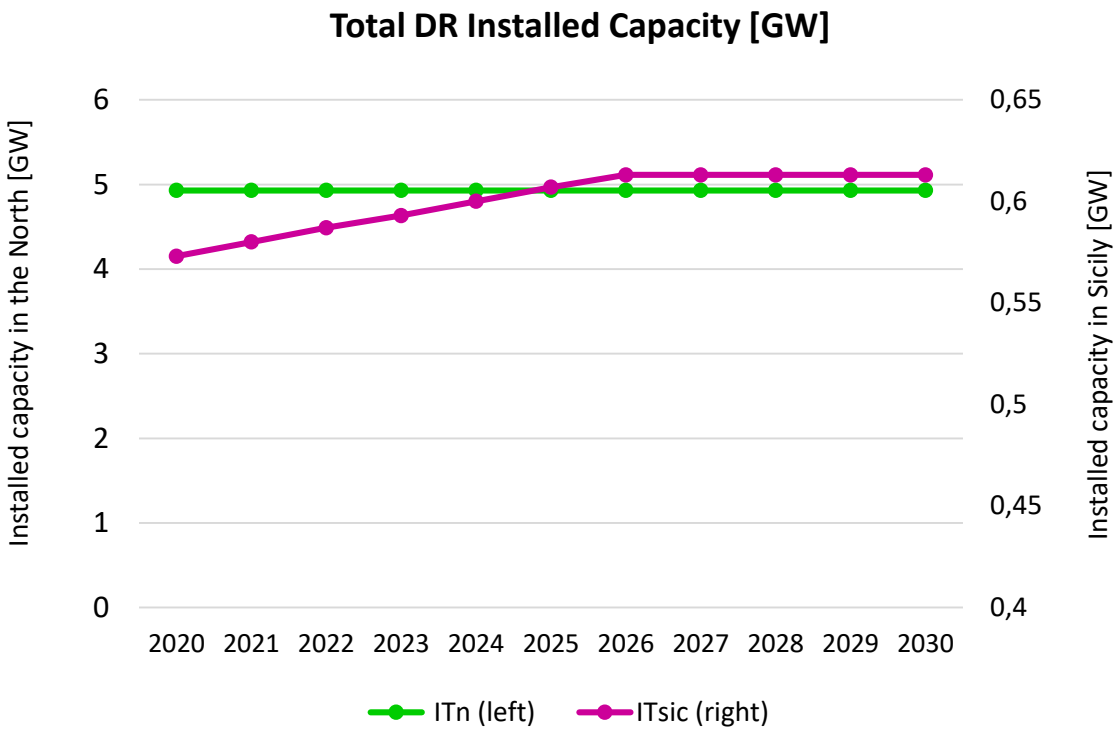


Fig. 6.8 Installed capacity [GW] for electrical devices performing load shifting

For each year of the planning horizon, the average relative demand variations induced by demand response devices installed in the North zone are represented in Fig. 6.9. As can be observed, investments in demand response devices increase at 17% the hourly limit to load variations in the North, where the demand for electricity is mainly shifted from the daytime hours to the night hours

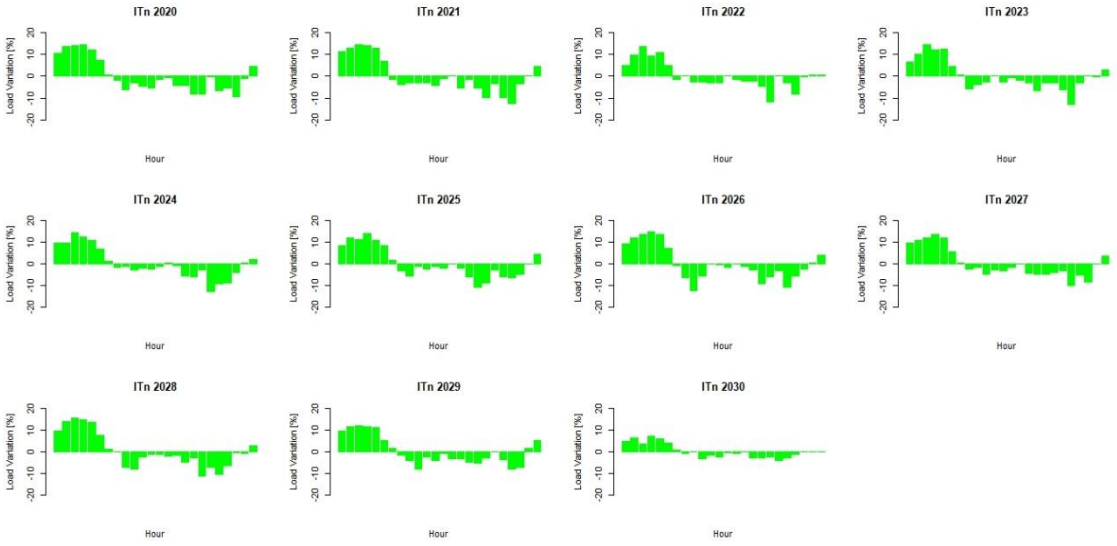


Fig. 6.9 Average relative demand variations in the North in each year of the planning horizon

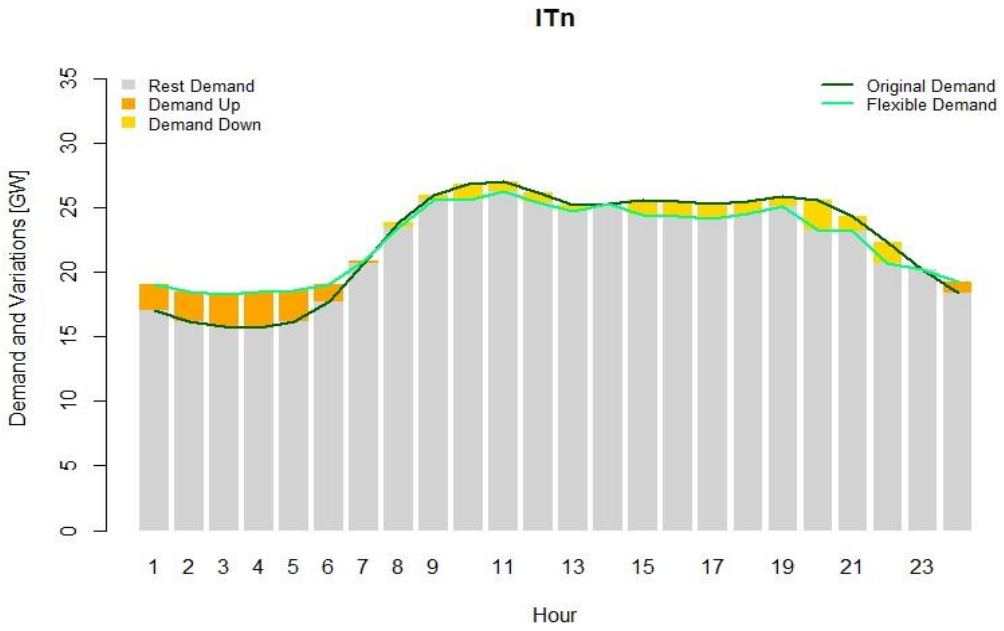


Fig. 6.10 Demand variations in the North

Fig. 6.10 shows the total amount of demand shifted over an average day in the North, by highlighting the original load, the flexible demand and upward and downward load variations. As can be noticed, demand response programs in the North aims at flattening the demand profile so as to obtain a significant cost saving by substituting more expensive energy with cheaper energy and by reducing load ramps, obtaining a smoother profile.

Instead, Fig. 6.11 illustrates for each year of the planning horizon the average relative demand variations induced by demand response devices installed in Sicily.

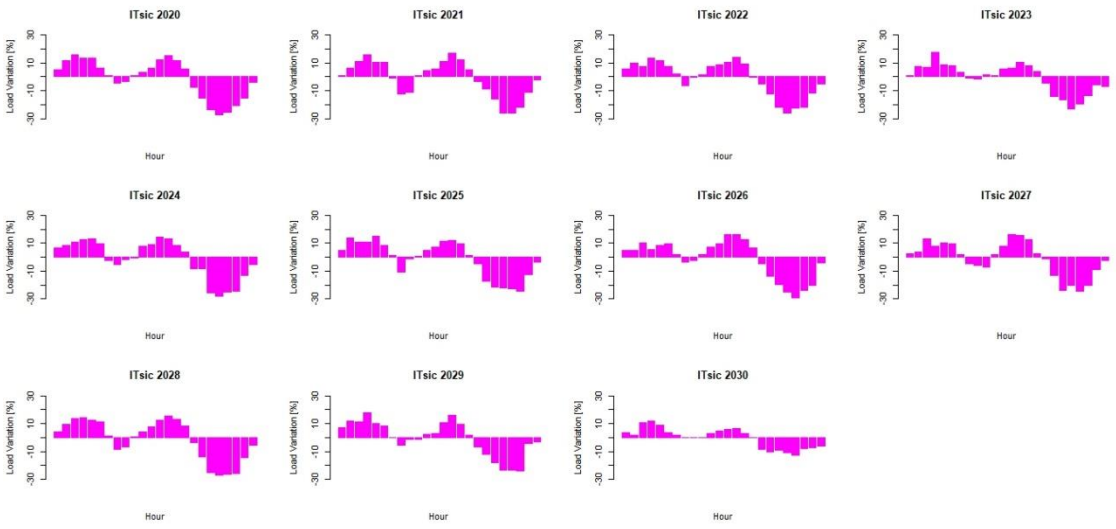


Fig. 6.11 Average relative demand variations in Sicily in each year of the planning horizon

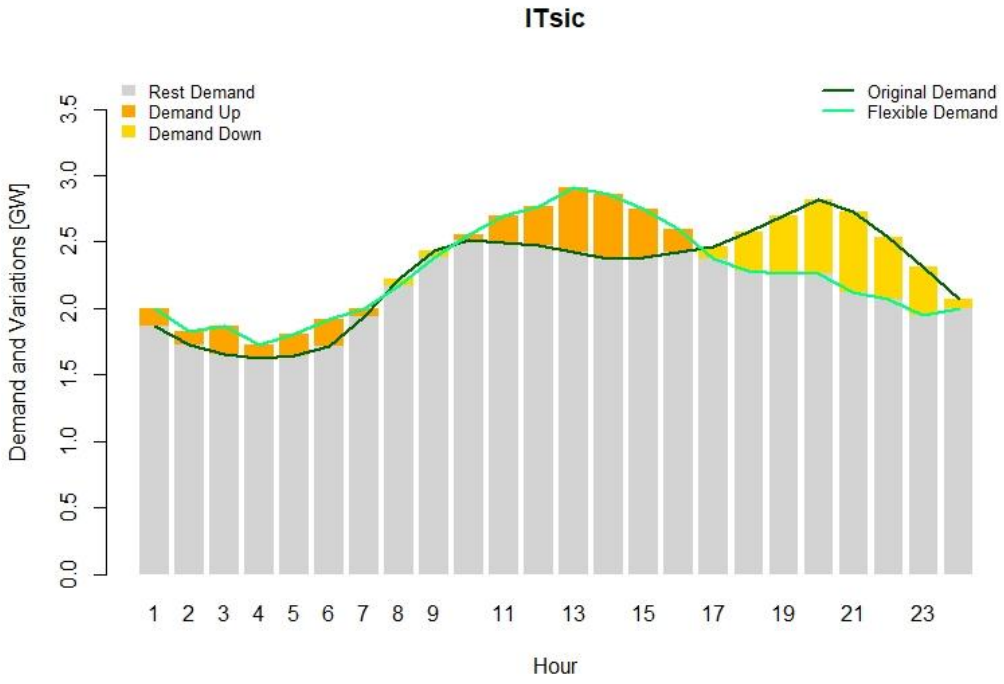


Fig. 6.12 Demand variations in Sicily

As can be noticed, investments in demand response devices increase at 30% the hourly limit to load variations, being most of the demand for electricity shifted between the second load peak and the valley between the two peaks during the day. The total amount of demand shifted over an average day in Sicily, as well as the original load and the flexible demand, are shown in Fig. 6.12. As can be observed, demand response programs dramatically change the demand curve in Sicily, by creating a profile with a central peak. These results may be explained by analysing investment decisions in new renewable power generation provided by the model, which consist in installing large shares of solar power in Sicily. By shifting the demand from the second daily peak to the valley between the two original peaks, demand response devices transform the load profile, which becomes similar to the solar power generation, promoting the integration of the large share of solar PV technology installed in the zone.

6.5 Chapter conclusions

In this chapter, the problem of including in the expansion planning framework demand variations to electricity prices has been addressed. First, a definition of demand response has been provided and several benefits related to the possibility of change load profiles have been pointed out. Such benefits include expected bill savings for all electrical

consumers, lower variability in the spot market and a reduction of CO_2 emissions, which could facilitate the optimal transition to a European low emission power system.

Then, a comprehensive review of works integrating the demand response into energy planning models has been provided. All the studies in the literature show how demand response could decrease peak demand, emissions and system costs, while increasing the integration of renewables in power system. However, similar outcomes may be provided by other technologies which may compete with demand response, such as batteries, hydro pumping units or fast ramping power plants. The lack of articles on this topic calls for the development of an integrated planning framework that could accurately consider the competition between different resources providing flexibility to power systems.

Thus, our approach to include customers' reactions to price signals in the GTEP analysis by considering elasticities and demand functions has been presented. Specifically, three different models have been proposed in the chapter. First, we modeled the load shifting objective, which consists in flattening the demand profile by transferring load from peak hours to off-peak hours. Second, the peak shaving strategy has been considered. Third, we presented an optimization model to plan investments in demand response devices.

Finally, some tests have been conducted to validate the consistency of the equations describing the demand reactions to electricity signals. The obtained results show peak reduction and valley filling effects, demonstrating that demand response could facilitate the integration of renewable energy and reduce system costs. However, numerical tests described in this section have been conducted for validation purposes only and do not represent a real analysis of any specific energy scenario. Indeed, the application to a real case study of the models introduced in this chapter requires performing extensive studies to evaluate potential and costs for demand response programs for the Italian power system. Such an analysis will represent a natural evolution of the research activities described in this dissertation.

Chapter 7

Conclusions

In this thesis, the GTEP problem to facilitate the transition to low emission power systems has been addressed. Meeting the challenging sustainability goals set by the European Commission requires increasing the share of power capacity from renewable power sources, as well as their penetration in the energy mix. However, in the presence of high levels of renewable penetration, power systems face great challenges to meet the demand, because of the unpredictable daily and seasonal nature of renewable generation. The variability of these intermittent power sources has to be managed in the short-term to ensure that power systems are operated in an efficient and reliable way.

In this dissertation, we have presented several optimization models to plan the joint expansion of power systems so as to achieve long-term policy goals, while considering the challenges related to integrating large shares of renewables, which require considering a very detailed representation of power systems short-term operations. Such models optimize strategic decisions including retirement of existing capacity and investments in new generation, transmission and storage facilities, as well as operational decisions. However, due to the high costs and the long lifetimes, investment decisions in the power sector are long-term decisions, which require considering planning horizon of several decades. Due to the long-term horizon, providing an hourly solution to the expansion planning problem is computationally infeasible.

To provide reliable expansion plans for large-scale energy systems with high shares of renewables, we included in the analysis a high level of technical detail by considering a computational efficient formulation for the unit commitment problem, while a high level of temporal detail has been obtained by working with representative days, discretized in hours, selected through a novel approach. The use of representative days raises the crucial issues regarding how to consider the seasonality of hydroelectric dispatch and how to set the initial ON/OFF status of thermal power plants in representative days. Both these crucial issues have been addressed in this thesis by proposing different novel methods.

Specifically, as regards to the problem of capturing the seasonality of the hydroelectric dispatch when working with representative days, two different approaches

have been proposed. First, we designed a QMIP model to perform a preliminary hydroelectric dispatch. Second, we proposed to include inter-day equations in the optimization model to create continuity in storage operation across the entire planning horizon. Instead, as regards to the determination of the initial statuses of thermal power plants in representative days, we developed a novel approach based on the application of classification techniques. Decision trees showed to be the best classifier, providing accurate estimates of initial statuses also in scenarios very different from the dataset used for the tree induction.

Another distinct feature of the GTEP problem is the high level of uncertainty. Indeed, since expansion plans are usually provided for a long-term planning horizon, the future system conditions are generally uncertain at the time the expansion plans are decided. Different sources of uncertainty may affect planning decisions and must be considered in the decision-making process. Specifically, while short-term uncertainty can be captured by accurately selecting the representative days, long-term uncertainty can be included in the decision making framework by means of a two-stage stochastic programming approach. Thus, in this thesis we proposed a two-stage stochastic MILP model to plan the investment decisions in the power sector, while providing more reliable decisions considering different scenarios for fuel prices and CO_2 price. In such an approach, investment and decommissioning decisions represent the first stage, while operational decisions are second-stage variables.

However, the inclusion of the long-term uncertainty in the analysis further increases the complexity of the problem, making the real-scale GTEP problems computationally intractable. To solve such problems, efficient solution algorithms that exploit the decomposable structure of two-stage stochastic programming models have to be applied. Thus, in this thesis we also implemented a solution algorithm based on the multi-cut Benders decomposition strategy to decompose the stochastic model both by year and by scenario. In our model, the first-stage variables represent investment and decommissioning decisions. If these variables are fixed, the original problem decomposes into a set of independent subproblems, one per year and scenario, each representing the operation in the second stage. In our solution algorithm, the two-stage stochastic problem is replaced by an iterative collection of smaller problems. At each iteration, the so-called master problem is solved first to determine suitable values for the first-stage variables. Once the investment schedule is determined, the subproblems are solved. Finally, the dual information of the subproblems is sent to the master problem employing a cut to update the master problem solution.

Beside storage technologies, natural gas represents an existing option to cope with the variability of renewable power sources. Specifically, the energy conversion between electricity and natural gas is bi-directional. First, natural gas can be converted into electricity by gas-fired power plants that can be fired up in just a few minutes, therefore being suited to compensate for the variability of renewable generation. Secondly, electricity can be transformed into gas by the PtG technology, which provides an opportunity to integrate large shares of renewables by converting surplus renewable power generation to gas fuel that can be stored locally to be used later or injected in the natural gas network. The deployment of both gas-fired power plants and PtG increases the interconnection between electricity and gas systems and requires an integrated planning framework that could accurately consider this coupling.

Thus, in this dissertation also the joint expansion planning problem for real-scale integrated electricity and gas systems has been addressed, by proposing a novel stochastic long-term expansion co-planning framework for integrated systems with bi-directional energy conversion. Such a model co-optimizes strategic investment decisions in new generating capacity, transmission lines, storage facilities, pipelines, and PtG capacity, as well as power system and gas system operational decisions. Interconnections between electricity and gas systems are taken into account by considering both the gas demand not related to the thermal power generation (which is an input parameter) and the gas demand related to thermal power generation (which instead is a model variable) and by modeling the energy conversion of electricity into gas operated by the PtG plants.

Moreover, a detailed and realistic case study concerning the Italian energy system has been presented in this thesis. In our numerical study, we simulated a planning horizon of 21 years, from 2020 up to 2040, and we planned the joint expansion of the Italian electricity and gas system under CO_2 and fuel prices uncertainty. Empirical results show the need to install large shares of renewables to achieve policy goals set by the European Commission, with the solar technology preferred to the wind power technology due to the lower investment costs. The huge solar penetration would require installing new storage systems and reinforcing the transmission network, by building new national and international cross border transmission lines, so as to better exploit the intermittent renewable energy sources and compensate for the decommissioning of some Italian thermoelectric plants. As regards to the natural gas system, the decommissioning of all coal and oil thermal power plants would require the availability of an additional amount of gas and therefore the development of the so called “GALSI”, from Algeria to Sardinia and to Italy mainland, starting from year 2022.

Finally, also the possibility to control the variability of intermittent renewable power sources by adjusting the demand for electricity has been analysed in this dissertation. Specifically, we considered in the GTEP framework customers' reactions to electricity prices, which include the reduction of the electricity consumption during peak hours and the demand shifting from peak hours to off-peak hours if consumption is inevitable. We then tried to fill the existing gap in the literature by proposing an integrated planning framework that could accurately consider the competition between different resources providing flexibility to power systems, i.e., storage systems, transmission lines, fast-ramping power plants and demand response devices. Results of several tests performed showed how demand response could facilitate the integration of renewable energy and reduce system costs.

However, numerical tests for demand response have been conducted for validation purposes only and do not represent a real analysis of any specific energy scenario. Indeed, extensive preliminary studies are needed to apply the proposed method to a real energy scenario in order to evaluate potential and costs for demand response programs as well as to determine suitable values for elasticities according to the type of consumer (e.g., domestic, commercial or residential) and to the direction of the demand variation. Studies to assess demand response potential and costs in Italy so as to perform a realistic GTEP analysis for the Italian energy system while considering the competition between different resources providing flexibility will represent a natural evolution of the research activities described in this dissertation.

Moreover, optimization models presented in this work are based on a centralized approach: by assuming the perspective of a single central entity, we plan the joint expansion of generation and transmission facilities so as to minimize the total system costs. Such an approach is justified by the objective of our research, which is to develop a computational tool that could support regulators in searching for optimal policies, focusing on the inclusion of as many engineering details as possible, while neglecting market aspects. A different approach to the GTEP problem will be examined in a future work, by developing decentralized models that could represent the interactions between different agents involved in the liberalized power sector.

Another suggestion for future research is the acceleration of the Benders algorithm. Indeed, it is well known that the application of the classical Benders decomposition sometimes leads to slow convergence and long computing times. Several techniques have been proposed to deal with this phenomenon and to accelerate the standard Benders method [126]. These techniques can be divided into two categories. The first approach aims at reducing the cost of each iteration, by reducing the time spent solving the master

problem or the subproblem for instance by only approximately solving the master problem. The second strategy instead aims at reducing the number of iterations by generating more efficient cuts. Both approaches will be investigated in a future research.

Appendix

Dissemination Activities

This section provides the list of dissemination activities connected to the thesis work, which include publications in international journals and conference proceedings, as well as presentations at international conferences.

Specifically, paper already published or accepted for publication include:

- Micheli, G., Soda, E., Vespucci, M., Gobbi, M. and Bertani, A., "Big data analytics: an aid to detection of non-technical losses in power utilities", Computational Management Science, June 2018, doi:10.1007/s10287-018-0325-x.

- Micheli, G., Vespucci, M.T., Stabile M., Puglisi C. and Ramos, A., "A two-stage stochastic MILP model for generation and transmission expansion planning with high shares of renewables", Energy Systems, October 2020, doi:10.1007/s12667-020-00404-w.

- Micheli, G., Vespucci, M.T., Stabile, M. and Cortazzi, A., "Selecting and initializing representative days for generation and transmission expansion planning with high shares of renewables". In: Gentile, C., Stecca, G., Ventura, P (eds.) Graphs and combinatorial optimization: from theory to applications - CTW2020 proceedings. Springer, April 2021, ISBN 978-3-030-63071-3.

- Micheli, G. and Vespucci, M.T., "A Survey on Modeling Approaches for Generation and Transmission Expansion Planning Analysis", Accepted for publication in NAOV-2020 proceedings, 2021.

Moreover, the following paper based on the thesis work is in preparation for submission to an international journal:

- Micheli, G., Vespucci, M.T., Puglisi, C. and Cortazzi, A., "Long-Term Expansion Planning of Integrated Electricity and Gas Systems with High Shares of Renewables and Bi-Directional Energy Flows".

Presentations of the thesis work at international conferences include:

- “A deterministic model for generation and transmission expansion planning with high shares of renewables”. Presented at Energy Finance 4, February 4-5, 2019, Milano.

- “Generation and transmission expansion planning with high shares of renewables”. Presented at Computational Management Science, March 27-29, 2019, Chemnitz.

- “A two-stage stochastic programming approach for generation and transmission expansion planning with high shares of renewables”. Presented at International Conference on Stochastic Programming, July 29 – August 2, 2019, Trondheim.

- “A two-stage stochastic programming approach for generation and transmission expansion planning with high shares of renewables”. Presented at International Conference on Optimization and Decision Science, September 4-7, 2019, Genova.

- “A two-stage stochastic MILP model for generation and transmission expansion planning with high shares of renewables”. Presented at AIROYoung 2020, February 5-7, 2020, Bolzano.

- “Selecting and Initializing Representative Days for Generation and Transmission Expansion Planning with High Shares of Renewables”. Presented at 18th Cologne-Twente Workshop on Graphs and Combinatorial Optimization, September 14-16, 2020, Online.

- “Long-Term Expansion Planning of Integrated Electricity and Gas Systems with High Shares of Renewables and Bi-Directional Energy Flows”. Presented at Energy Finance 6, February 22-23, 2021, Online.

Bibliography

[1] E. Sauma and S. Oren, "Economic Criteria for Planning Transmission Investment in Restructured Electricity Market," *IEEE Transactions on Power Systems,* vol. 22, no. 4, pp. 1394-1405, 2007.

[2] D. Pozo, J. Contreras and E. Sauma, "If you build it, he will come: Anticipative power transmission planning," *Energy Economics,* vol. 36, pp. 135-146, 2013.

[3] V. Krishnan, J. Ho, F. Hobbs, A. L. Liu, J. D. McCalley, M. Shahidehpour and Q. P. Zheng, "Co-optimization of electricity transmission and generation resources for planning and policy analysis: review of concepts and modeling approaches," *Energy Systems,* vol. 7, no. 2, pp. 297-332, 2016.

[4] D. Huppmann and R. Egging-Bratseth, "Market power, fuel substitution and infrastructure – A large-scale equilibrium model of global energy markets," *Energy,* vol. 75, no. C, pp. 483-500, 2014.

[5] P. Seljom and A. Tomasgard, "Short-term uncertainty in long-term energy system models — A case study of wind power in Denmark," *Energy Economics,* vol. 49, pp. 157-167, 2015.

[6] L. L. Garver, "Transmission Network Estimation Using Linear Programming," *IEEE Transactions on Power Apparatus and Systems,* vol. 89, no. 7, pp. 1686-1697, 1970.

[7] Y. Gu, J. McCalley and M. Ni, "Coordinating large-scale wind integration and transmission planning," *IEEE Transactions on Sustainable Energy,* vol. 3, no. 4, pp. 652-659, October 2011.

[8] "Bilevel Conic Transmission Expansion Planning," *IEEE Transactions on Power Systems,* vol. 33, no. 4, pp. 4640-4642, 2018.

[9] N. E. Koltsaklis and A. S. Dagoumas, "State-of-the-art generation expansion planning: A review," *Applied Energy,* vol. 230, pp. 563-589, 2018.

[10] R. Hemmati, H. Saboori and M. A. Jirdehi, "Multistage generation expansion planning incorporating large scale energy storage systems and environmental pollution," *Renewable Energy,* vol. 97, pp. 636-645, 2016.

[11] J. Meza, M. Yildirim and A. Masud, "A model for the multiperiod multiobjective power generation expansion problem," *IEEE Transactions on Power Systems,* vol. 22, no. 2, pp. 871-878, May 2007.

[12] S. J. Kazempour, A. J. Conejo and C. Ruiz, "Strategic Generation Investment Using a Complementarity Approach," *IEEE Transactions on Power Systems,* vol. 26, no. 2, pp. 940-948, 2011.

[13] D. Pozo, E. Sauma and J. Contreras, "A three-level static MILP model for generation and transmission expansion planning," *IEEE Transactions on Power Systems,* vol. 28, pp. 201-210, 2013.

[14] Y. Tohidi, M. R. Hesamzadeh and F. Regairaz, "Sequential Coordination of Transmission Expansion Planning With Strategic Generation Investments," *IEEE Transactions on Power Systems,* vol. 32, no. 4, pp. 2521-2534, 2017.

[15] J. Aghaei, N. Amjady, A. Baharvandi and M. Akbari, "Generation and Transmission Expansion Planning: MILP-Based Probabilistic Model," *IEEE Transactions on Power Systems,* vol. 29, pp. 1592-1601, 2014.

[16] S. Z. Moghaddam, "Generation and transmission expansion planning with high penetration of wind farms considering spatial distribution of wind speed," *International Journal of Electrical Power & Energy Systems,* vol. 106, pp. 232-241, 2019.

[17] S. You, S. W. Hadley, M. Shankar and Y. Liu, "Co-optimizing Generation and Transmission Expansion with Wind Power in Large-Scale Power Grids – Implementation in the US Eastern Interconnection," *Electric Power System Research,* vol. 133, pp. 209-218, 2016.

[18] B. Alizadeh and S. Jadid, "Reliability constrained coordination of generation and transmission expansion planning in power systems using mixed integer programming," *IET Generation, Transmission & Distribution,* vol. 5, pp. 948-960, 2011.

[19] B. Alizadeh and S. Jadid, "A dynamic model for coordination of generation and transmission expansion planning in power systems," *International Journal of Electrical Power & Energy Systems,* vol. 65, pp. 408-418, 2015.

[20] F. Fallahi, M. Nick, G. H. Riahy, S. H. Hosseinian and A. Doroudi, "The value of energy storage in optimal non-firm wind capacity connection to power systems," *Renewable Energy,* vol. 64, pp. 34-42, 2014.

[21] C. Unsihuay-Vila, J. W. Marangon-Lima, A. C. Zambroni de Souza and I. J. Perez-Arriaga, "Multistage expansion planning of generation and interconnections with sustainable energy development criteria: A multiobjective model," *International Journal of Electrical Power & Energy Systems,* vol. 33, no. 2, pp. 258-270, 2011.

[22] M. Hesamzadeh, J. Rosellon and I. Vogelsang, Transmission Network Investment in Liberalized Power Markets, Springer, 2020.

[23] A. Schwele, J. Kazempour and P. Pinson, "Do unit commitment constraints affect generation expansion? A scalable stochastic model," *Energy Systems,* 2019.

[24] M. Fürsch, S. Nagl and D. Lindenberg, "Optimization of power plant investments under uncertain renewable energy deployment paths: a multistage stochastic programming approach," *Energy Systems,* vol. 5, pp. 85-121, 2014.

[25] K. Poncelet, E. Delarue, D. Six, J. Duerinck and W. D'haeseleer, "Impact of the level of temporal and operational detail in energy-system planning models," *Applied Energy,* vol. 162, pp. 631-643, 2016.

[26] B. Saravanan, S. Das, S. Sikiri and D. P. Kothari, "A solution to the unit commitment problem—a review," *Frontiers in Energy,* vol. 7, no. 2, pp. 223-236, 2013.

[27] A. Taverna, "Algorithms for the large-scale Unit Commitment Problem in the simulation of Power Systems," PhD thesis. Università degli Studi di Milano, Milan, Italy, 2016.

[28] K. W. Hedman, R. P. O'Neill and S. S. Oren, "Analyzing valid inequalities of the generation unit commitment problem," in *2009 IEEE/PES Power Systems Conference and Exposition*, Seattle, WA, 2009.

[29] C. Gentile, G. Morales-España and A. Ramos, "A tight MIP formulation of the unit commitment problem with start-up and shut-down constraints," *EURO Journal on Computational Optimization,* vol. 5, pp. 177-201, 2017.

[30] G. Morales-España, J. M. Latorre and A. Ramos, "Tight and Compact MILP Formulation for the Thermal Unit Commitment Problem," *IEEE Transactions on Power Systems,* vol. 28, no. 4, pp. 4897-4908, 2013.

[31] M. Carrion and J. M. Arroyo, "A computationally efficient mixed-integer linear formulation for the thermal unit commitment problem," *IEEE Transactions on Power Systems,* vol. 21, no. 3, pp. 1371-1378, 2006.

[32] A. Belderbos and E. Delarue, "Accounting for flexibility in power system planning with renewables," *International Journal of Electrical Power & Energy Systems,* vol. 71, pp. 33-41, 2015.

[33] D. S. Kirschen, J. Ma, V. Silva and R. Belhomme, "Optimizing the flexibility of a portfolio of generating plants to deal with wind generation," in *2011 IEEE Power and Energy Society General Meeting*, Detroit, MI, USA, 2011.

[34] M. Fripp, "Switch: A Planning Tool for Power Systems with Large Shares of Intermittent Renewable Energy," *Environmental Science & Technology,* vol. 46, no. 11, pp. 6371-6378, 2012.

[35] E. K. Hart and M. Z. Jacobson, "A Monte Carlo approach to generator portfolio planning and carbon emissions," *Renewable Energy,* vol. 36, no. 8, pp. 2278-2286, 2011.

[36] M. Nick, R. Cherkaoui and M. Paolone, "Optimal Allocation of Dispersed Energy Storage Systems in Active Distribution Networks for Energy Balance and Grid Support," *IEEE Transactions on Power Systems,* vol. 29, no. 5, pp. 2300-2310, 2014.

[37] P. Nahmmacher, E. Schmid, L. Hirth and B. Knopf, "Carpe diem: A novel approach to select representative days for long-term power system models with high shares of renewable energy sources," *Energy,* vol. 112, pp. 430-442, 2016.

[38] M. S. ElNozahy, M. M. A. Salama and R. Seethapathy, "A probabilistic load modelling approach using clustering algorithms," in *2013 IEEE Power & Energy Society General Meeting*, 2013.

[39] S. Fazlollahi, S. L. Bungener, P. Mandel, G. Becker and F. Maréchal, "Multi-objectives, multi-period optimization of district energy systems: I. Selection of typical operating periods," *Computers & Chemical Engineering,* vol. 65, pp. 54-66, 2014.

[40] K. Poncelet, H. Höschle, E. Delarue, A. Virag and W. D'haeseleer, "Selecting Representative Days for Capturing the Implications of Integrating Intermittent Renewables in Generation Expansion Planning Problems," *IEEE Transactions on Power Systems,* vol. 32, no. 3, pp. 1936-1948, 2017.

[41] P. Arora, V. Deepali and S. Varshney, "Analysis of k-means and k-medoids algorithm for big data," *Procedia Computer Science,* vol. 78, pp. 507-512, 2016.

[42] K. Poncelet, A. van Stiphout, J. Meus, E. Delarue and W. D'haeseleer, "LUSYM Invest: a generation expansion planning model with a high level of temporal and technical detail," *KU Leuven, TME Working Paper WP EN2018-07,* 2018.

[43] D. A. Tejada-Arango, M. Domeshek, S. Wogrin and E. Centeno, "Enhanced Representative Days and System States Modeling for Energy Storage Investment Analysis," *IEEE Transactions on Power Systems,* vol. 33, no. 6, pp. 6534-6544, 2018.

[44] D. A. Tejada-Arango, "Co-optimization of energy storage technologies in tactical and strategic planning models," PhD thesis. Comillas University, Madrid, Spain, 2019.

[45] G. Rodriguez, "Lecture Notes on Generalized Linear Models," Princeton University, 2007. [Online]. Available: http://data.princeton.edu/wws509/notes/.

[46] B. Kröse and P. van der Smagt, An Introduction to Neural Networks, University of Amsterdam, 1996.

[47] H. Dahan, S. Cohen, L. Rokach and O. Maimon, Proactive Data Mining with Decision Trees, Springer Publishing Company, Incorporated, 2014.

[48] L. Rokach and O. Maimon, Data mining with decision trees: theory and applications, Singapore: World Scientific Publishing, 2008.

[49] V. Vapnik, Statistical Learning Theory, John Wiley & Sons, 1998.

[50] J. Suykens, "Nonlinear modelling and support vector machines," in *Proceedings of the 18th IEEE Instrumentation and Measurement Technology Conference. Rediscovering Measurement in the Age of Informatics (Cat. No.01CH 37188)*, Budapest, 2001.

[51] ENTSO-E; ENTSOG, "TYNDP 2020 Scenario Report," 2019.

[52] Terna S.p.A; Gruppo Terna, "Piano di Sviluppo," 2019.

[53] W. Cole and A. W. Frazier, "Cost Projections for Utility-Scale Battery Storage," National Renewable Energy Laboratory, 2019.

[54] International Renewable Energy Agency (IRENA), "Electricity storage and renewables: Costs and markets to 2030," 2017.

[55] European Energy Research Alliance (EERA), "Fact Sheet 1 - Pumped Hydro Energy Storage," Brussels, 2016.

[56] M. Moeini-Aghtaie, A. Abbaspour and M. Fotuhi-Firuzabad, "Incorporating Large-Scale Distant Wind Farms in Probabilistic Transmission Expansion Planning—Part I: Theory and Algorithm," *IEEE Transactions on Power Systems,* vol. 27, no. 3, pp. 1585-1593, 2012.

[57] H. Yu, C. Y. Chung, K. P. Wong and J. H. Zhang, "A Chance Constrained Transmission Network Expansion Planning Method With Consideration of Load and Wind Farm Uncertainties," *IEEE Transactions on Power Systems,* vol. 24, no. 3, pp. 1568-1576, 2009.

[58] G. A. Orfanos, P. S. Georgilakis and N. D. Hatziargyriou, "Transmission Expansion Planning of Systems With Increasing Wind Power Integration," *IEEE Transactions on Power Systems,* vol. 28, no. 2, pp. 1355-1362, 2012.

[59] C. Ruiz and A. J. Conejo, "Robust transmission expansion planning," *European Journal of Operational Research,* vol. 242, no. 2, pp. 390-401, 2015.

[60] J. Li, Z. Li, F. Liu, H. Ye, X. Zhang, S. Mei and N. Chang, "Robust Coordinated Transmission and Generation Expansion Planning Considering Ramping Requirements and Construction Periods," *IEEE Transactions on Power Systems,* vol. 33, no. 1, pp. 268-280, 2018.

[61] S. Dehghan, N. Amjady and K. Ahad, "Two-Stage Robust Generation Expansion Planning: A Mixed Integer Linear Programming Model," *IEEE Transactions on Power Systems,* vol. 29, no. 2, pp. 584-597, 2014.

[62] B. Chen, J. Wang, L. Wang, Y. He and Z. Wang, "Robust Optimization for Transmission Expansion Planning: Minimax Cost vs. Minimax Regret," *IEEE Transactions on Power Systems,* vol. 29, no. 6, pp. 3069-3077, 2014.

[63] D. Mejía-Giraldo and J. D. McCalley, "Maximizing Future Flexibility in Electric Generation Portfolios," *IEEE Transactions on Power Systems,* vol. 29, pp. 279-288, 2014.

[64] P. Maloney, O. Olatujoye, A. J. Ardakani, D. Mejïa-Giraldo and J. McCalley, "A Comparison of Stochastic and Adaptation programming methods for long term generation and transmission co-optimization under uncertainty," in *2016 North American Power Symposium (NAPS)*, Denver, CO, USA, 2016.

[65] J. Birge and F. Louveaux, Introduction to Stochastic Programming, 1997.

[66] J. Alvarez Lopez, K. Ponnambalam and V. H. Quintana, "Generation and Transmission Expansion Under Risk Using Stochastic Programming," *IEEE Transactions on Power Systems,* vol. 22, no. 3, pp. 1369-1378, 2007.

[67] A. H. Van der Weijde and B. F. Hobbs, "The economics of planning electricity transmission to accommodate renewables: Using two-stage optimisation to evaluate flexibility and the cost of disregarding uncertainty," *Energy Economics,* vol. 34, no. 6, pp. 2089-2101, 2012.

[68] S. Jin, S. M. Ryan, J.-P. Watson and D. L. Woodruff, "Modeling and solving a large-scale generation expansion planning problem under uncertainty," *Energy Systems,* vol. 2, pp. 209-242, 2011.

[69] S. Jin, A. Botterud and S. M. Ryan, "Temporal Versus Stochastic Granularity in Thermal Generation Capacity Planning With Wind Power," *IEEE Transactions on Power Systems,* vol. 29, no. 5, pp. 2033-2041, 2014.

[70] A. J. Conejo, L. Baringo, S. J. Kazempour and A. S. Siddiqui, Investment in electricity generation and transmission: Decision making under uncertainty, Springer, 2016.

[71] Y. Liu, R. Sioshansi and A. J. Conejo, "Multistage Stochastic Investment Planning With Multiscale Representation of Uncertainties and Decisions," *IEEE Transactions on Power Systems,* vol. 33, no. 1, pp. 781-791, 2018.

[72] L. Baringo and A. J. Conejo, "Risk-Constrained Multi-Stage Wind Power Investment," *IEEE Transactions on Power Systems,* vol. 28, no. 1, pp. 401-411, 2013.

[73] J. F. Benders, "Partitioning procedures for solving mixed-variables programming problems," *Numerische Mathematik,* vol. 4, no. 1, pp. 238-252, 1962.

[74] A. J. Conejo, E. Castillo, R. García-Bertrand and R. Mínguez, Decomposition Techniques in Mathematical Programming: Engineering and Science Applications, Berlin: Springer, 2006.

[75] V. Zverovich, C. I. Fábián, E. F. D. Ellison and G. Mitra, "A computational study of a solver system for processing two-stage stochastic LPs with enhanced Benders decomposition," *Mathematical Programming Computation,* vol. 4, pp. 211-238, 2012.

[76] J. Murphy, "Benders, nested Benders and stochastic programming: An intuitive introduction," 2013.

[77] J. R. Birge and F. V. Louveaux, "A multicut algorithm for two-stage stochastic linear programs," *European Journal of Operational Research,* vol. 34, no. 3, pp. 384-392, 1988.

[78] J. A. Bloom, “Solving an Electricity Generating Capacity Expansion Planning Problem by Generalized Benders' Decomposition,” *Operations Research,* vol. 31, no. 1, pp. 84-100, 1983.

[79] C. Skar, G. Doorman and A. Tomasgard, “Large-scale power system planning using enhanced Benders decomposition,” *2014 Power Systems Computation Conference (PSCC),* pp. 1-7, 2014.

[80] A. Nasri, S. J. Kazempour and A. J. Conejo, “Network-Constrained AC Unit Commitment Under Uncertainty: A Benders’ Decomposition Approach,” *IEEE Transactions on Power Systems,* vol. 31, no. 1, pp. 412-422, 2016.

[81] S. Lumbreras and A. Ramos, “Transmission expansion planning using an efficient version of Benders' decomposition. A case study,” in *IEEE Grenoble Conference*, Grenoble, 2013.

[82] S. Lumbreras and A. Ramos, “How to solve the transmission expansion planning problem faster: acceleration techniques applied to Benders’ decomposition,” *IET Generation, Transmission & Distribution,* vol. 10, no. 10, pp. 2351-2359, 2016.

[83] S. Trukhanov, L. Ntaimo and A. Schaefer, “Adaptive multicut aggregation for two-stage stochastic linear programs with recourse,” *European Journal of Operational Research,* vol. 206, no. 2, pp. 395-406, 2010.

[84] F. You and I. E. Grossmann, “Multicut Benders decomposition algorithm for process supply chain planning under uncertainty,” *Annals of Operations Research,* vol. 210, no. 1, pp. 191-211, 2013.

[85] M. R. Bussieck, M. C. Ferris and T. Lohmann, “GUSS: Solving Collections of Data Related Models Within GAMS,” in *Algebraic Modeling Systems*, Berlin, Springer, 2012, pp. 35-56.

[86] Ministero dello Sviluppo Economico; Ministero dell'Ambiente e della Tutela del Territorio e del Mare; Ministero delle Infrastrutture e dei Trasporti, “Piano Nazionale Integrato per l'Energia e il Clima,” 2019.

[87] ISPRA - Istituto Superiore per la Protezione e la Ricerca Ambientale, “Fattori di emissione atmosferica di gas a effetto serra nel settore elettrico nazionale e nei principali Paesi Europei,” 2020.

[88] J. Devlin, K. Li, P. Higgins and A. Foley, “The importance of gas infrastructure in power systems with high wind power penetrations,” *Applied Energy,* vol. 167, pp. 294-304, 2016.

[89] M. Jentsch, T. Trost and M. Sterner, "Optimal Use of Power-to-Gas Energy Storage Systems in an 85% Renewable Energy Scenario," *Energy Procedia,* vol. 46, pp. 254-261, 2014.

[90] Q. Zeng, J. Fang, J. Li and Z. Chen, "Steady-state analysis of the integrated natural gas and electric power system with bi-directional energy conversion," *Applied Energy,* vol. 184, pp. 1483-1492, 2016.

[91] F. Fallahi and P. Maghouli, "Integrated unit commitment and natural gas network operational planning under renewable generation uncertainty," *International Journal of Electrical Power & Energy Systems,* vol. 117, 2020.

[92] D. Wang, J. Qiu, K. Meng, X. Gao and Z. Dong, "Coordinated expansion co-planning of integrated gas and power systems," *Journal of Modern Power Systems and Clean Energy,* vol. 5, no. 3, pp. 314-325, 2017.

[93] C. Unsihuay-Vila, J. W. Marangon-Lima, A. C. Zambroni de Souza, I. J. Perez-Arriaga and P. P. Balestrassi, "A Model to Long-Term, Multiarea, Multistage, and Integrated Expansion Planning of Electricity and Natural Gas Systems," *IEEE Transactions on Power Systems,* vol. 25, no. 2, pp. 1154-1168, 2010.

[94] V. Khaligh, M. O. Buygi, A. Anvari-Moghaddam and J. M. Guerrero, "A Multi-Attribute Expansion Planning Model for Integrated Gas–Electricity System," *Energies,* vol. 11, no. 10, 2018.

[95] M. Chaudry, N. Jenkins, M. Qardan and J. Wu, "Combined gas and electricity network expansion planning," *Applied Energy,* vol. 113, pp. 1171-1187, 2014.

[96] B. Odetayo, M. Kazemi, J. MacCormack, W. D. Rosehart, H. Zareipour and A. R. Seifi, "A Chance Constrained Programming Approach to the Integrated Planning of Electric Power Generation, Natural Gas Network and Storage," *IEEE Transactions on Power Systems,* vol. 33, no. 6, pp. 6883-6893, 2018.

[97] B. Zhao, A. J. Conejo and R. Sioshansi, "Coordinated Expansion Planning of Natural Gas and Electric Power Systems," *IEEE Transactions on Power Systems,* vol. 33, no. 3, pp. 3064-3075, 2018.

[98] J. B. Nunes, N. Mahmoudi, T. K. Saha and D. Chattopadhyay, "A stochastic integrated planning of electricity and natural gas networks for Queensland, Australia considering high renewable penetration," *Energy,* vol. 153, pp. 539-553, 2018.

[99] T. Ding, Y. Hu and Z. Bie, "Multi-Stage Stochastic Programming With Nonanticipativity Constraints for Expansion of Combined Power and Natural Gas Systems," *IEEE Transactions on Power Systems,* vol. 33, no. 1, pp. 317-328, 2018.

[100] S. Chen, A. J. Conejo, R. Sioshansi and Z. Wei, "Operational Equilibria of Electric and Natural Gas Systems With Limited Information Interchange," *IEEE Transactions on Power Systems,* vol. 35, no. 1, pp. 662-671, 2020.

[101] V. Khaligh, M. O. Buygi, A. A. Moghaddam and J. M. Guerrero, "Leader-Follower Approach to Gas-Electricity Expansion Planning Problem," in *2018 IEEE International Conference on Environment and Electrical Engineering and 2018 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I&CPS Europe)*, Palermo, 2018.

[102] J. Qiu, H. Yang, Z. Y. Dong, J. H. Zhao, K. Meng, F. J. Luo and K. P. Wong, "A Linear Programming Approach to Expansion Co-Planning in Gas and Electricity Markets," *IEEE Transactions on Power Systems,* vol. 31, no. 5, pp. 3594-3606, 2016.

[103] M. Qadrdan, M. Abeysekera, M. Chaudry, J. Wu and N. Jenkins, "Role of power-to-gas in an integrated gas and electricity system in Great Britain," *International Journal of Hydrogen Energy,* vol. 40, no. 17, pp. 5763-5775, 2015.

[104] S. Clegg and P. Mancarella, "Integrated Modeling and Assessment of the Operational Impact of Power-to-Gas (P2G) on Electrical and Gas Transmission Networks," *IEEE Transactions on Sustainable Energy,* vol. 6, no. 4, pp. 1234-1244, 2015.

[105] Y. Li, W. Liu, M. Shahidehpour, F. Wen, K. Wang and Y. Huang, "Optimal Operation Strategy for Integrated Natural Gas Generating Unit and Power-to-Gas Conversion Facilities," *IEEE Transactions on Sustainable Energy,* vol. 9, no. 4, pp. 1870-1879, 2018.

[106] C. He, L. Wu, T. Liu and Z. Bie, "Robust Co-Optimization Planning of Interdependent Electricity and Natural Gas Systems With a Joint N-1 and Probabilistic Reliability Criterion," *IEEE Transactions on Power Systems,* vol. 33, no. 2, pp. 2140-2154, 2018.

[107] Q. Zeng, B. Zhang, J. Fang and Z. Chen, "A bi-level programming for multistage co-expansion planning of the integrated gas and electricity system," *Applied Energy,* vol. 200, pp. 192-203, 2017.

[108] X. Luo, J. Wang, M. Dooner and J. Clarke, "Overview of current development in electrical energy storage technologies and the application potential in power system operation," *Applied Energy,* vol. 137, pp. 511-536, 2015.

[109] B. Palmintier and M. Webster, "Impact of unit commitment constraints on generation expansion planning with renewables," in *2011 IEEE Power and Energy Society General Meeting*, Detroit, MI, USA, 2011.

[110] J. Meus, K. Poncelet and E. Delarue, "Applicability of a Clustered Unit Commitment Model in Power System Modeling," *IEEE Transactions on Power Systems,* vol. 33, no. 2, pp. 2195-2204, 2018.

[111] Snam; Terna, "Documento di Descrizione degli Scenari 2019," 2019.

[112] H. Marañón-Ledesma and A. Tomasgard, "Long-Term Electricity Investments Accounting for Demand and Supply Side Flexibility," MPRA Paper 93341, University Library of Munich, Germany, 2019.

[113] M. Albadi and E. El-Saadany, "A summary of demand response in electricity markets," *Electric Power Systems Research,* vol. 78, no. 11, pp. 1989-1996, 2008.

[114] L. Bastida, J. J. Cohen, A. Kollmann, A. Moya and J. Reichl, "Exploring the role of ICT on household behavioural energy efficiency to mitigate global warming," *Renewable and Sustainable Energy Reviews,* vol. 103, pp. 455-462, 2019.

[115] K. Dietrich, J. M. Latorre, L. Olmos and A. Ramos, "Demand Response in an Isolated System With High Wind Integration," *IEEE Transactions on Power Systems,* vol. 27, no. 1, pp. 20-29, 2012.

[116] C. De Jonghe, B. F. Hobbs and R. Belmans, "Optimal Generation Mix With Short-Term Demand Response and Wind Penetration," *IEEE Transactions on Power Systems,* vol. 27, no. 2, pp. 830-839, 2012.

[117] R. Sioshansi and W. Short, "Evaluating the Impacts of Real-Time Pricing on the Usage of Wind Generation," *IEEE Transactions on Power Systems,* vol. 24, no. 2, pp. 516-524, 2009.

[118] A. Nikoobakht, J. Aghaei, M. Shafie-Khah and J. P. S. Shafie-Khah, "Assessing Increased Flexibility of Energy Storage and Demand Response to Accommodate a High Penetration of Renewable Energy Sources," *IEEE Transactions on Sustainable Energy,* vol. 10, no. 2, pp. 659-669, 2019.

[119] T. Lohmann and S. Rebennack, "Tailored Benders Decomposition for a Long-Term Power Expansion Model with Short-Term Demand Response," *Management Science,* vol. 63, no. 6, 2016.

[120] M. Samadi, M. H. Javidi and M. S. Ghazizadeh, "Modeling the effects of demand response on generation expansion planning in restructured power system," *Journal of Zhejiang University SCIENCE C,* vol. 14, pp. 966-976, 2013.

[121] A. K. Kazerooni and J. Mutale, "Network investment planning for high penetration of wind energy under demand response program," in *2010 IEEE 11th International Conference on Probabilistic Methods Applied to Power Systems*, Singapore, 2010.

[122] D. Z. Fitiwi and V. Bertsch, "Interactions between Demand Response and Network Expansion Planning: A Quantitative Analysis," in *2018 15th International Conference on the European Energy Market (EEM)*, Lodz, 2018.

[123] R. Hejeejo and J. Qiu, "Probabilistic Transmission Expansion Planning Considering Distributed Generation and Demand Response," *IET Renewable Power Generation,* vol. 11, no. 5, pp. 650-658, 2017.

[124] M. Asensio, P. Meneses de Quevedo, G. Muñoz-Delgado and J. Contreras, "Joint Distribution Network and Renewable Energy Expansion Planning Considering Demand Response and Energy Storage—Part I: Stochastic Programming Model," *IEEE Transactions on Smart Grid,* vol. 9, no. 2, pp. 655-666, 2018.

[125] S. L. Gbadamosi and N. I. Nwulu, "A multi-period composite generation and transmission expansion planning model incorporating renewable energy sources and demand response," *Sustainable Energy Technologies and Assessments,* vol. 39, 2020.

[126] S. Zaourar and J. Malick, "Quadratic stabilization of Benders decomposition," Working paper, 2014.