



A two-stage stochastic MILP model for generation and transmission expansion planning with high shares of renewables

Giovanni Micheli, et al. *[full author details at the end of the article]*

Received: 31 December 2019 / Accepted: 2 September 2020
© The Author(s) 2020

Abstract

This paper is concerned with the generation and transmission expansion planning of large-scale energy systems with high penetration of renewable energy sources. Since expansion plans are usually provided for a long-term planning horizon, the system conditions are generally uncertain at the time the expansion plans are decided. In this work, we 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 plants and the economic viability of renewable generation. To deal with this long-term uncertainty, we consider different scenarios and we define capacity expansion decisions using a two-stage stochastic programming model that aims at minimizing the sum of investment, decommissioning and fixed costs and the expected value of operational costs. To be computationally tractable most of the existing expansion planning models employ a low level of temporal and technical detail. However, this approach is no more an appropriate approximation for power systems analysis, since it does not allow to accurately study all the challenges related to integrating high shares of intermittent energy sources, underestimating the need for flexible resources and the expected costs. To provide more accurate expansion plans for power systems with large penetration of renewables, in our analysis, we consider a high level of temporal detail and we include unit commitment constraints on a plant-by-plant level into the expansion planning framework. To maintain the problem computationally tractable, we use representative days and we implement a multi-cut Benders decomposition algorithm, decomposing the original problem both by year and by scenario. Results obtained with our methodology in the Italian energy system under a 21-year planning horizon show how the proposed model can offer professional guidance and support in strategic decisions to the different actors involved in electricity transmission and generation.

Keywords Generation and transmission expansion planning · Two-stage stochastic programming · Multi-cut Benders decomposition · Large-scale power systems · Unit commitment · Representative days

1 Introduction

Generation and transmission expansion planning models aim at determining the least-cost investment schedule for constructing new power generation capacity, building new electrical interconnections and decommissioning thermal power plants. 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 CO₂ 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 system operators (e.g. TSO) 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 [1–3].

Since generation and transmission expansion plans are usually provided for a long-term planning horizon, the 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. These uncertainties can be classified into 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 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 accurately represent the uncertainty framework in the planning decisions, several approaches have been developed in the literature, including a fuzzy decision approach, chance-constrained models, Monte Carlo simulation, robust optimization, adaptation 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 [4] 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 [5] 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 [6], 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 [7] 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 [8] 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 [9, 10] to address generation expansion and transmission expansion, respectively. Specifically, reference [9] introduces a multiyear 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 [10] 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 overall uncertainties are minimized, respectively.

Adaptation programming represents another method to cope with uncertainty. As described in [11], adaptation programming designs a flexible system by minimizing the sum of investment and operational cost and system future adaptation cost to the conditions of other identified scenarios. Reference [12] further explores the adaptation programming method by applying this kind of model to a small system over a 40-years planning horizon, considering wind and solar investment 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 [13], 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. 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 the realization of parameter values. For instance, in [14] 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 [15] presents a stochastic two-stage optimization model to evaluate interregional grid reinforcements in Great Britain. The same approach is also used in [16] to determine the type and quantity of power

plants to be constructed in each year of an extended planning horizon, considering the uncertainty regarding future demand and fuel prices. Authors in [17] 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 [18], 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 [19–21]. For instance, in [19] a multi-stage multi-scale linear stochastic model is presented to optimize electricity generation, storage and transmission investments over a long planning horizon. 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 [20] 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 [21] 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 the multi-stage approach better represents long-term dynamics than the two-stage method, the complexity of the problem in multi-stage models is further increased. Finding the right balance between modeling accuracy and computation tractability remains an open research topic. Specifically, to limit computational costs most of the existing planning models employ a low level of temporal detail. However, as shown in [22], the low level of temporal detail can significantly affect the results, 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 modelling 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 flexibility and, consequently, to underestimate the required new generation capacity as well as the system costs.

Table 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 paper are compared.

Specifically, to provide reliable expansion plans for large-scale energy systems with high shares of renewables, in this paper we propose a two-stage stochastic model for joint generation and transmission expansion planning, 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 distinct feature of the proposed model is a detailed representation of the short-term operation. The choice of a two-stage model rather than a multi-stage approach is due to computational restrictions. Indeed, as can be noticed in Table 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 [19, 21], 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 approach differs from multi-stage models by allowing investment decisions in each year of the planning horizon.

Contributions of this paper are as follows:

- High level of temporal detail by applying a clustering analysis on different time series (i.e., load, wind capacity factors, and solar capacity factors) to identify a small number of representative days.
- High level of technical detail, including unit commitment constraints on a plant-by-plant level and providing an accurate representation of both short-term and long-term storage.
- Modelling of long-term dynamics, determining the investment schedule along the planning horizon considering multiple investment periods and considering long-term policy goals and environmental targets.
- Modelling of long-term uncertainty, considering different scenarios for production costs of thermal power plants.
- Application of Benders Decomposition, decomposing the original model both by year and by scenario to maintain the stochastic problem computationally tractable.
- Application to a real-case study related to the Italian energy system, to evaluate the achievement of policy goals set by the European Commission.

Table 1 Comparison between relevant works in the literature and the proposed model

References	Modeling Approach	Generation Expansion		Storage Expansion		Transmission Expansion	Operation Evaluation	UC Constraints	Policy Goals	Investment Periods
		Thermal	Solar	Wind	Long Term					
[8]	RO	Yes	Yes	No	No	No	LDC	No	None	5
[11]	AP	Yes	Yes	Yes	Yes	No	LDC	No	RES penetration	20
[12]	AP	Yes	Yes	Yes	No	Yes	LDC	No	None	2
[15]	2-Stage SP	Yes	Yes	No	No	Yes	Sampled hours	No	None	2
[16]	2-Stage SP	Yes	Yes	No	No	No	LDC	No	None	10
[17]	2-Stage SP	Yes	No	No	No	No	Sampled weeks	Yes	None	1
[19]	Multi-Stage SP	Yes	Yes	Yes	No	Yes	Rep. days	No	None	4
[21]	Multi-Stage SP	Yes	No	No	Yes	No	Rep. days	No	None	4
[23]	2-Stage SP	Yes	Yes	No	No	No	Rep. days	Yes	RES penetration	1
This paper	2-Stage SP	Yes	Yes	Yes	Yes	Yes	Rep. days	Yes	RES penetration, CO ₂ emissions, Fuels availability	21

RO robust optimization, AP adaptation programming, SP stochastic programming, LDC load duration curve.

To the best of our knowledge, there are no other works with the same characteristics in the literature. The remainder of the paper is organized as follows. In Sect. 2 we describe the clustering analysis performed to identify the representative days. The mathematical formulation of the expansion planning model as a two-stage stochastic mixed-integer linear programming problem is introduced in Sect. 3. Section 4 describes the decomposition algorithm. In Sect. 5 we present a real case study concerning the expansion of the Italian energy system. Finally, conclusions are drawn in Sect. 6.

2 Clustering analysis

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 days instead of considering every hour of the planning horizon. Different approaches have been proposed to identify the representative days. Some authors use simple heuristics, selecting the days that contain the minimum demand, the maximum demand or the largest daily demand spread [24]. Other works combine heuristic approaches with the random selection of some additional days [25, 26]. More advanced methods apply clustering algorithms to group days with similar load, wind production or solar production into clusters [27–29]. For each group either the cluster centroid or a specific historical day is then taken as the representative day. Finally, some works select representative days by minimizing the difference between the load duration curve and the one constructed by using the representative days [30, 31]. In this paper, we identify representative days through a hybrid approach based on the iterative application of the k -medoids algorithm and on the selection of the most suited number of representative days to be used by considering load duration curves. We refer the reader to [32] for a full description of the proposed approach, hereafter summarized.

Specifically, this analysis aims to extract representative days from larger sets of days, consisting of multiple time series of electricity demand and renewables capacity factors. Therefore, a very important feature to be considered in the selection process is the correlation between different time series. In our approach, we apply the following procedure, designed to capture the spatial correlation between time series:

Vectors creation We consider the first year of the planning horizon and for each day d we define a vector V_d containing values of load, wind capacity factors and solar capacity factors for every zone and every hour of the day. Therefore, the dimension of each vector V_d equals 72 times the number of zones, while the number of vectors is equal to 365.

Normalization Since wind and solar capacity factors are parameters in the range $[0;1]$, to properly evaluate the distance between vectors, hourly zonal load profiles are normalized, by subtracting the minimum load and dividing by the difference between maximum and minimum load.

k-medoids algorithm application The objective of the k -medoids algorithm is to compute k clusters to minimize the deviation between observations V_d and their representative V_c^* , which it is also called quantization error or intracluster distance:

$$\min_c \sum_{d \in D_c} \|V_d - V_c^*\|^2$$

with c being the index for clusters, d the index for days and D_c the set of all days d within cluster c . While for k -means parameters V_c^* are computed as the mean of all the points within the cluster c , for k -medoids the representative of each cluster is a vector belonging to that group. As discussed in [29], since using as representative days specific historic days rather than clusters' means usually provide better results, in our approach, we suggest applying k -medoids rather than k -means.

Representative days weighting In our approach, representative days are weighted using cluster sizes, which are the number of observations grouped within a specific cluster.

Selection of the number of representative days To define the best number of representative days to be used we perform the k -medoids algorithm for increasing values of the parameter k (starting with $k = 2$), comparing the obtained results in terms of load duration curves. Specifically, for each partition, we compute for each zone the mean absolute percentage error (MAPE) between the original zonal load duration curve and the one obtained by using the representative days and their associated weights. We then compute the system average MAPE (i.e., the average between the MAPEs of zonal load duration curves) and we select the minimum number of representative days that allows obtaining a system average MAPE lower than an input threshold.

Application of annual load growth factors 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 identification of representative days allows reducing computational time while maintaining a high level of temporal detail, but it introduces a strong limitation in storage operation modelling. Indeed, as explained in [33], this approach allows providing a good representation of storage operation within a day, but since it does not preserve the chronology among representative days, any energy storage system with a cycle longer than 24 h cannot be modelled with the highest accuracy. To provide an accurate representation of long-term storage, in this paper we follow the same approach described in [29] applying constraints that guarantee some continuity between representative days. These constraints, along with the other equations that allow identifying the expansion plans, are described in the following section.

3 Mathematical formulation

Investments in the new generation, transmission and storage facilities as well as decommissioning of thermal plants are defined using a mixed-integer linear programming model. This section presents the modelling assumptions as well as the notation and the mathematical formulation of the optimization model.

3.1 Modelling assumptions

The mathematical model that defines investment and decommissioning decisions is based on the system costs minimization formulation. Thus, adopting a centralized approach, the objective of the proposed model is to determine the expansion plan that is optimal for the whole energy system under study. Bidding strategies and competition between generators are beyond the scope of this work. In developing countries, where the unbundling of the energy sector is still ongoing, the centralized approach reflects the decision making framework, since expansion decisions are taken by vertically integrated monopoly utilities. Instead, in an unbundled energy environment, generation expansion pertains to profit-seeking agents, while transmission expansion pertains to the transmission system operator, which is a welfare-focused agent. In this case, the centralized approach reflects the *anticipative planning* analysis performed by transmission system operators to identify the best network reinforcement and to set incentives that could induce generation companies to invest in a socially efficient manner.

As regards to investment decisions, in this paper we manage investments in new generating facilities differently. Specifically, while we use continuous variables to model wind and solar capacity expansion, investments in new thermal plants are managed through a set of candidate projects and binary variables describing whether the units are realized or not. Indeed, while it is possible to build wind and solar plants of any capacity, thermal units usually present a specified size, not allowing the modelling of thermal capacity expansion through continuous variables. Moreover, binary variables are needed to model some logical conditions between investment decisions, such as the existence of optional, mandatory, mutually exclusive, associate or upgrade projects. For the same reasons, binary variables model investments in new transmission lines. We introduce binary variables to represent also the building of new hydropower plants, usually characterized by predefined sizes, while investments in batteries capacity are modeled through continuous variables.

To represent the transmission network, we adopt a transportation model, imposing the fulfilment of nodal balance equations and transmission flow limits constraints, without incorporating voltage variables and the relationship of real power transfers with the bus angle difference and line impedance. Finally, the proposed optimization model is based on a perfect knowledge of investment costs, while demand is supposed to be inelastic.

3.2 Notation

To describe the two-stage mixed integer linear stochastic programming model for the power generation and transmission expansion problem, the following notation is introduced.

3.2.1 Sets

\mathcal{Y}	Set of years
\mathcal{Z}	Set of zones
\mathcal{M}	Set of macro-areas
\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 ($\mathcal{K} = \mathcal{K}_E \cup \mathcal{K}_C$)
$\Omega_z^k \subset \mathcal{K}$	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 ($\mathcal{L} = \mathcal{L}_E \cup \mathcal{L}_C$)
$\mathcal{L}_{C1} \subset \mathcal{L}_C$	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{F}	Set of fuels
$\Phi_{z,f} \subset \Omega_z^k$	Set of thermal power plants located in zone z using fuel f
\mathcal{H}_E	Set of existing hydropower plants
\mathcal{H}_C	Set of candidate hydropower plants
\mathcal{H}	Set of hydropower plants ($\mathcal{H} = \mathcal{H}_E \cup \mathcal{H}_C$)
$\Omega_z^h \subset \mathcal{H}$	Set of hydropower plants located in zone z
\mathcal{B}	Set of batteries

$\Omega_z^b \subset \mathcal{B}$	Set of batteries located in zone Z
\mathcal{C}	Set of representative days (from 1 to $n_{cluster}$)
\mathcal{T}	Set of hours (from 1 to $24 \cdot n_{cluster}$)
$\mathcal{T}_{in} \subset \mathcal{T}$	First hours of representative days
\mathcal{OT}	Set of all the hours in the planning horizon ($ \mathcal{OT} = 8760 \cdot \mathcal{C} $)
$\mathcal{OT}_M \subset \mathcal{OT}$	Set of hours in the planning horizon multiple of parameter M
\mathcal{W}	Set of scenarios
$ma(z)$	Macro-area that contains zone z
$UP(k)$	Upgrade project of existing thermal power plant k
$r(l)$	Receiving-end zone of transmission line l
$s(l)$	Sending-end zone of transmission line l
$fuel(k)$	Fuel used in thermal power plant k
$year(c)$	Year that contains representative day c
$cl(t)$	Representative day that contains hour t
$yr(t)$	Year that contains hour t
$Map_{ot,t}$	Injective map of each hour ot to a representative hour t

3.2.2 Parameters

y_0	Reference year to which all investment costs are discounted
r	Annual discount rate
c_{ENP}	Penalty for energy not provided, set to 10^4 €/MWh
c_{OG}	Penalty for over-generation, set to 200 €/MWh
$prob_w$	Probability of scenario w
$freq_c$	Weight of cluster c
DC_k	Decommissioning cost of existing thermal power plant $k \in \mathcal{K}_E$
IC_k^{th}	Investment cost of candidate thermal power plant $k \in \mathcal{K}_C$
FC_k	Annual fixed costs of thermal power plant $k \in \mathcal{K}$
$CM_{k,y,w}$	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	Minimum power output of thermal power plant k
\overline{P}_k	Maximum power produced by thermal plant k
SUC_k	Start-up cost of thermal power plant k
MUT_k	Minimum uptime of thermal power plant k
MDT_k	Minimum downtime of thermal power plant k
$\gamma_{k,c,w0}$	Initial state of thermal power plant k in cluster c in scenario w
hh_t	Hour of the day (from 1 to 24) related to hour t (e.g. $hh_{25'} = 1, hh_{48'} = 24$)
$IC_{z,y}^{sol}$	Investment cost of new solar power capacity in zone z in year y
$sol_{z,0}$	Solar power capacity installed in zone z at the beginning of the planning horizon

$\underline{PV}_{z,y}$	Lower bound for solar power capacity in zone z in year y
$\overline{PV}_{z,y}$	Upper bound for solar power capacity in zone z in year y
$IC_{z,y}^{wind}$	Investment cost of new wind power capacity in zone z in year y
$wind_{z,0}$	Wind power capacity installed in zone z at the beginning of the planning horizon
$\underline{W}_{z,y}$	Lower bound for wind power capacity in zone z in year y
$\overline{W}_{z,y}$	Upper bound for wind power capacity in zone z in year y
$D_{z,t}$	Load in zone z in hour t
$R_{z,t}$	Reserve requirement for zone z in hour t
$\mu_{z,t}$	Solar power capacity factor for zone z in hour t
$\rho_{z,t}$	Wind power capacity factor for zone z in hour t
$IC_{h,y}^{hydro}$	Investment cost of candidate hydropower plant $h \in \mathcal{H}_C$ in year y
$Cvar_h$	Operating cost of hydropower plant h
$E_{h,c0}$	Initial energy content of hydropower plant h in cluster c
$F_{h,t}$	Hourly energy inflows for hydropower plant h at time t
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 ($0 \leq \lambda_h^{in} \leq 1$)
λ_h^{out}	Loss coefficient for hydro plant h power generation ($\lambda_h^{out} \geq 1$)
E_{-out}^h	Upper bound on hydro plant h pumping power
E_h	Upper bound on hydro plant h power output
EPR_h	Maximum energy to power ratio (in h) for hydropower plant h
E_{h0}	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 168 h
$IC_{b,y}^{batt}$	Investment cost for battery b in year y
$Cvar_b$	Operating cost of battery b
CAP_b	Upper bound on battery b installed capacity
$E_{b,c0}$	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 ($0 \leq \lambda_b^{in} \leq 1$)
λ_b^{out}	Loss coefficient for battery b discharge ($\lambda_b^{out} \geq 1$)
E_{-out}^b	Upper bound on battery b charge
E_b	Upper bound on battery b discharge
EPR_b	Maximum energy to power ratio (in hours) for battery b

IC_l^{line}	Investment cost of candidate transmission line $l \in \mathcal{L}_C$
$\underline{\tau}_l$	Earliest date for construction of candidate transmission line l
$\overline{\tau}_l$	Latest date for construction of candidate transmission line l
\underline{F}_l	Minimum capacity of transmission line l
\overline{F}_l	Maximum capacity of transmission line l
HR_k	Heat rate of thermal power plant k
$ECnt_f$	Energy content of fuel f
CO_{2f}	CO ₂ emission factor of fuel f
$\overline{FA}_{f,m,y}$	Upper bound on availability of fuel f for macro-area m in year y
$CO_{2m,y}$	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

3.2.3 Variables

1. First-stage variables

a. Continuous variables

- $sol_{z,y}$: New solar capacity installed in zone z in year y
- $wind_{z,y}$: New wind capacity installed in zone z in year y
- $cap_{b,y}$: Storage capacity of battery b installed in year y

b. Binary 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

2. Second-stage variables

a. Continuous variables

- $p_{k,t,w}$: Power production of thermal power plant k in hour t above its minimum output \underline{P}_k

- $E_{h,t,w}^{in}$: Pumping power of hydro reservoir h in hour t in scenario w
- $E_{h,t,w}^{out}$: Power output of hydro reservoir h in hour t in scenario w
- $E_{h,t,w}$: Energy level of hydro reservoir h in hour t in scenario w
- $sl_{h,t,w}$: Spillage from hydro reservoir h in hour t in scenario w
- $\bar{E}_{h,ot,w}$: Energy level of hydro reservoir h in hour ot in scenario w
- $E_{b,t,w}^{in}$: Charge of battery b in hour t in scenario w
- $E_{b,t,w}^{out}$: Discharge of battery b in hour t in scenario w
- $E_{b,t,w}$: Energy level of battery b in hour t in scenario w
- $x_{l,t,w}$: Energy flow on transmission line l in hour t in scenario w
- $ENP_{z,t,w}$: Energy not provided in zone z in hour t in scenario w
- $OG_{z,t,w}$: Over-generation in zone z in hour t in scenario w
- $RES_{z,t,w}$: Renewable generation in zone z in hour t in scenario w

b. Binary variables

- $\gamma_{k,t,w}$: 1: thermal power plant k is ON in hour t in scenario w ; 0: otherwise
- $\alpha_{k,t,w}$: 1: thermal power plant k is started up in hour t in scenario w ; 0: otherwise
- $\beta_{k,t,w}$: 1: thermal power plant k is shut down in hour t in scenario w ; 0: otherwise

3.2.4 Mathematical model

The expansion planning model can be formulated as the following two-stage stochastic mixed-integer linear programming problem:

$$\begin{aligned}
 \min z = & \sum_{y \in \mathcal{Y}} \left(\sum_{k \in \mathcal{K}_E} \frac{DC_k \delta_{k,y}^-}{(1+r)^{y-\gamma_0}} + \sum_{k \in \mathcal{K}_C} \frac{IC_k^+ \delta_{k,y}^+}{(1+r)^{y-\gamma_0}} \right) \\
 & + \sum_{y \in \mathcal{Y}} \left(\sum_{z \in \mathcal{Z}} \frac{IC_{z,y}^{sol} sol_{z,y}}{(1+r)^{y-\gamma_0}} + \sum_{z \in \mathcal{Z}} \frac{IC_{z,y}^{wind} wind_{z,y}}{(1+r)^{y-\gamma_0}} \right) \\
 & + \sum_{y \in \mathcal{Y}} \sum_{h \in \mathcal{H}_C} \frac{IC_{h,y}^{hydro} \delta_{h,y}}{(1+r)^{y-\gamma_0}} \\
 & + \sum_{y \in \mathcal{Y}} \sum_{b \in \mathcal{B}} \frac{IC_{b,y}^{batt} cap_{b,y}}{(1+r)^{y-\gamma_0}} \\
 & + \sum_{y \in \mathcal{Y}} \sum_{l \in \mathcal{L}_C} \frac{IC_l^{line} \delta_{l,y}}{(1+r)^{y-\gamma_0}} \\
 & + \sum_{y \in \mathcal{Y}} \left[\sum_{k \in \mathcal{K}_E} FC_k (1 - \theta_{k,y}^-) + \sum_{k \in \mathcal{K}_C} FC_k \theta_{k,y}^+ \right] + \sum_{w \in \mathcal{W}} prob_w \left[\sum_{y \in \mathcal{Y}} \sum_{c | year(c)=y} freq_c \sum_{l | c(l)=c} \left(\sum_{k \in \mathcal{K}} CM_{k,y,w} (P_k \gamma_{k,t,w} + P_{k,t,w}) \right. \right. \\
 & + SUC_k \sum_{k \in \mathcal{K}} \alpha_{k,t,w} + \sum_{h \in \mathcal{H}} Cvar_h E_{h,t,w}^{out} \\
 & \left. \left. + \sum_{b \in \mathcal{B}} Cvar_b E_{b,t,w}^{out} + c_{ENP} \sum_{z \in \mathcal{Z}} ENP_{z,t,w} + c_{OG} \sum_{z \in \mathcal{Z}} OG_{z,t,w} \right) \right] \tag{3.1}
 \end{aligned}$$

subject to

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

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

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

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

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

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

$$\sum_{y \in \mathcal{Y}} \delta_{k,y}^+ = \sum_{y \in \mathcal{Y}} \delta_{k1,y}^+ \quad k \in \mathcal{AK}_j, k1 \in \mathcal{AK}_j, j \in J^{AK} \tag{3.8}$$

$$\sum_{k \in \mathcal{M}\mathcal{E}\mathcal{K}_j} \sum_{y \in \mathcal{Y}} \delta_{k,y}^+ \leq 1 \quad j \in J^{\mathcal{M}\mathcal{E}\mathcal{K}} \tag{3.9}$$

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

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

$$\sum_{y \in \mathcal{Y}} \delta_{h,y} \leq 1 \quad h \in \mathcal{H}_C \tag{3.12}$$

$$\sum_{y \in \mathcal{Y}} cap_{b,y} \leq \overline{CAP}_b \quad b \in \mathcal{B} \tag{3.13}$$

$$\delta_{l,y} = 0 \quad l \in \mathcal{L}_C, y \notin [\underline{\tau}_l, \bar{\tau}_l] \tag{3.14}$$

$$\sum_{y \in \mathcal{Y}} \delta_{l,y} = 1 \quad l \in \mathcal{L}_{C1} \tag{3.15}$$

$$\sum_{y \in \mathcal{Y}} \delta_{l,y} \leq 1 \quad l \in \mathcal{L}_{C2} \quad (3.16)$$

$$\sum_{y \in \mathcal{Y}} \delta_{l,y} = \sum_{y \in \mathcal{Y}} \delta_{l_1,y} \quad l \in \mathcal{AL}_j, l_1 \in \mathcal{AL}_j, j \in J^{AC} \quad (3.17)$$

$$\sum_{l \in \mathcal{MEL}_j} \sum_{y \in \mathcal{Y}} \delta_{l,y} \leq 1 \quad j \in J^{MEL} \quad (3.18)$$

$$\theta_{k,y}^- = \sum_{i=1}^y \delta_{k,i}^- \quad k \in \mathcal{K}_E, y \in \mathcal{Y} \quad (3.19)$$

$$\theta_{k,y}^+ = \sum_{i=1}^y \delta_{k,i}^+ \quad k \in \mathcal{K}_C, y \in \mathcal{Y} \quad (3.20)$$

$$\theta_{h,y} = \sum_{i=1}^y \delta_{h,i} \quad h \in \mathcal{H}_C, y \in \mathcal{Y} \quad (3.21)$$

$$\theta_{l,y} = \sum_{i=1}^y \delta_{l,i} \quad l \in \mathcal{L}_C, y \in \mathcal{Y} \quad (3.22)$$

$$\gamma_{k,t,w} \leq 1 - \theta_{k,y}^- \quad k \in \mathcal{K}_E, t \in \mathcal{T}, y = yr(t), w \in \mathcal{W} \quad (3.23)$$

$$\gamma_{k,t,w} \leq \theta_{k,y}^+ \quad k \in \mathcal{K}_C, t \in \mathcal{T}, y = yr(t), w \in \mathcal{W} \quad (3.24)$$

$$\gamma_{k,t,w} \leq 1 - \theta_{k_1,y}^+ \quad k \in \mathcal{K}_E, k_1 = UP(k), t \in \mathcal{T}, y = yr(t), w \in \mathcal{W} \quad (3.25)$$

$$p_{k,t,w} \leq (\bar{P}_k - \underline{P}_k) \gamma_{k,t,w} \quad k \in \mathcal{K}, t \in \mathcal{T}, w \in \mathcal{W} \quad (3.26)$$

$$\gamma_{k,t,w} - \gamma_{k,t-1,w} = \alpha_{k,t,w} - \beta_{k,t,w} \quad k \in \mathcal{K}, t \in (\mathcal{T} - T_{in}), w \in \mathcal{W} \quad (3.27)$$

$$\gamma_{k,t,w} - \gamma_{k,c,w_0} = \alpha_{k,t,w} - \beta_{k,t,w} \quad k \in \mathcal{K}, t \in T_{in}, c = cl(t), w \in \mathcal{W} \quad (3.28)$$

$$\sum_{i=t-MUT_k+1}^t \alpha_{k,i,w} \leq \gamma_{k,t,w} \quad k \in \mathcal{K}, t : hh_t \geq MUT_k, w \in \mathcal{W} \quad (3.29)$$

$$\sum_{i=t-MDT_k+1}^t \beta_{k,i,w} \leq 1 - \gamma_{k,t,w} \quad k \in \mathcal{K}, t : hh_t \geq MDT_k, w \in \mathcal{W} \quad (3.30)$$

$$E_{h,t,w}^{in} \leq E_h^{-in} \quad h \in \mathcal{H}_E, t \in \mathcal{T}, w \in \mathcal{W} \quad (3.31)$$

$$E_{h,t,w}^{out} \leq E_h^{-out} \quad h \in \mathcal{H}_E, t \in \mathcal{T}, w \in \mathcal{W} \quad (3.32)$$

$$E_{h,t,w} \leq EPR_h^{-in} E_h \quad h \in \mathcal{H}_E, t \in \mathcal{T}, w \in \mathcal{W} \quad (3.33)$$

$$E_{h,t,w}^{in} \leq E_h^{-in} \theta_{h,y} \quad h \in \mathcal{H}_C, t \in \mathcal{T}, y = yr(t), w \in \mathcal{W} \quad (3.34)$$

$$E_{h,t,w}^{out} \leq E_h^{-out} \theta_{h,y} \quad h \in \mathcal{H}_C, t \in \mathcal{T}, y = yr(t), w \in \mathcal{W} \quad (3.35)$$

$$E_{h,t,w} \leq EPR_h^{-in} E_h \theta_{h,y} \quad h \in \mathcal{H}_C, t \in \mathcal{T}, y = yr(t), w \in \mathcal{W} \quad (3.36)$$

$$E_{h,t,w} = (1 - \lambda_h) E_{h,t-1,w} + F_{h,t} + \lambda_h^{in} E_{h,t,w}^{in} - \lambda_h^{out} E_{h,t,w}^{out} - sl_{h,t,w} \quad h \in \mathcal{H}, t \in (\mathcal{T} - \mathcal{T}_{in}), w \in \mathcal{W} \quad (3.37)$$

$$E_{h,t,w} = (1 - \lambda_h) E_{h,c0} + F_{h,t} + \lambda_h^{in} E_{h,t,w}^{in} - \lambda_h^{out} E_{h,t,w}^{out} - sl_{h,t,w} \quad h \in \mathcal{H}, t \in \mathcal{T}_{in}, c = cl(t), w \in \mathcal{W} \quad (3.38)$$

$$\hat{E}_{h,ot,w} = \hat{E}_{h,ot-M,w} + E_{h0} |_{ot=M} + \sum_{ot'=ot-M+1}^{ot} \sum_{t \in \text{Map}_{ot,t}} (F_{h,t} + \lambda_h^{in} E_{h,t,w}^{in} - \lambda_h^{out} E_{h,t,w}^{out} - sl_{h,t,w})$$

$$h \in \mathcal{H}, ot \in \mathcal{OT}_M, w \in \mathcal{W} \quad (3.39)$$

$$\hat{E}_{h,ot,w} \leq EPR_h^{-in} E_h \quad h \in \mathcal{H}, ot \in \mathcal{OT}_M, w \in \mathcal{W} \quad (3.40)$$

$$\hat{E}_{h,ot,w} = E_{h0} \quad h \in \mathcal{H}, ot = |\mathcal{OT}|, w \in \mathcal{W} \quad (3.41)$$

$$E_{b,t,w}^{in} \leq \sum_{i=1}^y cap_{b,i} \quad b \in \mathcal{B}, t \in \mathcal{T}, y = yr(t), w \in \mathcal{W} \quad (3.42)$$

$$E_{b,t,w}^{out} \leq \sum_{i=1}^y cap_{b,i} \quad b \in \mathcal{B}, t \in \mathcal{T}, y = yr(t), w \in \mathcal{W} \quad (3.43)$$

$$E_{b,t,w} \leq EPR_b \sum_{i=1}^y cap_{b,i} \quad b \in \mathcal{B}, t \in \mathcal{T}, y = yr(t), w \in \mathcal{W} \quad (3.44)$$

$$E_{b,t,w} = (1 - \lambda_b)E_{b,t-1,w} + \lambda_b^{in} E_{b,t,w}^{in} - \lambda_b^{out} E_{b,t,w}^{out} \quad b \in \mathcal{B}, t \in (\mathcal{T} - \mathcal{T}_{in}), w \in \mathcal{W} \quad (3.45)$$

$$E_{b,t,w} = (1 - \lambda_b)E_{b,c0} + \lambda_b^{in} E_{b,t,w}^{in} - \lambda_b^{out} E_{b,t,w}^{out} \quad b \in \mathcal{B}, t \in \mathcal{T}_{in}, c = cl(t), w \in \mathcal{W} \quad (3.46)$$

$$E_{b,t,w} = E_{b,c0} \quad b \in \mathcal{B}, t : hh_t = 24, c = cl(t), w \in \mathcal{W} \quad (3.47)$$

$$\underline{F}_l \leq x_{l,t,w} \leq \bar{F}_l \quad l \in \mathcal{L}_E, t \in \mathcal{T}, w \in \mathcal{W} \quad (3.48)$$

$$\theta_{l,y} \underline{F}_l \leq x_{l,t,w} \leq \theta_{l,y} \bar{F}_l \quad l \in \mathcal{L}_C, t \in \mathcal{T}, y = yr(t), w \in \mathcal{W} \quad (3.49)$$

$$\begin{aligned} & \sum_{k \in \Omega_z^k} (\underline{P}_k \gamma_{k,t,w} + p_{k,t,w}) + \mu_{z,t} \left(sol_{z,0} + \sum_{i=1}^y sol_{z,i} \right) \\ & + \rho_{z,t} \left(wind_{z,0} + \sum_{i=1}^y wind_{z,i} \right) + \sum_{l|r(t)=z} x_{l,t,w} \\ & + \sum_{h \in \Omega_z^h} E_{h,t,w}^{out} + \sum_{b \in \Omega_z^b} E_{b,t,w}^{out} + ENP_{z,t,w} = D_{z,t} \\ & + \sum_{l|s(t)=z} x_{l,t,w} + \sum_{h \in \Omega_z^h} E_{h,t,w}^{in} + \sum_{b \in \Omega_z^b} E_{b,t,w}^{in} + OG_{z,t,w} \\ & z \in \mathcal{Z}, t \in \mathcal{T}, y = yr(t), w \in \mathcal{W} \end{aligned} \quad (3.50)$$

$$\begin{aligned} & \sum_{k \in \mathcal{K}_E} \left[\bar{P}_k (1 - \theta_{k,y}^-) - \underline{P}_k \gamma_{k,t,w} - p_{k,t,w} \right] \\ & + \sum_{k \in \mathcal{K}_C} \left(\bar{P}_k \theta_{k,y}^+ - \underline{P}_k \gamma_{k,t,w} - p_{k,t,w} \right) \geq R_{z,t} \quad z \in \mathcal{Z}, t \in \mathcal{T}, w \in \mathcal{W} \end{aligned} \quad (3.51)$$

$$\begin{aligned} & \sum_{z|ma(z)=m} \sum_{k \in \Phi_{z,f}} \sum_{c|year(c)=y} freq_c \sum_{t|cl(t)=c} \frac{HR_k(\underline{P}_k \gamma_{k,t,w} + p_{k,t,w})}{ECnt_f} \leq \bar{FA}_{f,m,y} \\ & f \in \mathcal{F}, m \in \mathcal{M}, y \in \mathcal{Y}, w \in \mathcal{W} \end{aligned} \quad (3.52)$$

$$\begin{aligned} & \sum_{z|ma(z)=m} \sum_{f \in \mathcal{F}} \sum_{k \in \Phi_{z,f}} \sum_{c|year(c)=y} freq_c \sum_{t|cl(t)=c} HR_k(\underline{P}_k \gamma_{k,t,w} + p_{k,t,w}) co_{2f} \leq \bar{CO}_{2m,y} \\ & m \in \mathcal{M}, y \in \mathcal{Y}, w \in \mathcal{W} \end{aligned} \quad (3.53)$$

$$RES_{z,t,w} = \mu_{z,t} \left(sol_{z,0} + \sum_{i=1}^{yr(t)} sol_{z,i} \right) + \rho_{z,t} \left(wind_{z,0} + \sum_{i=1}^{yr(t)} wind_{z,i} \right) + \sum_{h \in \Omega_z^h} E_{h,t,w}^{out} \quad z \in \mathcal{Z}, t \in \mathcal{T}, w \in \mathcal{W} \tag{3.54}$$

$$\frac{\sum_{z|ma(z)=m} \sum_{c|year(c)=y} freq_c \sum_{t|cl(t)=c} RES_{z,t,w}}{\sum_{z|ma(z)=m} \sum_{c|year(c)=y} freq_c \sum_{t|cl(t)=c} D_{z,t}} \geq \varphi_{m,y} \quad m \in \mathcal{M}, y \in \mathcal{Y}, w \in \mathcal{W} \tag{3.55}$$

$$\delta_{k,y}^-, \theta_{k,y}^- \in \{0,1\} \quad k \in \mathcal{K}_E, y \in \mathcal{Y} \tag{3.56}$$

$$\delta_{k,y}^+, \theta_{k,y}^+ \in \{0,1\} \quad k \in \mathcal{K}_C, y \in \mathcal{Y} \tag{3.57}$$

$$\delta_{h,y}, \theta_{h,y} \in \{0,1\} \quad h \in \mathcal{H}_C, y \in \mathcal{Y} \tag{3.58}$$

$$\delta_{l,y}, \theta_{l,y} \in \{0,1\} \quad l \in \mathcal{L}_C, y \in \mathcal{Y} \tag{3.59}$$

$$sol_{z,y}, wind_{z,y} \geq 0 \quad z \in \mathcal{Z}, y \in \mathcal{Y} \tag{3.60}$$

$$cap_{b,y} \geq 0 \quad b \in \mathcal{B}, y \in \mathcal{Y} \tag{3.61}$$

$$\alpha_{k,t,w}, \beta_{k,t,w}, \gamma_{k,t,w} \in \{0,1\} \quad k \in \mathcal{K}, t \in \mathcal{T}, w \in \mathcal{W} \tag{3.62}$$

$$p_{k,t,w} \geq 0 \quad k \in \mathcal{K}, t \in \mathcal{T}, w \in \mathcal{W} \tag{3.63}$$

$$E_{h,t,w}, E_{h,t,w}^{in}, E_{h,t,w}^{out}, sl_{h,t,w} \geq 0 \quad h \in \mathcal{H}, t \in \mathcal{T}, w \in \mathcal{W} \tag{3.64}$$

$$\hat{E}_{h,ot,w} \geq 0 \quad h \in \mathcal{H}, ot \in \mathcal{OT}, w \in \mathcal{W} \tag{3.65}$$

$$E_{b,t,w}, E_{b,t,w}^{in}, E_{b,t,w}^{out} \geq 0 \quad b \in \mathcal{B}, t \in \mathcal{T}, w \in \mathcal{W} \tag{3.66}$$

$$ENP_{z,t,w}, OG_{z,t,w}, RES_{z,t,w} \geq 0 \quad z \in \mathcal{Z}, t \in \mathcal{T}, w \in \mathcal{W} \tag{3.67}$$

$$x_{l,t,w} \text{ free variable} \quad l \in \mathcal{L}, t \in \mathcal{T}, w \in \mathcal{W} \tag{3.68}$$

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

1. $\sum_{y \in \mathcal{Y}} \left(\sum_{k \in \mathcal{K}_E} \frac{DC_k \delta_{k,y}^-}{(1+r)^{y-\gamma_0}} + \sum_{k \in \mathcal{K}_C} \frac{IC_k^{th} \delta_{k,y}^+}{(1+r)^{y-\gamma_0}} \right)$ are the annualized decommissioning costs of existing thermal power plants and investment costs in new thermal power generation;
2. $\sum_{y \in \mathcal{Y}} \left(\sum_{z \in \mathcal{Z}} \frac{IC_{z,y}^{sol} \delta_{z,y}}{(1+r)^{y-\gamma_0}} + \sum_{z \in \mathcal{Z}} \frac{IC_{z,y}^{wind} \delta_{z,y}}{(1+r)^{y-\gamma_0}} \right)$ are the annualized investment costs in new solar and wind capacity;
3. $\sum_{y \in \mathcal{Y}} \left(\sum_{h \in \mathcal{H}_C} \frac{IC_{h,y}^{hydro} \delta_{h,y}}{(1+r)^{y-\gamma_0}} \right)$ are the annualized investment costs in new hydropower plants;
4. $\sum_{y \in \mathcal{Y}} \left(\sum_{b \in \mathcal{B}} \frac{IC_{b,y}^{batt} \delta_{b,y}}{(1+r)^{y-\gamma_0}} \right)$ are the annualized investment costs in new batteries capacity;
5. $\sum_{y \in \mathcal{Y}} \sum_{l \in \mathcal{L}_C} \frac{IC_l^{line} \delta_{l,y}}{(1+r)^{y-\gamma_0}}$ 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. $\sum_{w \in \mathcal{W}} prob_w \left[\sum_{y \in \mathcal{Y}} \sum_{c | year(c)=y} freq_c \sum_{t | cl(t)=c} \left(\sum_{k \in \mathcal{K}} CM_{k,y,w} \left(\underline{P}_k \gamma_{k,t,w} + P_{k,t,w} \right) + \sum_{k \in \mathcal{K}} SUC_k \alpha_{k,t,w} + \sum_{h \in \mathcal{H}} Cvar_h E_{h,t,w}^{out} + \sum_{b \in \mathcal{B}} Cvar_b E_{b,t,w}^{out} + c_{ENP} \sum_{z \in \mathcal{Z}} ENP_{z,t,w} + c_{OG} \sum_{z \in \mathcal{Z}} OG_{z,t,w} \right) \right]$ are the second-stage costs, i.e., the operating 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 $CM_{k,y,w}$ the slopes of these linear relationships. In each scenario w , the marginal cost of thermal plant k in year y is computed as:

$$CM_{k,y,w} = O\&M_k + HR_k (Fuel.Price_{fuel(k),y,w} + co_{2fuel(k)} CO_2.Price_{y,w})$$

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 , $Fuel.Price_{fuel(k),y,w}$ [€/Gcal] the price in year y under scenario w of fuel used by unit k , $CO_{2fuel(k)}$ [ton/Gcal] the CO_2 emission rate of fuel used by unit k and $CO_2.Price_{y,w}$ [€/ton] the emission cost in year y under scenario w .

In the proposed model there are three groups of constraints, namely investment constraints (3.2)–(3.22), operational constraints (3.23)–(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 Eqs. (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. Constraints (3.12) model investments in new hydropower plants as optional decisions, while inequalities (3.13) impose an upper bound on batteries installed capacity at the end of the planning horizon. The group of inequalities (3.14)–(3.18) model investment decisions on new transmission lines. Specifically, Eqs. (3.14) impose a temporal window for the introduction of a new transmission line, while constraints (3.15) and (3.16) model, respectively, mandatory and optional decisions. Equations (3.17) are the associate line constraints, indicating that a group of lines is subject to a single investment decision, while the mutually exclusive line constraints (3.18) ensure that only one interconnection (or none) of the transmission lines in a given group (\mathcal{MEL}_j) is built. Finally, Eqs. (3.19)–(3.22) 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 Eqs. (3.23)–(3.30) models the thermal component of the energy system. Constraints (3.23) ensure consistency between the binary variables representing the commitment status and those representing the decommissioning decisions, forcing the existing power plants decommissioned within year y ($\theta_{k,y}^- = 1$) to be offline in all hours after decommission. Constraints (3.24) and (3.25) enforce the consistency between the binary variables representing the commitment status and those representing investment decisions. Indeed, inequalities (3.24) 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 ($\gamma_{k,t,w} = 0$) in all hours of year y . Inequalities (3.25) model the reinforcements of existing thermal power plants. Specifically, let $k1 = UP(k)$ denote the new project that replaces the existing unit k when it starts operating: building project $k1$ within year y ($\theta_{k1,y}^+ = 1$) implies the permanent offline status of unit k ($\gamma_{k,t,w} = 0$). Inequalities (3.26) state that the power output $p_{k,t,w}$ above the minimum power output \underline{P}_k is either bounded above by $\overline{P}_k - \underline{P}_k$, if unit k is online ($\gamma_{k,t,w} = 1$), or zero if unit k is offline ($\gamma_{k,t,w} = 0$). Constraints (3.27) enforce consistency between the binary variables $\alpha_{k,t,w}$, $\beta_{k,t,w}$, $\gamma_{k,t-1,w}$ and $\gamma_{k,t,w}$ that represent start-up, shut down and status in adjacent hours, respectively, in all hours of the representative days, except the first hour of the day. Constraints (3.28) enforce consistency between $\alpha_{k,t,w}$, $\beta_{k,t,w}$, $\gamma_{k,c,w0}$ and $\gamma_{k,t,w}$, for the first hour of every representative day, where the parameter $\gamma_{k,c,w0}$ represents the status of unit k at the beginning of representative day c under scenario w . In our numerical experiments, the values assigned to $\gamma_{k,c,w0}$ are determined by applying a classification tree

trained on historical data. The description of this procedure is outside the scope of this work and we refer the reader to [32] for a detailed description of how the values of parameters γ_{k,c,w_0} are computed. The minimum uptime constraints (3.29) impose that unit k can be started-up at most once in an interval of MUT_k consecutive time periods. For each representative day, the minimum uptime constraints are enforced on the hours in the range from MUT_k to the final (the 24th) hour of the day, to keep the representative days separate in the description of future scenarios. The minimum downtime constraints (3.30) work similarly to the shut-down of thermal power plants.

Constraints (3.31)–(3.41) model the operation of hydro plants. Specifically, constraints (3.31)–(3.33) bound the pumping power, the power output and the energy level of the existing hydroelectric reservoirs below their respective upper limits. If candidate hydro plant $h \in \mathcal{H}_C$ is built within year y , inequalities (3.34)–(3.36) define the upper bounds to pumping power, power output and energy level of new hydroelectric reservoirs, otherwise set the corresponding variables to zero. Energy balances (3.37) apply to all hours but the first of the representative days: they ensure that the energy stored by hydro plant h at the end of hour t equals the energy stored at the end of hour $t - 1$ (reduced by the loss coefficient $\lambda_h \leq 1$), plus the natural inflows, plus the energy injected in h (reduced by the coefficient $\lambda_h^{in} \leq 1$), minus the energy released from the reservoir (reduced by the coefficient $\lambda_h^{out} \geq 1$), minus the spillage. Equations (3.38) impose the energy balance for the first hour of each representative day. Equations (3.39)–(3.41) model the long-term operations of hydro plants. Specifically, constraints (3.39) and (3.40) create the continuity in storage across the entire time horizon: this is done by checking at regular intervals (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. This is possible because, as a result of the clustering procedure that determines the representative days, we know the representative hour associated with each hour of the planning horizon. This information is included in the model using the set $\text{Map}_{ot,t}$. We refer the reader to [29] for a complete description of these long-term constraints. 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 the first-stage and second-stage 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 $\lambda_b^{in} \leq 1$), minus the energy released from the battery (reduced by the coefficient $\lambda_b^{out} \geq 1$). 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 energy 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 energy flows on candidate transmission lines and the binary variables related to investment decisions, not allowing energy flows on candidate lines which have not been built ($\theta_{l,y} = 0$). The zonal balance Eqs. (3.50) impose equality between energy sources and use in every zone and every hour. 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 sides describe the energy uses (represented by the load, outgoing energy flows, pumping power and energy absorbed by batteries). The variables $ENP_{z,t,w}$ and $OG_{z,t,w}$ 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 (either existing plants not yet decommissioned or newly constructed 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 , each fuel f , and each scenario w . In particular, they are computed by multiplying the daily consumption of fuel f under scenario w 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 CO₂ 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. Equation (3.54) compute the zonal hourly renewable production, by considering solar, wind and hydroelectric 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.

4 Decomposition 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 [34] that allows solving a linear programming problem with complicating variables in a distributed manner at the cost of iterations [35]. In recent years, this algorithm has been widely applied to two-stage stochastic programming models [36–38]. 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. For instance, in [39] 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 [40] a Benders decomposition algorithm is used to solve a network-constrained AC unit commitment problem under uncertainty. Finally, references [41] and [42] 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 $\theta_{k,y}^-$, $\theta_{k,y}^+$, $wind_{z,y}$, $sol_{z,y}$, $\theta_{h,y}$, $cap_{b,y}$ and $\theta_{l,y}$. 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.

4.1 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 mixed-integer linear programming model (referred to as master problem (4) in the sequel) whose optimization variables are $\delta_{k,y}^-$, $\theta_{k,y}^-$, $\delta_{k,y}^+$, $\theta_{k,y}^+$, $sol_{z,y}$, $wind_{z,y}$, $\delta_{h,y}$, $\theta_{h,y}$, $cap_{b,y}$, $\delta_{l,y}$, $\theta_{l,y}$ and σ_w .

$$\begin{aligned} \min z_{down} = & \sum_{y \in \mathcal{Y}} \left(\sum_{k \in \mathcal{K}_E} \frac{DC_k \delta_{k,y}^-}{(1+r)^{y-\gamma_0}} + \sum_{k \in \mathcal{K}_C} \frac{IC_k^{th} \delta_{k,y}^+}{(1+r)^{y-\gamma_0}} \right) \\ & + \sum_{y \in \mathcal{Y}} \left(\sum_{z \in \mathcal{Z}} \frac{IC_{z,y}^{sol} sol_{z,y}}{(1+r)^{y-\gamma_0}} + \sum_{z \in \mathcal{Z}} \frac{IC_{z,y}^{wind} wind_{z,y}}{(1+r)^{y-\gamma_0}} \right) \\ & + \sum_{y \in \mathcal{Y}} \sum_{h \in \mathcal{H}_C} \frac{IC_{h,y}^{hydro} \delta_{h,y}}{(1+r)^{y-\gamma_0}} + \sum_{y \in \mathcal{Y}} \sum_{b \in \mathcal{B}} \frac{IC_{b,y}^{batt} cap_{b,y}}{(1+r)^{y-\gamma_0}} \\ & + \sum_{y \in \mathcal{Y}} \sum_{l \in \mathcal{L}_C} \frac{IC_l^{line} \delta_{l,y}}{(1+r)^{y-\gamma_0}} + \sum_{y \in \mathcal{Y}} \left(\sum_{k \in \mathcal{K}_C} FC_k \cdot \theta_{k,y}^+ + \sum_{k \in \mathcal{K}_E} FC_k (1 - \theta_{k,y}^-) \right) + \sum_w prob_w \cdot \sigma_w \end{aligned} \quad (4.1)$$

subject to

$$\begin{aligned}
 \sigma_w \geq & \sum_y z_{y,w}^{(v)} + \sum_y \sum_{k \in \mathcal{K}_E} \lambda_{k,y,w}^{\theta^-(v)} \left(\theta_{k,y}^- - \theta_{k,y}^{-(v)} \right) \\
 & + \sum_y \sum_{k \in \mathcal{K}_C} \lambda_{k,y,w}^{\theta^+(v)} \left(\theta_{k,y}^+ - \theta_{k,y}^{+(v)} \right) + \sum_y \sum_z \lambda_{z,y,w}^{sol^{(v)}} \left(sol_{z,y} - sol_{z,y}^{(v)} \right) \\
 & + \sum_y \sum_z \lambda_{z,y,w}^{wind^{(v)}} \left(wind_{z,y} - wind_{z,y}^{(v)} \right) \tag{4.2} \\
 & + \sum_y \sum_{h \in \mathcal{H}_C} \lambda_{h,y,w}^{\theta_h^{(v)}} \left(\theta_{h,y} - \theta_{h,y}^{(v)} \right) + \sum_y \sum_{b \in \mathcal{B}} \lambda_{b,y,w}^{cap^{(v)}} \left(cap_{b,y} - cap_{b,y}^{(v)} \right) \\
 & + \sum_y \sum_{l \in \mathcal{L}_C} \lambda_{l,y,w}^{\theta_l^{(v)}} \left(\theta_{l,y} - \theta_{l,y}^{(v)} \right) w \in \mathcal{W}, v = 1, \dots, j-1
 \end{aligned}$$

$$\sigma_w \geq \sigma^{down} \quad w \in \mathcal{W} \tag{4.3}$$

$$(3.2) - (3.22) \tag{4.4}$$

$$(3.56) - (3.61) \tag{4.5}$$

The objective function (4.1) includes investment, decommissioning and fixed costs and the auxiliary variables σ_w that lower-approximate the operation cost under scenario w . Therefore, the solution of this master problem represents a lower bound for the optimal solution of the original problem. 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 $\theta_{k,y}^{-(j)}$, $\theta_{k,y}^{+(j)}$, $sol_{z,y}^{(j)}$, $wind_{z,y}^{(j)}$, $\theta_{h,y}^{(j)}$, $cap_{b,y}^{(j)}$ and $\theta_{l,y}^{(j)}$. These parameters will be used to build constraints (4.2) in subsequent iterations, along with the optimal value of subproblems objective function $z_{y,w}^{(j)}$ (5.1) and dual variable vectors $\lambda_{y,w}^{(j)}$, obtained from fixing constraints (5.3)–(5.9) in the subproblems at iteration j , as described in the next paragraph. Specifically, the constraints (4.2), 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 [34], in our approach we implement a multi-cut strategy, generating at each iteration one cut per scenario. As observed in [23, 43] and [44], also in our application, the multi-cut Benders decomposition showed a faster convergence than the mono-cut algorithm. Moreover, 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 $ENP_{z,t,w}$ and $OG_{z,t,w}$ that model energy not provided and overgeneration (i.e., energy in excess), respectively, the subproblems are always feasible. Lower bound constraints (4.3) on variables σ_w avoid the master problem being unbounded in the first iteration. Constraints (4.4) control investment and decommissioning decisions as in the original problem, while constraints (4.5) define optimization variables.

4.2 Subproblems

At each iteration j , for given values of the first-stage variables $\theta_{k,y}^{+(j)}$, $\theta_{k,y}^{-(j)}$, $\theta_{h,y}^{(j)}$, $sol_{z,y}^{(j)}$, $wind_{z,y}^{(j)}$, $cap_{b,y}^{(j)}$ and $\theta_{l,y}^{(j)}$ the subproblem associated with year y and scenario w (referred to as subproblem (5) in the sequel) is formulated as follows.

$$\begin{aligned} \min_{z,y,w} = & \sum_{c|year(c)=y} freq_c \left[\sum_{t|c(t)=c} \left(\sum_{k \in \mathcal{K}} (\gamma_{k,t,w} P_k + p_{k,t,w}) CM_{k,y,w} + SUC_k \sum_{k \in \mathcal{K}} \alpha_{k,t,w} \right. \right. \\ & + \sum_{h \in \mathcal{H}} Cvar_h \cdot E_{h,t,w}^{out} \\ & \left. \left. + \sum_{b \in \mathcal{B}} Cvar_b \cdot E_{b,t,w}^{out} + c_{ENP} \sum_{z \in \mathcal{Z}} ENP_{z,t,w} + c_{OG} \sum_{z \in \mathcal{Z}} OG_{z,t,w} \right) \right] \end{aligned} \tag{5.1}$$

subject to

$$(3.23) - (3.55) \quad \forall t | year(t) = y \tag{5.2}$$

$$\theta_{k,y}^+ = \theta_{k,y}^{+(j)} : \lambda_{k,y,w}^{\theta^+} \quad k \in \mathcal{K}_C \tag{5.3}$$

$$\theta_{k,y}^- = \theta_{k,y}^{-(j)} : \lambda_{k,y,w}^{\theta^-} \quad k \in \mathcal{K}_E \tag{5.4}$$

$$\theta_{l,y} = \theta_{l,y}^{(j)} : \lambda_{l,y,w}^{\theta_l} \quad l \in \mathcal{L}_C \tag{5.5}$$

$$wind_{z,y} = wind_{z,y}^{(j)} : \lambda_{z,y,w}^{wind} \quad z \in \mathcal{Z} \tag{5.6}$$

$$sol_{z,y} = sol_{z,y}^{(j)} : \lambda_{z,y,w}^{sol} \quad z \in \mathcal{Z} \tag{5.7}$$

$$cap_{b,y} = cap_{b,y}^{(j)} : \lambda_{b,y,w}^{cap} \quad b \in \mathcal{B} \tag{5.8}$$

$$\theta_{h,y} = \theta_{h,y}^{(j)} : \lambda_{h,y,w}^{\theta_h} \quad h \in \mathcal{H}_C \tag{5.9}$$

$$0 \leq \alpha_{k,t,w}, \beta_{k,t,w}, \gamma_{k,t,w} \leq 1 \quad k \in \mathcal{K}, t \in \mathcal{T}, w \in \mathcal{W} \tag{5.10}$$

$$(3.63) - (3.68) \tag{5.11}$$

Objective function (5.1) minimizes the operating cost of the system in year y under scenario w . Constraints (5.2) include all operating constraints in the original problem and are imposed for each hour t belonging to the considered year y . Equations (5.3)–(5.9) fix the complicating variables to values determined by the master

problem, while variables $\lambda_{k,y,w}^{\theta^+}$, $\lambda_{k,y,w}^{\theta^-}$, $\lambda_{l,y,w}^{\theta_l}$, $\lambda_{z,y,w}^{wind}$, $\lambda_{z,y,w}^{sol}$, $\lambda_{p,y,w}^{cap}$ and $\lambda_{h,y,w}^{\theta_h}$ 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 $z_{y,w}^{(j)}$ and $\lambda_{y,w}^{(j)}$, respectively. Both these parameters are needed to add Benders optimality cuts (4.2) to the master problem. Finally, constraints (5.10) and (5.11) define the optimization variables.

It is worth mentioning that constraints (5.10) replace the original definition (3.62) imposed on thermal commitment variables, i.e., in the subproblems the second-stage binary variables that describe thermal power plants activation patterns (i.e., $\alpha_{k,t,w}$, $\beta_{k,t,w}$ and $\gamma_{k,t,w}$) 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. We consider a reasonable optimality gap tolerance of 0.1% (i.e., $\frac{z_{MILP}^* - z_{LP}^*}{z_{LP}^*} \leq 0.001$). Empirical results show how the two solutions z_{LP}^* and z_{MILP} are very close, being the relative distance in the current application 0.02% 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 [45] and [46], 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

$$z_{up}^{(j)} = z_{down}^{(j)} - \sum_w prob_w \cdot \sigma_w^{(j)} + \sum_w prob_w \sum_y z_{y,w}^{(j)}$$

4.3 Solution algorithm based on Benders decomposition

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

1. *Initialization* Initialize the iteration counter, set $j = 1$. Set $z_{up}^{(j)} = \infty$ and $z_{down}^{(j)} = -\infty$.

2. *Master problem solution* Solve the master problem (4). Update $z_{down}^{(j)}$ and the values of first-stage variables.
3. *First year* Consider the first year of the planning horizon, i.e., $y = 1$.
4. *First scenario* Consider the first scenario, i.e., $w = 1$.
5. *Subproblem solution* Solve subproblem (5) for year y and scenario w . Compute $z_{y,w}^{(j)}$ and store the dual variables of the fixing constraints.
6. *Scenario update* Consider the next scenario and repeat step 4. If all scenarios have been considered go to step 6.
7. *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.
8. *Convergence checking* Compute $z_{up}^{(j)}$. If $\frac{|z_{up}^{(j)} - z_{down}^{(j)}|}{z_{up}^{(j)}} < \varepsilon$, 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.
9. *Subproblems final integer solution* For each year y and each scenario w , solve subproblems (5) replacing constraints (5.10) with (3.62), i.e., considering mixed-integer linear problems. The solution obtained is now feasible for the original problem.

5 Numerical study

In this section, we describe the analysis conducted with the proposed model on the Italian energy system to assess the achievement at 2040 of policy goals set by the European Commission. A period of 21 years (planning horizon from 2020 to 2040) is simulated on the system divided into seven market zones. The simulation assumptions hereafter summarized and described are based on public information provided by ENTSO-E, ENTSOG, Terna, GME and EERA [47–51]. Input data and numerical results of the case study are described in the following subsections.

5.1 Data for Italian energy system

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. Figure 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.

For modelling short-term operation, we have applied the procedure described in Sect. 2 by fixing at 5% the threshold for the system average MAPE 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 2.

Figure 2 illustrates the five representative days in year 2020 for the North zone, which has the highest electricity demand. As can be noted, each of the representative days is characterized by 24 values for load, solar capacity factors, and wind capacity factors and by a specific weight.

Representative days for the following years of the planning horizon have been obtained by applying an annual average demand growth of 1%, as shown in Fig. 3.

As far as renewable installed capacity growth is concerned, a lower bound of 50% of RES penetration (calculated as the ratio between renewable production and expected demand) has been imposed for year 2040. Based on this lower bound set by the Italian Energy Plan [48], 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

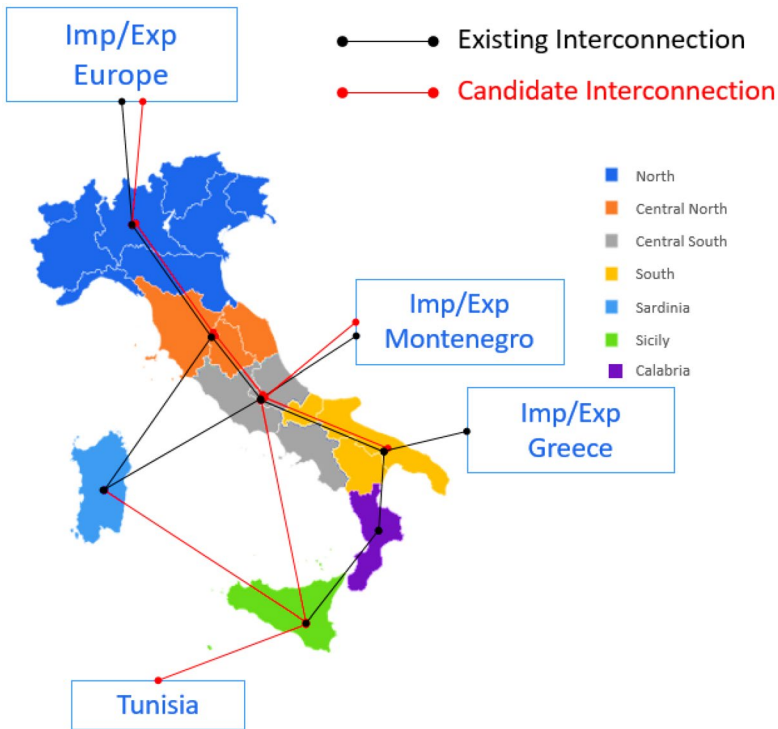


Fig. 1 Existing and candidate interconnections in the Italian power system

information and decommissioning details. 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 resumes the above-mentioned assumptions.

Regarding fuel consumption, we have considered a CO₂ 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 4 resumes the technical data related to candidate storage projects.

The investment cost is assumed to decrease in the period 2020–2040, as stated in [50], with an exponentially decreasing trend as shown in Fig. 4.

For the economic factors, the values related to Investment Costs (IC), Fixed Costs and Decommissioning Costs are shown below, for both conventional thermal power plants and photovoltaic and wind power plants (see Table 5).

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 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 CO₂ 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 CO₂ 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 CO₂ and the fuel price. Indeed, we have used prices scenarios described in [47], 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 CO₂ prices). These scenarios are summarized in Tables 6 and 7. It is worth mentioning that codes Gas, Gasoil, and Coal in Table 7 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 2 Number of representative days for different threshold values

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

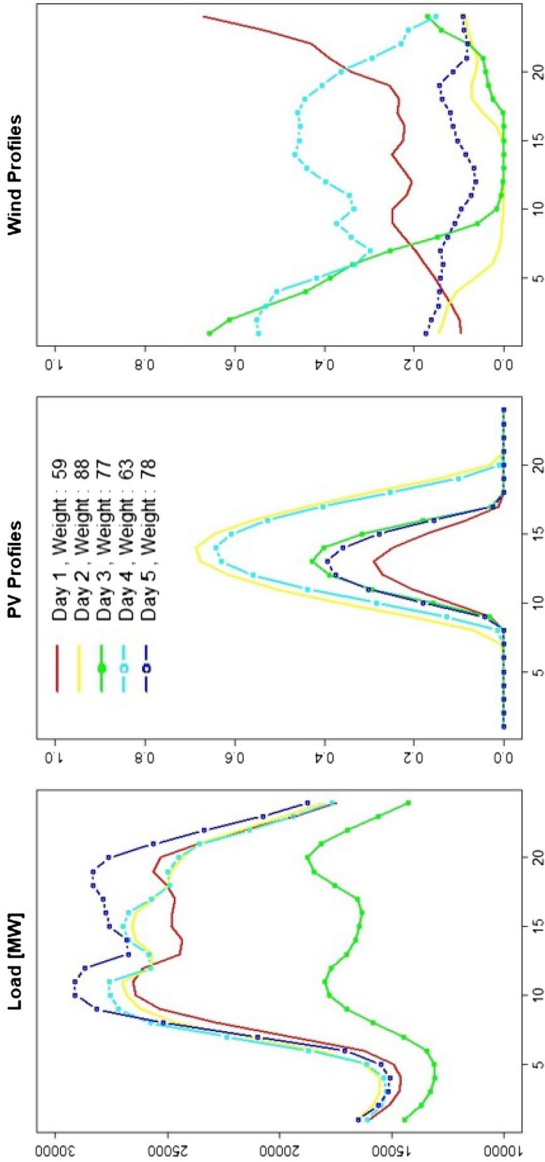


Fig. 2 Daily load, solar and wind profiles for the North zone in the five representative days in year 2020

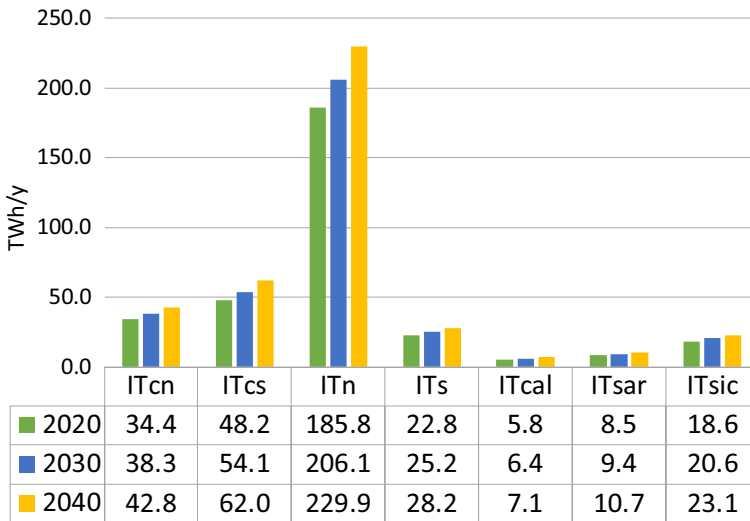


Fig. 3 Zonal load [TWh/year]

Table 3 Installed, outgoing and incoming capacity [GW]

[GW]	CCGT	COAL	GT	TOT
Installed in 2020	37.75	8.15	2.57	48.47
Outgoing capacity 2025/2040	3.57	8.15	1.45	13.17
Incoming capacity 2025/2040	11.20	–	15.45	26.65

Table 4 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
\bar{E}^{out} [MW]	1000	450	1250	800	480	600	600
\bar{E}^{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

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 [52] integrated with solver Gurobi under GAMS 24.7.4. We considered $\varepsilon = 10^{-4}$ as tolerance for Benders decomposition convergence, while in each

Fig. 4 Batteries investment cost trend

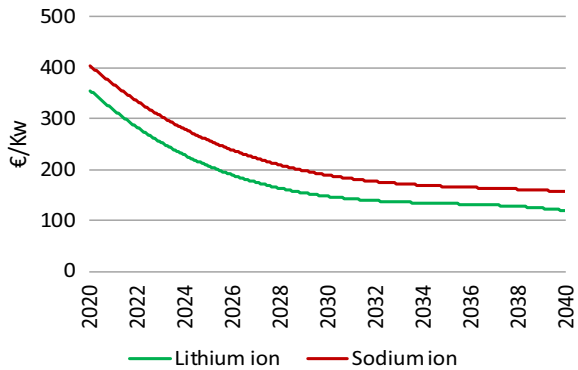


Table 5 Economic factors for conventional power plants

	CCGT	COAL	GT	PV	Wind
IC [€/kW]	800	2600	400	1000	1300
Fixed Cost [€/MW/year]	10	32	5	-	-
Decommissioning [€/kW]	6	20	3	-	-

Table 6 CO₂ prices [€/ton] in different scenarios

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

iteration we solved the master problem up to optimality. Figure 5 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.

The total time needed to solve the problem is 19,032 s, corresponding to 5 h 17 min and 12 s. Table 8 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 8, while the base case solution requires almost two minutes, each of the updated subproblems is

Table 7 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

solved in only 2.622 s on average. Since the number of subproblems equals 126, at each iteration, the average time required to solve all the subproblems is 440.242 s.

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: a higher number of scenarios would not exponentially increase computational times. Specifically, Table 9 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 h, a result compatible with time requirements in investment studies.

As explained in the previous section, since in each iteration subproblems are defined as linear problems, second-stage decisions $\alpha_{k,t,w}$, $\beta_{k,t,w}$ and $\gamma_{k,t,w}$, 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

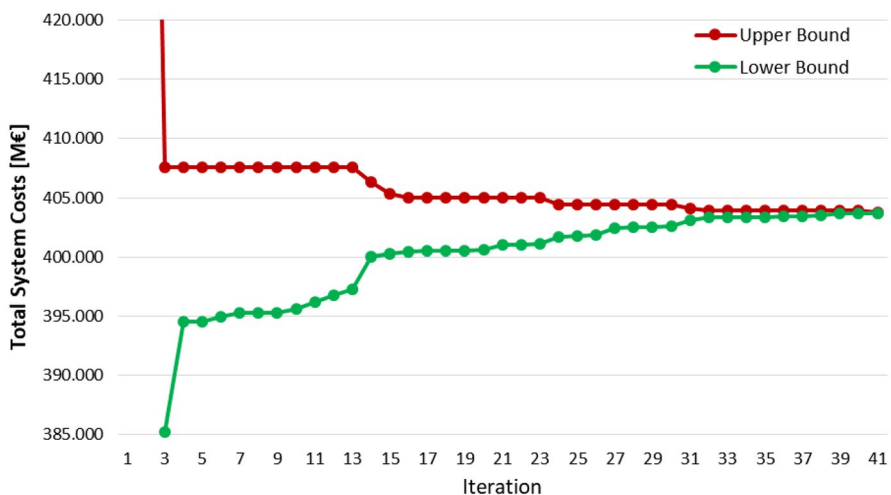
**Fig. 5** Upper and lower bounds values over iterations in multi-cut Benders algorithm

Table 8 Size and solution time of master problem and subproblems at the last iteration of Benders algorithm

	# Constraints	# Decision variables	# Discrete variables	CPU time [s]
Master problem	9665	12,859	5670	27.408
first subproblem (base case)	182,179	149,408	0	112.492
Updated subproblem	182,179	149,408	0	2.622

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€.

The system expected costs for the whole expansion planning period are shown in Table 10. 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, which include O&M and fuel consumption, 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 CO₂ and fuel prices. Specifically, Table 11 describes how system costs vary depending on stochastic prices.

As can be noticed, for each fuel price scenario A and B, the three CO₂ 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 CO₂ 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 CO₂ prices to 382,048 M€ in the fuel price scenario B with high CO₂ prices.

Table 12 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 9 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 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 13.

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 14. 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.

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.

Figure 6 reports the expected energy generation in 2040 divided by energy source, for each Italian zone. As can be noticed, since the RES production (i.e., from wind, solar and hydro sources) is 218 TWh/year while the load is 404 TWh/year, the target of reaching 50% 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, which is almost 60% of the Italian electricity demand. Even though generation exceeds demand in some market zones, in 2040 Italy will import 17 TWh from neighboring countries.

Table 10 Breakdown of system costs

Costs	M€
Thermal capacity expansion	2988
Wind capacity expansion	14,128
Solar capacity expansion	33,832
Transmission capacity expansion	1627
Pump units capacity expansion	3480
Batteries capacity expansion	4554
Decommissioning costs	211
Thermal fixed costs	8
Expected thermal production cost	341,619
Expected start-up costs	399
Expected hydro operation costs	170
Expected batteries operation costs	766
Penalties for overgeneration	28
Penalties for energy not provided	0
Total system expected costs	403,810

Table 11 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

Table 12 Renewable generation capacity expansion [GW] divided by source and implementation year

Year	Wind	PV
2020	3.76	7.67
2021	0.77	8.31
2022	0.77	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

Figure 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

Table 13 Candidate interconnections selected by the model

From	To	Transmission limits	Year of intervention
Montenegro	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

wind expansion is mainly located in southern regions, which are characterized by the highest wind capacity factors.

6 Conclusions

In this paper, we have introduced a two-stage stochastic mixed-integer linear programming model for power generation and transmission expansion planning. The main characteristics of the proposed model are:

Table 14 Installed capacity of energy storage systems [MW]

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

- Identification of the least-cost investment schedule for investment in generating facilities, transmission lines, and storage systems and for decommissioning of existing thermal power plants.
- High level of temporal detail to accurately capture short-term volatility of intermittent energy sources.
- High level of technical detail, integrating into the expansion model the unit commitment constraints on a plant-by-plant level and accounting for both short-term and long-term storage to provide an accurate representation of the system's operation.
- Modelling of long-term uncertainty, considering three scenarios for CO₂ prices combined with two scenarios for fuel prices, therefore considering six scenarios for production costs of thermal units.
- Consideration of policy goals such as CO₂ emission limits and renewable penetration targets in expansion planning.

To be computationally tractable, the optimization model has been integrated with a clustering analysis to select representative days considering the correlation between different time series (i.e., daily zonal electricity demands and renewable capacity factors). To further reduce computational costs, a multi-cut Benders decomposition algorithm has been implemented, decomposing the original two-stage stochastic programming model by year and by scenario.

The proposed model presents many applications in power systems analysis, allowing in general to address what-if questions. For example, in this paper, we have applied the proposed model to evaluate the possibility for the Italian energy system to achieve by 2040 a 50% renewable penetration target and a sustainable reduction of CO₂ emissions, according to the goals set by the European Commission. Empirical results show how solar PV technology could play a key role in achieving these

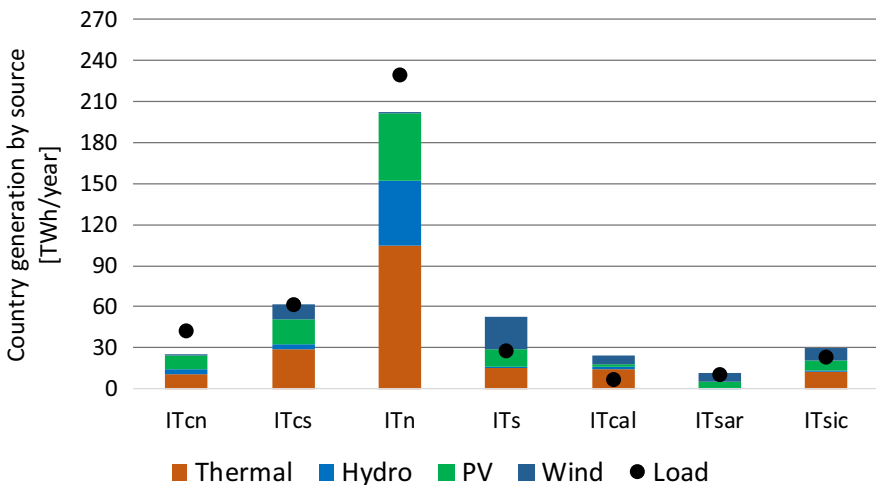


Fig. 6 Expected energy generation by source and domestic demand for each Italian market zone in year 2040

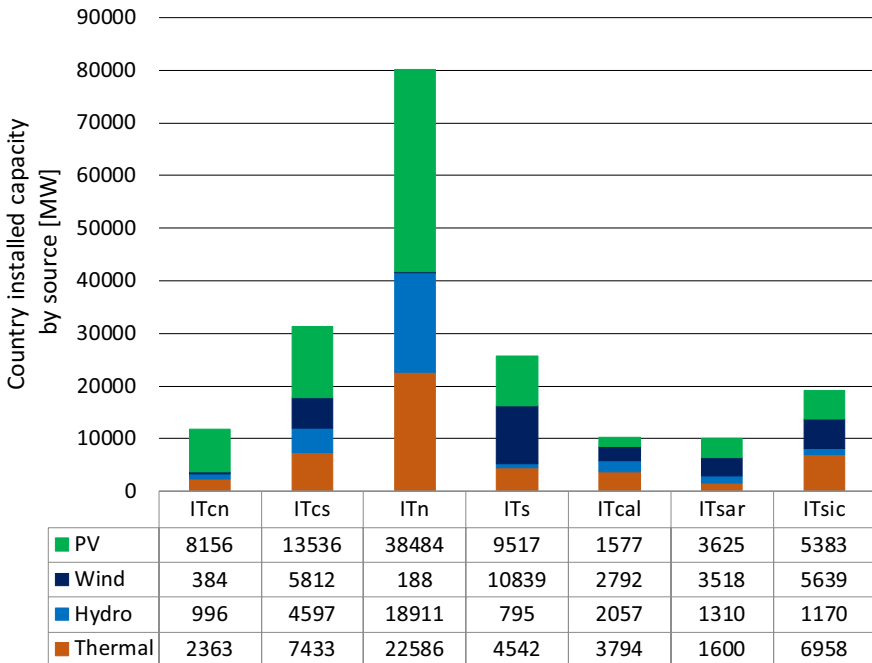


Fig. 7 Installed capacity by source for each market zone in year 2040

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.

Funding Open access funding provided by Università degli Studi di Bergamo within the CRUI-CARE Agreement.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Sauma, E., Oren, S.: Economic criteria for planning transmission investment in restructured electricity market. *IEEE Trans. Power Syst.* **22**(4), 1394–1405 (2007)
2. Pozo, D., Contreras, J., Sauma, E.: If you build it, he will come: anticipative power transmission planning. *Energy Econ.* **36**, 135–146 (2013)
3. Krishnan, V., Ho, J., Hobbs, F., Liu, A.L., McCalley, J.D., Shahidehpour, M., Zheng, Q.P.: Co-optimization of electricity transmission and generation resources for planning and policy analysis: review of concepts and modeling approaches. *Energy Syst.* **7**(2), 297–332 (2016)
4. Moeini-Aghtaie, M., Abbaspour, A., Fotuhi-Firuzabad, M.: Incorporating large-scale distant wind farms in probabilistic transmission expansion planning—part I: theory and algorithm. *IEEE Trans. Power Syst.* **27**(3), 1585–1593 (2012)
5. Yu, H., Chung, C.Y., Wong, K.P., Zhang, J.H.: A chance constrained transmission network expansion planning method with consideration of load and wind farm uncertainties. *IEEE Trans. Power Syst.* **24**(3), 1568–1576 (2009)
6. Orfanos, G.A., Georgilakis, P.S., Hatzigiorgiou, N.D.: Transmission expansion planning of systems with increasing wind power integration. *IEEE Trans. Power Syst.* **28**(2), 1355–1362 (2012)
7. Ruiz, C., Conejo, A.J.: Robust transmission expansion planning. *Eur. J. Oper. Res.* **242**(2), 390–401 (2015)
8. Li, J., Li, Z., Liu, F., Ye, H., Zhang, X., Mei, S., Chang, N.: Robust coordinated transmission and generation expansion planning considering ramping requirements and construction periods. *IEEE Trans. Power Syst.* **33**(1), 268–280 (2018)
9. Dehghan, S., Amjadi, N., Ahad, K.: Two-stage robust generation expansion planning: a mixed integer linear programming model. *IEEE Trans. Power Syst.* **29**(2), 584–597 (2014)
10. Chen, B., Wang, J., Wang, L., He, Y., Wang, Z.: Robust optimization for transmission expansion planning: minimax cost vs. minimax regret. *IEEE Trans. Power Syst.* **29**(6), 3069–3077 (2014)
11. Mejía-Giraldo, D., McCalley, J.D.: Maximizing future flexibility in electric generation portfolios. *IEEE Trans. Power Syst.* **29**, 279–288 (2014)
12. Maloney, P., Olatujoye, O., Ardakani, A.J., Mejía-Giraldo, D., McCalley, J.: 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)
13. Birge, J., Louveaux, F.: Introduction to stochastic programming (1997)
14. Alvarez Lopez, J., Ponnambalam, K., Quintana, V.H.: Generation and transmission expansion under risk using stochastic programming. *IEEE Trans. Power Syst.* **22**(3), 1369–1378 (2007)
15. Van der Weijde, A.H., Hobbs, B.F.: The economics of planning electricity transmission to accommodate renewables: Using two-stage optimisation to evaluate flexibility and the cost of disregarding uncertainty. *Energy Econ.* **34**(6), 2089–2101 (2012)
16. Jin, S., Ryan, S.M., Watson, J.-P., Woodruff, D.L.: Modeling and solving a large-scale generation expansion planning problem under uncertainty. *Energy Syst.* **2**, 209–242 (2011)
17. Jin, S., Botterud, A., Ryan, S.M.: Temporal versus stochastic granularity in thermal generation capacity planning with wind power. *IEEE Trans. Power Syst.* **29**(5), 2033–2041 (2014)
18. Conejo, A.J., Baringo, L., Kazempour, S.J., Siddiqui, A.S.: Investment in electricity generation and transmission: decision making under uncertainty. Springer, Berlin (2016)
19. Liu, Y., Sioshansi, R., Conejo, A.J.: Multistage stochastic investment planning with multiscale representation of uncertainties and decisions. *IEEE Trans. Power Syst.* **33**(1), 781–791 (2018)
20. Baringo, L., Conejo, A.J.: Risk-constrained multi-stage wind power investment. *IEEE Trans. Power Syst.* **28**(1), 401–411 (2013)
21. Fürsch, M., Nagl, S., Lindenberg, D.: Optimization of power plant investments under uncertain renewable energy deployment paths: a multistage stochastic programming approach. *Energy Syst.* **5**, 85–121 (2014)
22. Poncelet, K., Delarue, E., Six, D., Duerinck, J., D’haeseleer, W.: Impact of the level of temporal and operational detail in energy-system planning models. *Appl. Energy* **162**, 631–643 (2016)
23. Schwele, A., Kazempour, J., Pinson, P.: Do unit commitment constraints affect generation expansion? A scalable stochastic model. *Energy Syst.* (2019)

24. Belderbos, A., Delarue, E.: Accounting for flexibility in power system planning with renewables. *Int. J. Electr. Power Energy Syst.* **71**, 33–41 (2015)
25. Frupp, M.: Switch: a planning tool for power systems with large shares of intermittent renewable energy. *Environ. Sci. Technol.* **46**(11), 6371–6378 (2012)
26. Hart, E.K., Jacobson, M.Z.: A Monte Carlo approach to generator portfolio planning and carbon emissions. *Renew. Energy* **36**(8), 2278–2286 (2011)
27. ElNozahy, M.S., Salama, M.M.A., Seethapathy, R.: A probabilistic load modelling approach using clustering algorithms. In: 2013 IEEE Power & Energy Society General Meeting (2013)
28. Nick, M., Cherkaoui, R., Paolone, M.: Optimal allocation of dispersed energy storage systems in active distribution networks for energy balance and grid support. *IEEE Trans. Power Syst.* **29**(5), 2300–2310 (2014)
29. Nahmmacher, P., Schmid, E., Hirth, L., Knopf, B.: Carpe diem: a novel approach to select representative days for long-term power system models with high shares of renewable energy sources. *Energy* **112**, 430–442 (2016)
30. Fazlollahi, S., Bungener, S.L., Mandel, P., Becker, G., Maréchal, F.: Multi-objectives, multi-period optimization of district energy systems: I. Selection of typical operating periods. *Comput. Chem. Eng.* **65**, 54–66 (2014)
31. Poncelet, K., Höschle, H., Delarue, E., Virag, A., Dhaeseleer, W.: Selecting representative days for capturing the implications of integrating intermittent renewables in generation expansion planning problems. *IEEE Trans. Power Syst.* **32**(3), 1936–1948 (2017)
32. Micheli, G., Vespucci, M.T., Stabile, M., 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 (2020)
33. Tejada-Arango, D.A., Domeshek, M., Wogrin, S., Centeno, E.: Enhanced representative days and system states modeling for energy storage investment analysis. *IEEE Trans. Power Syst.* **33**(6), 6534–6544 (2018)
34. Benders, J.F.: Partitioning procedures for solving mixed-variables programming problems. *Numer. Math.* **4**(1), 238–252 (1962)
35. Conejo, A.J., Castillo, E., García-Bertrand, R., Mínguez, R.: *Decomposition techniques in mathematical programming: engineering and science applications*. Springer, Berlin (2006)
36. Zverovich, V., Fábíán, C.I., Ellison, E.F.D., Mitra, G.: A computational study of a solver system for processing two-stage stochastic LPs with enhanced Benders decomposition. *Math. Programm. Comput.* **4**, 211–238 (2012)
37. Murphy, J.: *Benders, nested Benders and stochastic programming: an intuitive introduction* (2013)
38. Birge, J.R., Louveaux, F.V.: A multicut algorithm for two-stage stochastic linear programs. *Eur. J. Oper. Res.* **34**(3), 384–392 (1988)
39. Skar, C., Doorman, G., Tomasgard, A.: Large-scale power system planning using enhanced Benders decomposition. In: 2014 Power Systems Computation Conference (PSCC), pp. 1–7 (2014)
40. Nasri, A., Kazempour, S.J., Conejo, A.J.: Network-constrained AC unit commitment under uncertainty: a benders’ decomposition approach. *IEEE Trans. Power Syst.* **31**(1), 412–422 (2016)
41. Lumbreras, S., Ramos, A.: Transmission expansion planning using an efficient version of Benders’ decomposition. A case study. In: *IEEE Grenoble Conference*, Grenoble (2013)
42. Lumbreras, S., Ramos, A.: How to solve the transmission expansion planning problem faster: acceleration techniques applied to Benders’ decomposition. *IET Gener. Transm. Distrib.* **10**(10), 2351–2359 (2016)
43. Trukhanov, S., Ntaimo, L., Schaefer, A.: Adaptive multicut aggregation for two-stage stochastic linear programs with recourse. *Eur. J. Oper. Res.* **206**(2), 395–406 (2010)
44. You, F., Grossmann, I.E.: Multicut Benders decomposition algorithm for process supply chain planning under uncertainty. *Ann. Oper. Res.* **210**(1), 191–211 (2013)
45. Morales-España, G., Latorre, J.M., Ramos, A.: Tight and compact MILP formulation for the thermal unit commitment problem. *IEEE Trans. Power Syst.* **28**(4), 4897–4908 (2013)
46. Gentile, C., Morales-España, G., Ramos, A.: A tight MIP formulation of the unit commitment problem with start-up and shut-down constraints. *EURO J. Comput. Optimiz.* **5**, 177–201 (2017)
47. ENTSO-E; ENTSOG: *TYNDP 2020 Scenario Report* (2019)
48. Terna S.p.A: *Gruppo Terna: Piano di Sviluppo* (2019)
49. Cole, W., Frazier, A.W.: *Cost projections for utility-scale battery storage*. National Renewable Energy Laboratory, Golden (2019)

50. International Renewable Energy Agency (IRENA): Electricity storage and renewables: costs and markets to 2030 (2017)
51. European Energy Research Alliance (EERA): Fact Sheet 1—pumped hydro energy storage. Brussels (2016)
52. Bussieck, M.R., Ferris, M.C., Lohmann, T.: GUSS: solving collections of data related models within GAMS. In: Algebraic modeling systems, pp. 35–56. Springer, Berlin (2012)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Giovanni Micheli¹  · **Maria Teresa Vespucci**¹ · **Marco Stabile**² · **Cinzia Puglisi**² · **Andres Ramos**³

✉ Giovanni Micheli
giovanni.micheli@unibg.it

Maria Teresa Vespucci
maria-teresa.vespucci@unibg.it

Marco Stabile
marco.stabile@cesi.it

Cinzia Puglisi
cinzia.puglisi@cesi.it

Andres Ramos
andres.ramos@comillas.edu

¹ Department of Management, Information and Production Engineering, University of Bergamo, Bergamo, Italy

² CESI, Milano, Italy

³ Instituto de Investigación Tecnológica, ICAI, Universidad Pontificia Comillas, Madrid, Spain