Electric Power Infrastructure Planning: Mixed-Integer Programming Model and Nested Decomposition Algorithm

Lara, C. L.[a], Mallapragada, D.[b], Papageorgiou, D.[b], Venkatesh, A.[b], Grossmann, I. E.[a]

[a] *Department of Chemical Engineering, Carnegie Mellon University*
5000 Forbes Avenue, Pittsburgh, PA 15213
[b] *Corporate Strategic Research, ExxonMobil Research and Engineering Company*
1545 Route 22 East, Annandale, NJ 08801, USA

Abstract

This paper addresses the long-term planning of electric power infrastructures considering high renewable penetration. To capture the intermittency of these sources, we propose a deterministic multi-scale Mixed-Integer Linear Programming (MILP) formulation that simultaneously considers annual generation investment decisions and hourly operational decisions. We adopt judicious approximations and aggregations to improve its tractability. Moreover, to overcome the computational challenges of treating hourly operational decisions within a monolithic multi-year planning horizon, we propose a decomposition algorithm based on Nested Benders Decomposition for multi-period MILP problems to allow the solution of larger instances. Our decomposition extends previous nested Benders methods by handling integer and continuous state variables. We apply the proposed modeling framework to a case study in the ERCOT region, and demonstrate massive computational savings from our decomposition.

Keywords: strategic planning, OR in energy, large-scale optimization

1. Introduction

Energy systems planning models allow the evaluation of alternate scenarios for future growth, providing information to support the decision-making process and the selection of technologies in the power sector [57, 52, 14]. Generation and transmission expansion models can vary widely in scope (local versus regional) as well as in resolution of time and space. These models can be used to study the impact of new technology developments, resource cost trends, and policy shifts on the projected generation mix in order to meet future demand.

Although transmission expansion is not considered in this work, it is important to be aware of its impact on long-term planning decisions, and thus we discuss it here. Traditionally, generation and transmission expansion are modeled separately: the generation is planned first and the transmission network is designed

*Corresponding author
Email address: grossmann@cmu.edu (Grossmann, I. E.)

to meet this supply [60, 5, 29, 4, 6]. Their simultaneous optimization is, however, a better way of capturing the trade-off between investing in local generation or transmission from remote supplies [27].

There is growing interest to use planning models to study scenarios with increasing penetration of solar and wind generation [31]. Historically, since power systems were dominated by *dispatchable* thermal resources, planning models could ignore short-term operating constraints and have longer time periods without impacting much the quality of the results. However, in a system deriving a large proportion of generation from intermittent resources, it is critical to include/consider hourly or subhourly operational decisions to assess the flexibility of the system [35, 3, 28]. Only then it is possible to systematically/rigorously assess the trade-off between long-term investment decisions and short-term operating decisions. Accordingly, several papers have studied the impact of including operating constraints such as unit commitment, ramping limits and operating reserves in long term planning models [53, 12, 39, 45, 46, 26, 19].

Other directions recently taken by researchers in the area of generation and transmission expansion are to represent competitive market behavior between generation entities/participants using game theory and multi-level optimization [50, 48, 9], and to model the uncertainties [8, 41, 11, 30]. O'Neill et al. [38] present a comprehensive formulation including detailed generator, topology, transmission and operational aspects from production cost planning models into a long-term stochastic two-stage mixed-integer planning framework. The complexity in the formulation makes the model computationally intractable and poses new challenges to the power systems community.

In this paper, we propose an optimization modeling framework to evaluate the changes in the power systems infrastructure required to meet the projected electricity demand over the next few decades, while taking into account detailed operating constraints, and the variability and intermittency of renewable generation sources. The modeling framework, which is based on mixed-integer linear programming (MILP), takes the viewpoint of a central planning entity whose goal is to identify the source (nuclear , coal, natural gas, wind and solar), generation technology (e.g., steam, combustion and wind turbines, photo-voltaic and concentrated solar panels), location (regions), and capacity of future power generation technologies that can meet the projected electricity demand, while minimizing the amortized capital investment of all new generating units, the operating costs of both new and existing units, and corresponding environmental costs (e.g. carbon tax and renewable generation quota).

The major challenge lies in the multi-scale integration of detailed operation decisions at the hourly (or subhourly) level with investment planning decisions over a few decades. In order to improve its computational tractability, judicious modeling approximations and aggregations are considered. Although there is novelty in the formulation of the problem, the main contribution of this work is the solution strategy. We propose a decomposition algorithm based on Nested Benders Decomposition for mixed-integer multi-period problems to solve large-scale models. This framework was originally developed for stochastic programming [61], but

we have adapted it to deterministic multi-period problems. We have extended it to handle integer and continuous state variables, and have applied acceleration techniques to improve the overall performance of the algorithm.

In Section 2, a formal problem statement is given, and in Section 3 we describe the modeling strategies adopted to handle the spatial and temporal multi-scale aspect of the problem. The MILP formulation is presented in Section 4. Section 5 describes the proposed decomposition algorithm. Section 6 shows the results for a real-world case study for the Electric Reliability Council of Texas (ERCOT) region, and a comparison between the performance of the full size MILP formulation and the proposed algorithm.

2. Problem statement

The proposed planning problem involves choosing the optimal investment strategy and operating schedule for the power system in order to meet the projected load demand over the time-horizon for each location.

A set of existing and potential generators is given, and for which the energy source (nuclear, coal, natural gas, wind or solar)[1] and the generation technology are known.

- For the existing generators we consider: (a) **coal:** steam turbine (coal-st-old); (b) **natural gas:** boiler plants with steam turbine (ng-st-old), combustion turbine (ng-ct-old), and combined-cycle (ng-cc-old); (c) **nuclear:** steam turbine (nuc-st-old); (d) **solar:** photo-voltaic (pv-old); (e) **wind:** wind turbine (wind-old);
- For the potential generators we consider: (a) **coal:** without (coal-new) and with carbon capture (coal-ccs-new); (b) **natural gas:** combustion turbine (ng-ct-new), combined-cycle without (ng-cc-new) and with carbon capture (ng-cc-ccs-new); (c) **nuclear:** steam turbine (nuc-st-new); (d) **solar:** photo-voltaic (pv-new) and concentrated solar panel (csp-new); (e) **wind:** wind turbine (wind-new);

Also known are: their nameplate (maximum) capacity; expected lifetime; fixed and variable operating costs; start-up cost (fixed and variable); cost for extending their lifetimes; CO_2 emission factor and carbon tax, if applicable; fuel price, if applicable; and operating characteristics such as ramp-up/ramp-down rates, operating limits, contribution to spinning and quick start fraction for thermal generators, and capacity factor for renewable generators.

For the case of existing generators, their age at the beginning of the time-horizon and location are also known. For the case of potential generators, the capital cost and the maximum yearly installation of each generation technology are also given. Additionally, the projected load demand is given for each location, as well as the distance between locations, the transmission loss per mile, and the transmission line capacity between locations.

[1] In this paper we do not consider hydroelectric power as it is available in very limited amounts in the ERCOT region.

The problem is then to determine: a) when, where, which type, and in how many new generators to invest; b) whether or not to extend the life of the generators that reached their expected lifetime; and c) an approximate operating schedule for each installed generator, and the approximate power flow between each location in order to meet the projected load demand while minimizing the overall operating, investment, and environmental costs.

3. Modeling strategies and assumptions

The mix of combinatorial and operational elements of the problem described in Section 2 means that, depending on the time horizon and area considered, the corresponding optimization problem may be too large and intractable for current commercial general purpose MILP solvers. Therefore, in order to solve the resulting deterministic MILP to provable optimality for large areas and over a few decades, it is key to explore judicious modeling aggregations and approximations to address the multi-scale aspects, both in its spatial and temporal dimensions. In order to significantly reduce computation time, generator clustering [40] and time sampling [44] approaches are adopted.

3.1. Spatial representation

In order to allow the solution of large-scale instances, the area of scope is divided into regions that have similar climate (e.g., wind speed and solar incidence over time), and load demand profiles. It is assumed that the potential locations for the generators are the midpoints of each region r. Additionally, based on the work of Palmintier and Webster [40], generators that have the same characteristics, such as generation technology and operating status (i.e., existing or potential), are aggregated into clusters i for each region r. The spatial configuration of the problem is shown in Figure 1 for the ERCOT region.

The major impact of this approximation in the model formulation is that the discrete variables associated with generators correspond to integer rather than binary variables to represent the number of generators under a specific status in cluster i.

3.2. Temporal representation

It is crucial to include hourly level information to evaluate scenarios with increasing renewable energy generation [45], because of the variability in that resource, as well as the changes in load. On the other hand, strategic capacity and transmission expansion decisions must be optimized over a long-term horizon (e.g., a few decades). Therefore, the investment decisions are made on a yearly basis, while operating decisions are made at the hourly level. To tackle the problem's multi-scale nature and reduce computation time, each year is modeled using d representative days with hourly resolution resulting in 24 subperiods. We employ a k-means clustering approach to select these days from historical data (see Section 6), where the goal of the clustering procedure is to select representative days to approximate: (i) the "duration curves"

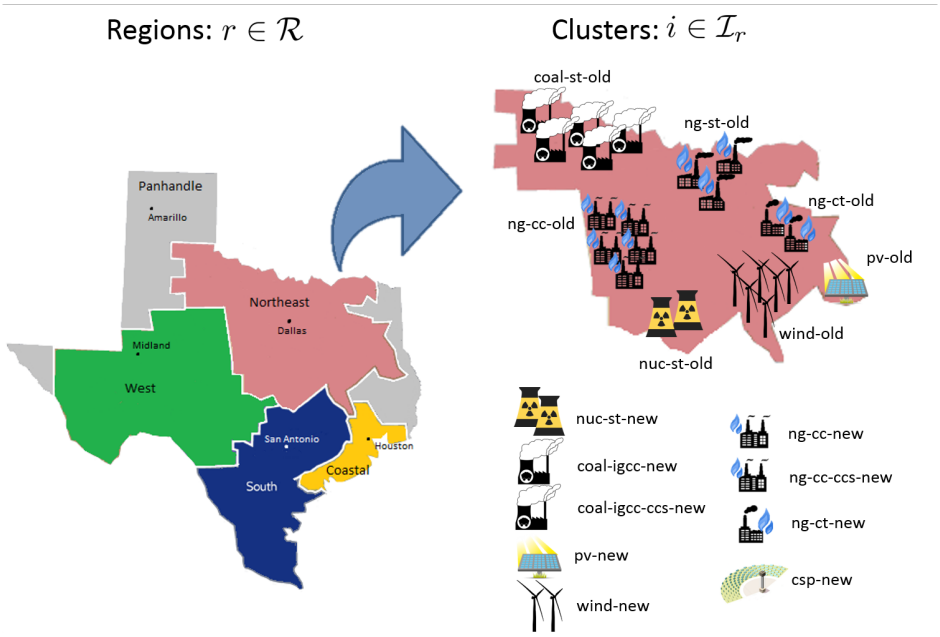

Figure 1: Model representation of regions and clusters (regional map modified from [16])

of historical load and renewables time series, (ii) the temporal correlation of each time series, and (iii) the hourly correlation between each time series. The temporal configuration of the problem is shown in Figure 2.

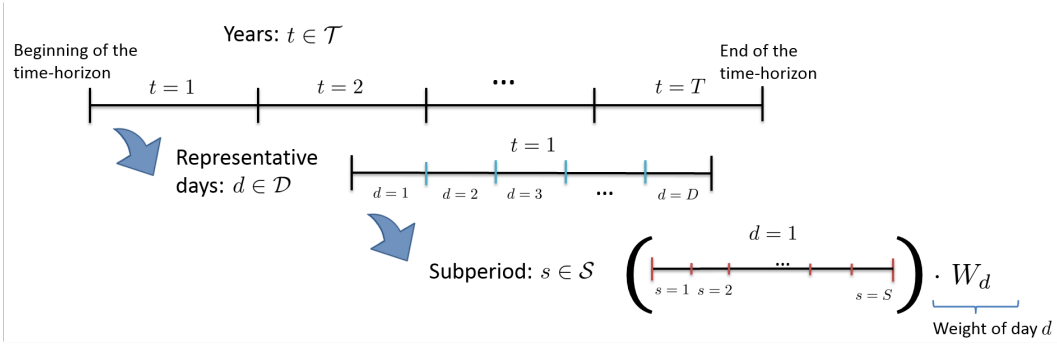

Figure 2: Multi-scale representation

3.3. Transmission representation

Transmission is also an important aspect of a power systems infrastructure, influencing where to build power plants, which ones to operate, and how much power to be generated by each of them. The rigorous way of representing transmission between generation and load nodes in the system is through optimal power flow models [22, 21]. As explained in Section 3.1, the proposed model uses a reduced network, which for the example in Figure 1 has only 5 nodes representing the 5 regions. Additionally, in order to further

simplify the transmission model, the "truck-route" representation is adopted as described in [52] and [27]. The transmission network is represented similarly to pipelines, assuming that the flow in each line can be determined by an energy balance between nodes. This approximation ignores Kirchhoff's voltage law, which dictates that the power will flow along the path of least impedance. It is also assumed that the transmission lines have a maximum capacity, and that transmission expansion is not considered. Additionally, the transmission losses are characterized by a fraction loss per mile, and are not endogeneously calculated.

4. MILP Formulation

This section presents a deterministic MILP formulation organized into 3 groups of constraints: operational, investment-related, and generator balances. Note that if an index appears in a summation or next to a $\forall$ symbol without a corresponding set, all elements in that set are assumed.

4.1. Operational constraints

The energy balance (1) ensures that the sum of instantaneous power $p_{i,r,t,d,s}$ generated by generator cluster i in region r plus the difference between the power flow going from regions r' to region r, $p^{\text{flow}}_{r',r,t,d,s}$, and the power flowing from region r to regions r', $p^{\text{flow}}_{r,r',t,d,s}$, equals the load demand $L_{r,t,d,s}$ at that region r plus a slack for curtailment of renewable generation $cu_{r,t,d,s}$, at all times.

$$\sum_{i}(p_{i,r,t,d,s}) + \sum_{r' \neq r}\left(p^{\text{flow}}_{r',r,t,d,s} \cdot (1 - T^{\text{loss}}_{r,r'} \cdot D_{r,r'}) - p^{flow}_{r,r',t,d,s}\right) = L_{r,t,d,s} + cu_{r,t,d,s} \quad \forall\ r,t,d,s \tag{1}$$

The distance between regions $D_{r,r'}$ assumes the midpoint for each region, and the transmission loss $T^{\text{loss}}_{r,r'}$ is approximated by a fraction loss per mile.

The capacity factor constraint (2) limits the power outlet $p_{i,r,t,d,s}$ of renewable generators to be equal to a fraction $Cf_{i,r,t,d,s}$ of the nameplate capacity $Qg^{\text{np}}_{i,r}$ in each sub-period s. $ngo^{\text{rn}}_{i,r,t}$ represents the number of renewable generators that are operational in year t. Due to the flexibility in sizes for renewable generators, $ngo^{\text{rn}}_{i,r,t}$ is relaxed to be continuous.

$$p_{i,r,t,d,s} = Qg^{\text{np}}_{i,r} \cdot Cf_{i,r,t,d,s} \cdot ngo^{\text{rn}}_{i,r,t} \qquad \forall\ i \in \mathcal{I}^{\text{RN}}_{r}, r,t,d,s \tag{2}$$

The unit commitment constraint (3) computes the number of generators that are ON, $u_{i,r,t,d,s}$, or in startup, $su_{i,r,t,d,s}$, and shutdown, $sd_{i,r,t,d,s}$, modes in cluster i in sub-period s of representative day d of year t, and treated as integer variables.

$$u_{i,r,t,d,s} = u_{i,r,t,d,s-1} + su_{i,r,t,d,s} - sd_{i,r,t,d,s} \qquad \forall\ i \in \mathcal{I}^{\text{TH}}_{r}, r,t,d,s \tag{3}$$

The ramping limit constraints (4)-(5) capture the limitation on how fast thermal units can adjust their output power, $p_{i,r,t,d,s}$, where $Ru_i^{\max}$ is the maximum ramp-up rate, $Rd_i^{\max}$ is the maximum ramp-down rate, and $Pg_i^{\min}$ is the minimum operating limit [40].

$$\begin{aligned} p_{i,r,t,d,s} - p_{i,r,t,d,s-1} \leq \quad & Ru_i^{\max} \cdot Hs \cdot Qg_{i,r}^{\text{np}} \cdot (u_{i,r,t,d,s} - su_{i,r,t,d,s}) \\ & + \max\left(Pg_i^{\min}, Ru_i^{\max} \cdot Hs\right) \cdot Qg_{i,r}^{\text{np}} \cdot su_{i,r,t,d,s} \qquad \forall\ i \in \mathcal{I}_r^{\text{TH}}, r, t, d, s \end{aligned} \tag{4}$$

$$\begin{aligned} p_{i,r,t,d,s-1} - p_{i,r,t,d,s} \leq \quad & Rd_i^{\max} \cdot Hs \cdot Qg_{i,r}^{\text{np}} \cdot (u_{i,r,t,d,s} - su_{i,r,t,d,s}) \\ & + \max\left(Pg_i^{\min}, Rd_i^{\max} \cdot Hs\right) \cdot Qg_{i,r}^{\text{np}} \cdot sd_{i,r,t,d,s} \qquad \forall\ i \in \mathcal{I}_r^{\text{TH}}, r, t, d, s \end{aligned} \tag{5}$$

Note that the first terms on the right hand side of (4) and (5) apply only for normal operating mode (i.e., generator is ON), while the second terms apply for the startup and shutdown modes.

The operating limits constraints (6)-(7) specify that each thermal generator is either OFF and outputting zero power, or ON and running within the operating limits $Pg_i^{\min} \cdot Qg_{i,r}^{\text{np}}$ and $Qg_{i,r}^{\text{np}}$. The variable $u_{i,r,t,d,s}$ (integer variable) represents the number of generators that are ON in cluster $i \in \mathcal{I}_r^{\text{TH}}$ at the time period t, representative day d, and sub-period s.

$$u_{i,r,t,d,s} \cdot Pg_i^{\min} \cdot Qg_{i,r}^{\text{np}} \leq p_{i,r,t,d,s} \qquad \forall\ i \in \mathcal{I}_r^{\text{TH}}, r, t, d, s \tag{6}$$

$$p_{i,r,t,d,s} + q_{i,r,t,d,s}^{\text{spin}} \leq u_{i,r,t,d,s} \cdot Qg_{i,r}^{\text{np}} \qquad \forall\ i \in \mathcal{I}_r^{\text{TH}}, r, t, d, s \tag{7}$$

The upper limit constraint is modified in order to capture the need for generators to run below the maximum considering operating reserves, where $q_{i,r,t,d,s}^{\text{spin}}$ is a variable representing the spinning reserve capacity.

The total operating reserve constraint (8) dictates that the total spinning reserve, $q_{i,r,t,d,s}^{\text{spin}}$, plus quick-start reserve, $q_{i,r,t,d,s}^{\text{Qstart}}$, must exceed the minimum operating reserve, $Op^{\min}$, which is a percentage of the load $L_{r,t,d,s}$ in a reserve sharing region r at each sub-period s.

$$\sum_{i \in \mathcal{I}_r^{\text{TH}}} \left(q_{i,r,t,d,s}^{\text{spin}} + q_{i,r,t,d,s}^{\text{Qstart}}\right) \geq Op^{\min} \cdot L_{r,t,d,s} \qquad \forall\ i \in \mathcal{I}_r^{\text{TH}}, r, t, d, s \tag{8}$$

Spinning Reserve is the on-line reserve capacity that is synchronized to the grid system and ready to meet electric demand within 10 minutes of a dispatch instruction by the independent system operator (ISO). Quick-start (or non-spinning) reserve is the extra generation capacity that is not currently connected to the system but can be brought on-line after a short delay.

The total spinning reserve constraint (9) specifies that the total spinning reserve $q_{i,r,t,d,s}^{\text{spin}}$ must exceed the minimum spinning reserve, $Spin^{\min}$, which is a percentage of the load $L_{r,t,d,s}$ in a reserve sharing region r

at each sub-period s.

$$\sum_{i \in \mathcal{I}_r^{\mathrm{TH}}} q_{i,r,t,d,s}^{\mathrm{spin}} \geq Spin^{\min} \cdot L_{r,t,d,s} \qquad \forall\, i \in \mathcal{I}_r^{\mathrm{TH}}, r, t, d, s \tag{9}$$

The maximum spinning reserve constraint (10) states that the maximum fraction of capacity of each generator cluster that can contribute to spinning reserves is given by $Frac_i^{\mathrm{spin}}$, which is a fraction of the nameplate capacity $Qg_{i,r}^{\mathrm{np}}$.

$$q_{i,r,t,d,s}^{\mathrm{spin}} \leq u_{i,r,t,d,s} \cdot Qg_{i,r}^{\mathrm{np}} \cdot Frac_i^{\mathrm{spin}} \qquad \forall\, i \in \mathcal{I}_r^{\mathrm{TH}}, r, t, d, s \tag{10}$$

The maximum quick-start reserve constraint dictates that the maximum fraction of the capacity of each generator cluster that can contribute to quick-start reserves is given by $Frac_i^{\mathrm{Qstart}}$ (fraction of the nameplate capacity $Qg_{i,r}^{\mathrm{np}}$), and that quick-start reserves can only be provided by the generators that are OFF, i.e., not active.

$$q_{i,r,t,d,s}^{\mathrm{Qstart}} \leq (ngo_{i,r,t}^{\mathrm{th}} - u_{i,r,t,d,s}) \cdot Qg_{i,r}^{\mathrm{np}} \cdot Frac_i^{\mathrm{Qstart}} \qquad \forall\, i \in \mathcal{I}_r^{\mathrm{TH}}, r, t, d, s \tag{11}$$

Here the integer variable $ngo_{i,r,t}^{\mathrm{th}}$ represents the number of thermal generators that are operational (i.e., installed and ready to operate) at year t.

4.2. Investment-related constraints

The planning reserve requirement (12) ensures that the operating capacity is greater than or equal to the annual peak load $L_t^{\max}$, plus a predefined fraction of reserve margin $R_t^{\min}$ of the annual peak load $L_t^{\max}$.

$$\sum_{i \in \mathcal{I}_r^{\mathrm{RN}}} \sum_{r} \left(Qg_{i,r}^{\mathrm{np}} \cdot Q_i^{\mathrm{v}} \cdot ngo_{i,r,t}^{\mathrm{rn}}\right) + \sum_{i \in \mathcal{I}_r^{\mathrm{TH}}} \sum_{r} \left(Qg_{i,r}^{\mathrm{np}} \cdot ngo_{i,r,t}^{\mathrm{th}}\right) \geq (1 + R_t^{\min}) \cdot L_t^{\max} \qquad \forall\, t \tag{12}$$

For all thermal generators, their full nameplate capacity $Qg_{i,r}^{\mathrm{np}}$ counts towards the planning reserve requirement. However, for the renewable technologies (wind, PV and CSP), their contribution is less than the nameplate due to the inability to control dispatch and the uncertainty of the output [52]. Therefore, the fraction of the capacity that can be reliably counted towards the planning reserve requirement is referred to as the capacity value Q_i^{v}.

The minimum annual renewable generation requirement (13) ensures that, in case of policy mandates, the renewable generation quota target, $RN_t^{\min}$, which is a fraction of the energy demand ED_t, is satisfied. If not, i.e, if there is a deficit def_t^{rn} from the quota, this is subjected to a penalty that is included later in

the objective function.

$$\sum_{d}\sum_{s}\left[W_d \cdot Hs \cdot \left(\sum_{i\in\mathcal{I}_r^{\mathrm{RN}}}\sum_{r} p_{i,r,t,d,s} - cu_{r,t,d,s}\right)\right] + def_t^{\mathrm{rn}} \geq RN_t^{\min}\cdot ED_t \qquad \forall\ t \tag{13}$$

Here W_d represents the weight of the representative day d, Hs is the length of the sub-period, $cu_{r,t,d,s}$ is the curtailment of renewable generation, and ED_t represent the energy demand in year t:

$$ED_t = \sum_{r}\sum_{d}\sum_{s}(W_d \cdot Hs \cdot L_{r,t,d,s})$$

The maximum yearly installation constraints (14)-(15) limit the yearly installation per generation type in each region r to an upper bound $Q_{i,t}^{\mathrm{inst,UB}}$ in MW/year. Here $ngb_{i,r,t}^{\mathrm{rn}}$ and $ngb_{i,r,t}^{\mathrm{th}}$ represent the number of renewable and thermal generators built in region r in year t, respectively. Note that due to the flexibility in sizes for renewable generators, $ngb_{i,r,t}^{\mathrm{rn}}$ is relaxed to be continuous.

$$\sum_{r} ngb_{i,r,t}^{\mathrm{rn}} \leq Q_{i,t}^{\mathrm{inst,UB}}/Qg_{i,r}^{\mathrm{np}} \qquad \forall\ i\in\mathcal{I}_r^{\mathrm{Rnew}}, t \tag{14}$$

$$\sum_{r} ngb_{i,r,t}^{\mathrm{th}} \leq Q_{i,t}^{\mathrm{inst,UB}}/Qg_{i,r}^{\mathrm{np}} \qquad \forall\ i\in\mathcal{I}_r^{\mathrm{Tnew}}, t \tag{15}$$

4.3. Generator balance constraints

Concerning renewable generator clusters, we define a set of constraints (16)-(17) to compute the number of generators in cluster i that are ready to operate $ngo_{i,r,t}^{\mathrm{rn}}$, taking into account the generators that were already existing at the beginning of the planning horizon $Ng_{i,r}^{\mathrm{Rold}}$, the generators built $ngb_{i,r,t}^{\mathrm{rn}}$, and the generators retired $ngr_{i,r,t}^{\mathrm{rn}}$ at year t. It is important to highlight that we assume *no lead time* between the decision to build/install a generator and the moment it can begin producing electricity.

$$ngo_{i,r,t}^{\mathrm{rn}} = Ng_{i,r}^{\mathrm{Rold}} + ngb_{i,r,t}^{\mathrm{rn}} - ngr_{i,r,t}^{\mathrm{rn}} \qquad \forall\ i\in\mathcal{I}_r^{\mathrm{RN}}, t=1 \tag{16}$$

$$ngo_{i,r,t}^{\mathrm{rn}} = ngo_{i,r,t-1}^{\mathrm{rn}} + ngb_{i,r,t}^{\mathrm{rn}} - ngr_{i,r,t}^{\mathrm{rn}} \qquad \forall\ i\in\mathcal{I}_r^{\mathrm{RN}}, t>1 \tag{17}$$

As aforementioned, due to the flexibility in sizes for renewable generators, $ngo_{i,r,t}^{\mathrm{rn}}$, $ngb_{i,r,t}^{\mathrm{rn}}$, and $ngr_{i,r,t}^{\mathrm{rn}}$ are relaxed to be continuous. Note that $ngb_{i,r,t}^{\mathrm{rn}}$ for $i\in\mathcal{I}_r^{\mathrm{Rold}}$ is fixed to zero in all time periods, i.e., the clusters of existing renewable generators cannot have any new additions during the time horizon considered.

We also define a set of constraints (18)-(19) to enforce the generators that reached the end of their lifetime to either retire, $ngr_{i,r,t}^{\mathrm{rn}}$, or have their life extended, $nge_{i,r,t}^{\mathrm{rn}}$. $Ng_{i,r,t}^{\mathrm{r}}$ is a parameter that represents

the number of old generators (i.e., $i \in \mathcal{I}_r^{\text{old}}$) that reached the end of their lifetime, LT_i, at year t.

$$Ng^{\text{r}}_{i,r,t} = ngr^{\text{rn}}_{i,r,t} + nge^{\text{rn}}_{i,r,t} \qquad \forall\, i \in \mathcal{I}_r^{\text{Rold}}, r, t \tag{18}$$

$$\sum_{t'' \leq t - LT_i} ngb^{\text{rn}}_{i,r,t''} = \sum_{t' \leq t} \left(ngr^{\text{rn}}_{i,r,t'} + nge^{\text{rn}}_{i,r,t'} \right) \qquad \forall\, i \in \mathcal{I}_r^{\text{Rnew}}, r, t \tag{19}$$

Concerning thermal generator clusters, we define a set of constraints (20)-(21) to compute the number of generators in cluster i that are ready to operate $ngo^{\text{th}}_{i,r,t}$, taking into account the generators that were already existing at the beginning of the planning horizon $Ng^{\text{Told}}_{i,r}$, the generators built $ngb^{\text{th}}_{i,r,t}$, and the generators retired $ngr^{\text{th}}_{i,r,t}$ at year t.

$$ngo^{\text{th}}_{i,r,t} = Ng^{\text{Told}}_{i,r} + ngb^{\text{th}}_{i,r,t} - ngr^{\text{th}}_{i,r,t} \qquad \forall\, i \in \mathcal{I}_r^{\text{TH}}, t = 1 \tag{20}$$

$$ngo^{\text{th}}_{i,r,t} = ngo^{\text{th}}_{i,r,t-1} + ngb^{\text{th}}_{i,r,t} - ngr^{\text{th}}_{i,r,t} \qquad \forall\, i \in \mathcal{I}_r^{\text{TH}}, t > 1 \tag{21}$$

Note that $ngb^{\text{th}}_{i,r,t}$ for $i \in \mathcal{I}_r^{\text{Told}}$ is fixed to zero in all time periods, i.e., the clusters of existing thermal generators cannot have any new additions during the time horizon considered.

We also define a set of constraints (22)-(23) to enforce the generators that reached the end of their lifetime to either retire, $ngr^{\text{th}}_{i,r,t}$, or have their life extended $nge^{\text{th}}_{i,r,t}$.

$$Ng^{\text{r}}_{i,r,t} = ngr^{\text{th}}_{i,r,t} + nge^{\text{th}}_{i,r,t} \qquad \forall\, i \in \mathcal{I}_r^{\text{Told}}, r, t \tag{22}$$

$$\sum_{t'' \leq t - LT_i} ngb^{\text{th}}_{i,r,t''} = \sum_{t' \leq t} \left(ngr^{\text{th}}_{i,r,t'} + nge^{\text{th}}_{i,r,t'} \right) \qquad \forall\, i \in \mathcal{I}_r^{\text{Tnew}}, r, t \tag{23}$$

Finally, we have constraint (24) that ensures that only installed generators can be in operation:

$$u_{i,r,t,d,s} \leq ngo^{\text{th}}_{i,r,t} \qquad \forall\, i \in \mathcal{I}_r^{\text{Tnew}}, r, t, d, s \tag{24}$$

4.4. Objective Function

The objective of this model is to minimize the net present cost, Φ, over the planning horizon, which includes operating costs Φ^{opex}, investment costs Φ^{capex}, and potential penalties Φ^{PEN} for not meeting the the targets on renewables.

$$\min \quad \Phi = \sum_{t} \left(\Phi^{\text{opex}}_t + \Phi^{\text{capex}}_t + \Phi^{\text{PEN}}_t \right) \tag{25}$$

The operating expenditure, Φ^{opex}_t, comprises the variable $VOC_{i,t}$ and fixed $FOC_{i,t}$ operating costs, as well as fuel cost P^{fuel}_i per heat rate HR_i, carbon tax $Tx^{\text{CO}_2}_t$ for CO_2 emissions $EF^{CO_2}_i$, and start-up cost

(variable cost P_i^{fuel} that depends on the amount of fuel burned for startup F_i^{start}, and fixed cost C_i^{start}).

$$\begin{aligned}\Phi_t^{\text{opex}} = \; & If_t \cdot \Bigg[\sum_d \sum_s W_d \cdot hs \cdot \Bigg(\sum_i \sum_r (VOC_{i,t} + P_i^{\text{fuel}} \cdot HR_i + Tx_t^{\text{CO}_2} \cdot EF_i^{CO_2} \cdot HR_i) \cdot p_{i,r,t,d,s} \Bigg) \\ & + \Bigg(\sum_{i \in \mathcal{I}_r^{\text{RN}}} \sum_r FOC_{i,t} \cdot Qg_{i,r}^{\text{np}} \cdot ngo_{i,r,t}^{\text{rn}} \Bigg) + \Bigg(\sum_{i \in \mathcal{I}_r^{\text{TH}}} \sum_r FOC_{i,t} \cdot Qg_{i,r}^{\text{np}} \cdot ngo_{i,r,t}^{\text{th}} \Bigg) \\ & + \sum_{i \in \mathcal{I}_r^{\text{TH}}} \sum_r \sum_d \sum_s W_d \cdot Hs \cdot su_{i,r,t,d,s} \cdot Qg_{i,r}^{\text{np}} \\ & \cdot \Big(F_i^{\text{start}} \cdot P_i^{\text{fuel}} + F_i^{\text{start}} \cdot EF^{CO_2} \cdot Tx_t^{\text{CO}_2} + C_i^{\text{start}} \Big) \Bigg] \end{aligned} \tag{26}$$

The capital expenditure, Φ_t^{capex}, includes the amortized cost of acquiring new generators, $DIC_{i,t}$, and the amortized cost of extending the life of generators that reached their expected lifetime. The latter is assumed to be a fraction LE_i of the investment cost, $DIC_{i,t}$, in a new generator with the same or equivalent generation technology. In this framework, the investment cost takes into account the remaining value at the end of the time horizon by considering the annualized capital cost and multiplying it by the number of years remaining in the planning horizon at the time of installation to calculate the $DIC_{i,t}$.

$$\begin{aligned}\Phi_t^{\text{capex}} = \; & If_t \cdot \Bigg[\sum_{i \in \mathcal{I}_r^{\text{Rnew}}} \sum_r DIC_{i,t} \cdot CC_i^{\text{m}} \cdot Qg_{i,r}^{\text{np}} \cdot ngb_{i,r,t}^{\text{rn}} + \sum_{i \in \mathcal{I}_r^{\text{Tnew}}} \sum_r DIC_{i,t} \cdot CC_i^{\text{m}} \cdot Qg_{i,r}^{\text{np}} \cdot ngb_{i,r,t}^{\text{th}} \\ & + \sum_{i \in \mathcal{I}_r^{\text{RN}}} \sum_r DIC_{i,t} \cdot LE_i \cdot Qg_{i,r}^{\text{np}} \cdot nge_{i,r,t}^{\text{rn}} + \sum_{i \in \mathcal{I}_r^{\text{TH}}} \sum_r DIC_{i,t} \cdot LE_i \cdot Qg_{i,r}^{\text{np}} \cdot nge_{i,r,t}^{\text{th}} \Bigg] \end{aligned} \tag{27}$$

The capital multiplier CC_i^{m} associated with new generator clusters is meant to account for differences in depreciation schedules applicable to each technology, with higher values being indicative of slower depreciating schedule and vice versa.

Lastly, the penalty cost, Φ_t^{PEN}, includes the potential fines for not meeting the renewable energy quota, PEN_t^{rn}, and curtailing the renewable generation.

$$\Phi_t^{\text{PEN}} = \; If_t \cdot \Bigg(PEN_t^{\text{rn}} \cdot def_t^{\text{rn}} + PEN^c \cdot \sum_r \sum_d \sum_s cu_{r,t,d,s} \Bigg) \tag{28}$$

The parameters If_t, $DIC_{i,t}$, $ACC_{i,t}$, and T_t^{remain} are defined in Appendix A.

The integrated planning and operations model for the electric power systems is then given by the multi-period MILP model defined by equations (1)-(28).

5. Nested Decomposition for Multiperiod MILP Problems

Even though the multi-period MILP formulation in 4 incorporates modeling strategies to reduce the size of the model, it can still be very expensive to solve and potentially intractable depending on the size of the area considered, and the time resolution of the representative periods per season. Therefore, we propose a decomposition algorithm based on Nested Benders Decomposition [7], Stochastic Dual Dynamic Programming (SDDP) [43], and Generalized Dual Dynamic Programming (GDDP) [58].

These methods are used in the context of Multistage Linear Stochastic Programming (MLSP), but their major limitation is that they can only be applied to convex subproblems. Thus, they are not suitable for our problem, which gives rise to mixed-integer subproblems. In this context, Cerisola et al. [10] propose a variant of Benders Decomposition for multistage stochastic integer programming and apply it to the stochastic unit commitment problem. Thome et al. [55] introduce an extension of the SDDP framework by using Lagrangean Relaxation to convexify the recourse function applied to nonconvex hydrothermal operation planning. Zou et al. [61] present a valid Stochastic Dual Dynamic Integer Programming (SDDiP) algorithm for Multistage Stochastic Integer Programming (MSIP) with binary state variables, and prove that for some of the cuts presented the algorithm converges in a finite number of steps. In their more recent paper, Zou et al. [62] apply their SDDiP algorithm to Stochastic Unit Commitment problems. Finally, Steeger and Rebennack [54] propose a dynamic convexification within Benders Decomposition using Lagrangean relaxation and apply it to the electricity market bidding problem.

In this work, we have adapted the algorithm proposed in [61] for deterministic multi-period MILP models and apply it to the formulation given in Section 4. Moreover, we extend the approach in [61] by allowing for integer and continuous state variables. To facilitate the understanding of the algorithm, we first describe how the multi-period MILP model defined by equations (1)-(28) is decomposed by time period (year). Then, we introduce a more concise notation to represent the decomposed 1-year-long MILPs, and use this notation to describe the Nested Decomposition algorithm in Section 5.2.

5.1. Decomposition by time period (year)

In our formulation, the only constraints that depend on more than one time period are equations (17), (19), (21), and (23). Therefore, these constraints have to be reformulated in order to be able to solve the problem separately for each time period, which is done by duplicating the linking variables, $ngo^{\mathrm{rn}}_{i,r,t}$, $ngo^{\mathrm{th}}_{i,r,t}$, $ngb^{\mathrm{rn}}_{i,r,t}$, $ngb^{\mathrm{th}}_{i,r,t}$.

Equation (17), which computes the number of renewable generators that are operational at time period t based on the number of generators built and retired at t and operational at $t-1$, is substituted by equations

(29) and (30).

$$ngo^{\mathrm{rn}}_{i,r,t} = ngo^{\mathrm{rn,prev}}_{i,r,t} + ngb^{\mathrm{rn}}_{i,r,t} - ngr^{\mathrm{rn}}_{i,r,t} \qquad \forall\, i \in \mathcal{I}^{\mathrm{RN}}_r, t > 1 \tag{29}$$

$$ngo^{\mathrm{rn,prev}}_{i,r,t} = n\hat{g}o^{\mathrm{rn}}_{i,r,t-1} \qquad \leftarrow \mu^{\mathrm{o,rn}}_{i,r,t} \in \mathbb{R}^{|\mathcal{I}^{\mathrm{RN}}_r|+|\mathcal{T}|-1} \qquad \forall\, i \in \mathcal{I}^{\mathrm{RN}}_r, t > 1 \tag{30}$$

Here $ngo^{\mathrm{rn,prev}}_{i,r,t}$ is the duplicated variable representing $ngo^{\mathrm{rn}}_{i,r,t-1}$, and $n\hat{g}o^{\mathrm{rn}}_{i,r,t-1}$ is the solution for $ngo^{\mathrm{rn}}_{i,r,t}$ at time period $t-1$, which is fixed when solving time period t. The Lagrange multiplier $\mu^{\mathrm{o,rn}}_{i,r,t}$ of equation (30) is unrestricted in sign.

Similarly, equation (21), which refers to thermal generators, is substituted by equations (31) and (32).

$$ngo^{\mathrm{th}}_{i,r,t} = ngo^{\mathrm{th,prev}}_{i,r,t} + ngb^{\mathrm{th}}_{i,r,t} - ngr^{\mathrm{th}}_{i,r,t} \qquad \forall\, i \in \mathcal{I}^{\mathrm{TH}}_r, t > 1 \tag{31}$$

$$ngo^{\mathrm{th,prev}}_{i,r,t} = n\hat{g}o^{\mathrm{th}}_{i,r,t-1} \qquad \leftarrow \mu^{\mathrm{o,th}}_{i,r,t} \in \mathbb{R}^{|\mathcal{I}^{\mathrm{TH}}_r|+|\mathcal{T}|-1} \qquad \forall\, i \in \mathcal{I}^{\mathrm{TH}}_r, t > 1 \tag{32}$$

Here $ngo^{\mathrm{th,prev}}_{i,r,t}$ is the duplicated variable representing $ngo^{\mathrm{th}}_{i,r,t-1}$, and $n\hat{g}o^{\mathrm{th}}_{i,r,t-1}$ is the solution for $ngo^{\mathrm{th}}_{i,r,t}$ at time period $t-1$, which is fixed when solving time period t. The Lagrange multiplier $\mu^{\mathrm{o,th}}_{i,r,t}$ of equation (32) is unrestricted in sign.

Equations (19) and (23) compute the age of the new generators that are built during the planning horizon to be able to enforce their retirement (or life extension) when they achieve the end of their lifetime. Hence, these constraints link time period t to time period $t - LT_i$, where LT_i is the expected lifetime of a generator in cluster i. In order to decouple those time periods, (19) has to be replaced by (33) and (34).

$$ngb^{\mathrm{rn,LT}}_{i,r,t} = ngr^{\mathrm{rn}}_{i,r,t} + nge^{\mathrm{rn}}_{i,r,t} \qquad \forall\, i \in \mathcal{I}^{\mathrm{Rnew}}_r, r, t \tag{33}$$

$$ngb^{\mathrm{rn,LT}}_{i,r,t} = \hat{ngb}^{\mathrm{rn}}_{i,r,t-LT_i} \qquad \leftarrow \mu^{\mathrm{b,rn}}_{i,r,t} \in \mathbb{R}^{|\mathcal{I}^{\mathrm{Rnew}}_r|+|\mathcal{R}|+|\mathcal{T}|} \qquad \forall\, i \in \mathcal{I}^{\mathrm{Rnew}}_r, r, t \tag{34}$$

Here $ngb^{\mathrm{rn,LT}}_{i,r,t}$ is the duplicated variable representing the renewable generators built in year $t-LT_i$, $ngb^{\mathrm{rn}}_{i,r,t-LT_i}$. Thus, the model is able to track when the end of the generators' lifetime is, which is the year they were built plus their lifetime, LT_i. $\hat{ngb}^{\mathrm{rn}}_{i,r,t-LT_i}$ is the solution for $ngb^{\mathrm{rn}}_{i,r,t}$ at year $t - LT_i$, which is fixed when solving time period t. The Lagrange multiplier $\mu^{\mathrm{b,rn}}_{i,r,t}$ of equation (34) is unrestricted in sign.

Similarly, constraint (23) is replaced by (35) and (36).

$$ngb^{\mathrm{th,LT}}_{i,r,t} = ngr^{\mathrm{th}}_{i,r,t} + nge^{\mathrm{th}}_{i,r,t} \qquad \forall\, i \in \mathcal{I}^{\mathrm{Tnew}}_r, r, t \tag{35}$$

$$ngb^{\mathrm{th,LT}}_{i,r,t} = \hat{ngb}^{\mathrm{th}}_{i,r,t-LT_i} \qquad \leftarrow \mu^{\mathrm{b,th}}_{i,r,t} \in \mathbb{R}^{|\mathcal{I}^{\mathrm{Tnew}}_r|+|\mathcal{R}|+|\mathcal{T}|} \qquad \forall\, i \in \mathcal{I}^{\mathrm{Tnew}}_r, r, t \tag{36}$$

Here $ngb^{\mathrm{th,LT}}_{i,r,t}$ is the duplicated variable representing the thermal generators built at year $t-LT_i$, $ngb^{\mathrm{th}}_{i,r,t-LT_i}$. $\hat{ngb}^{\mathrm{th}}_{i,r,t-LT_i}$ is the solution for $ngb^{\mathrm{th}}_{i,r,t}$ at time period $t - LT_i$, which is fixed when solving time period t. The

Lagrange multiplier $\mu^{\mathrm{b,th}}_{i,r,t}$ of equation (36) is unrestricted in sign.

Furthermore, the objective function for a given time period is solved independently, and incorporates the cuts for future cost that are added in the following iterations. These cuts, given by equation (38), project the problem onto the subspace defined by the linking variables, and will be explained in detail in Section 5.2.2. Hence, equation (25) is replaced by (37)-(38),

$$\min \Phi_t = \Phi^{\mathrm{opex}}_t + \Phi^{\mathrm{capex}}_t + \Phi^{\mathrm{PEN}}_t + \alpha_t \tag{37}$$

$$\begin{aligned}\alpha_t \quad \geq \hat{\Phi}_{t+1,k} &+ \sum_{i\in\mathcal{I}^{\mathrm{RN}}_r,r} \mu^{\mathrm{o,rn}}_{i,r,t+1,k}\cdot\left(\hat{ngo}^{\mathrm{rn}}_{i,r,t,k} - ngo^{\mathrm{rn}}_{i,r,t}\right) + \sum_{i\in\mathcal{I}^{\mathrm{TH}}_r,r} \mu^{\mathrm{o,th}}_{i,r,t+1,k}\cdot\left(\hat{ngo}^{\mathrm{th}}_{i,r,t,k} - ngo^{\mathrm{th}}_{i,r,t}\right)\\ &+ \sum_{i\in\mathcal{I}^{\mathrm{RN}}_r,r} \mu^{\mathrm{b,rn}}_{i,r,t+LT_i,k}\cdot\left(\hat{ngb}^{\mathrm{rn}}_{i,r,t,k} - ngb^{\mathrm{rn}}_{i,r,t}\right) + \sum_{i\in\mathcal{I}^{\mathrm{TH}}_r,r} \mu^{\mathrm{b,th}}_{i,r,t+LT_i,k}\cdot\left(\hat{ngb}^{\mathrm{th}}_{i,r,t,k} - ngb^{\mathrm{th}}_{i,r,t}\right)\end{aligned} \tag{38}$$

$$\forall\ k$$

where k is the iteration counter.

The MILP subproblem for a given time period t and iteration k, described by equations (1)-(16), (18), (20), (22), (24), (29)-(38), can be more concisely represented by $(\mathscr{P}_{t,k})$.

$$\mathscr{P}_{t,k} : \Phi_{t,k}(\hat{x}_{t-1,k}, \phi_{t,k}) = \min_{x_t,y_t,z_t} \quad f_t(x_t,y_t) + \phi_{t,k}(\hat{x}_{t,k}) \tag{39a}$$

$$\text{s.t.} \qquad z_t = \hat{x}_{t-1,k} \qquad \leftarrow \mu_{t,k}\in\mathbb{R}^n \tag{39b}$$

$$(x_t,y_t,z_t)\in\mathcal{X}_t \tag{39c}$$

where the feasible region $\mathcal{X}_t$ is the *mixed-integer* set given by

$$\mathcal{X}_t = \Big\{ (x_t,y_t,z_t) : (1)-(16),(18),(20),(22),(24),(29),(31),(33),(35) \tag{40a}$$

$$x_t\in\mathbb{Z}^{n_1}_+\times\mathbb{R}^{n_2}_+,\quad y_t\in\mathbb{Z}^{m_1}_+\times\mathbb{R}^{m_2}_+,\quad z_t\in\mathbb{R}^n \Big\} \tag{40b}$$

and $n = n_1 + n_2$, $m = m_1 + m_2$.

The components of $(\mathscr{P}_{t,k})$ map to our problem as follows:

- x_t represents the state (i.e., linking) variables, $ngo^{\mathrm{rn}}_{i,r,t}$, $ngo^{\mathrm{th}}_{i,r,t}$, $ngb^{\mathrm{rn}}_{i,r,t}$, $ngb^{\mathrm{th}}_{i,r,t}$. These are mixed integer variables since $ngo^{\mathrm{rn}}_{i,r,t}, ngb^{\mathrm{rn}}_{i,r,t}\in\mathbb{R}_+$, and $ngo^{\mathrm{th}}_{i,r,t}, ngb^{\mathrm{th}}_{i,r,t}\in\mathbb{Z}_+$.
- z_t represents the duplicated state variables, $ngo^{\mathrm{rn,prev}}_{i,r,t}$, $ngo^{\mathrm{th,prev}}_{i,r,t}$, $ngb^{\mathrm{rn,LT}}_{i,r,t}$, $ngb^{\mathrm{th,LT}}_{i,r,t}$, which are continuous variables.
- y_t represents the local variables, i.e. all the other variables not listed above. These are mixed-integer variables.

- $\hat{x}_{t-1,k}$ is the state of the system at the start of a time period t of iteration k, i.e., the solution for x_t obtained in a previous period of the Forward Pass that is linked to (thus fixed in) the current time period (both in iteration k). In our problem this corresponds to the fixed values of $\hat{ngo}^{\text{rn}}_{i,r,t-1}$, $\hat{ngo}^{\text{th}}_{i,r,t-1}$, $\hat{ngb}^{\text{rn}}_{i,r,t-LT_i}$, $\hat{ngb}^{\text{th}}_{i,r,t-LT_i}$ from equations (30), (32), (34), (36), such that $\hat{x}_{t-1,k} = (\hat{x}^{\text{o}}_{t-1,k}, \hat{x}^{\text{b}}_{t-LT,k})$, where $\hat{x}^{\text{o}}_{t-1,k}$ maps to ($\hat{ngo}^{\text{rn}}_{i,r,t-1}$, $\hat{ngo}^{\text{th}}_{i,r,t-1}$), and $\hat{x}^{\text{b}}_{t-LT,k}$ maps to ($\hat{ngb}^{\text{rn}}_{i,r,t-LT_i}$, $\hat{ngb}^{\text{th}}_{i,r,t-LT_i}$). Note that $\hat{x}^{b}_{t-LT,k}$ is a slight abuse of notation since the lifetimes LT_i of different generator clusters may differ. These fixed values are also used in the following Backward Pass.
- $f(x_t, y_t)$ is the objective function in terms of the state and local variables, x_t and y_t, respectively.
- $\phi_{t,k}$ is the cost-to-go as a function of $\hat{x}_{t,k}$.
- Constraint (39b) represents the equalities (30), (32), (34), (36), and $\mu_{t,k}$ represent their Lagrange multipliers $\mu^{\text{o,rn}}_{i,r,t}$, $\mu^{\text{o,th}}_{i,r,t}$, $\mu^{\text{b,rn}}_{i,r,t}$, $\mu^{\text{b,rn}}_{i,r,t}$, respectively, such that $\mu_{t,k} = (\mu^{\text{o}}_{t,k}, \mu^{\text{b}}_{t,k})$, where $\mu^{\text{o}}_{t,k}$ maps to ($\mu^{\text{o,rn}}_{i,r,t}$, $\mu^{\text{o,th}}_{i,r,t}$), and $\mu^{\text{b}}_{t,k}$ maps to ($\mu^{\text{b,rn}}_{i,r,t}$, $\mu^{\text{b,rn}}_{i,r,t}$).

The cost-to-go function, $\phi_{t,k}(\cdot)$, is defined as:

$$\phi_{t,k}(\hat{x}_{t,k}) := \min_{x_t, \alpha_t} \{\alpha_t : \ \alpha_t \geq \hat{\Phi}_{t+1,k} + \mu^{\intercal}_{t+1,k}(\hat{x}_{t,k} - x_t)\} \tag{41}$$

where $\hat{\Phi}_{t+1,k}$ is the optimal value for time period $t+1$, $\mu_{t+1,k}$ is the Lagrange multiplier for the equality constraint, both obtained in the Backward Pass of iteration k, such that $\mu_{t+1,k} = (\mu^{\text{o}}_{t+1,k}, \mu^{\text{b}}_{t+LT,k})$, where $\mu^{\text{o}}_{t+1,k}$ maps to ($\mu^{\text{o,rn}}_{i,r,t+1}$, $\mu^{\text{o,th}}_{i,r,t+1}$), and $\mu^{\text{b}}_{t+LT,k}$ maps to ($\mu^{\text{b,rn}}_{i,r,t+LT}$, $\mu^{\text{b,rn}}_{i,r,t+LT}$). Note that $\mu^{\text{b}}_{t+LT,k}$ is a slight abuse of notation since the lifetimes LT_i of different generator clusters may differ. $\hat{x}_{t,k}$ is the solution for x_t obtained in the Forward Pass of iteration k and fixed in the following Backward Pass.

5.2. Description of the algorithm

The Nested Decomposition algorithm consists of decomposing the problem per time period (year) and solving it iteratively in a forward and a backward fashion. The Forward Pass yields a feasible upper bound, while the Backward Pass, which generates cuts from the relaxed subproblems, provides a lower bound. New cuts are added in the Backward Pass of each iteration k, and are kept in the following Forward Pass, until the difference between the upper and lower bounds is less then a pre-specified tolerance, ϵ_1, as shown in Figure 3.

5.2.1. Forward Pass

The purpose of the forward pass is to generate a feasible solution to the full problem. It accomplishes this by making decisions in time period t, implementing the investment decisions for that period, and then repeating the process in the subsequent period. Therefore, this first step consists of solving the optimization

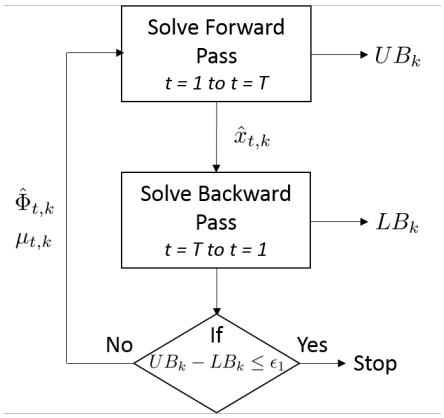

Figure 3: Steps at iteration k of the Nested Decomposition algorithm

problem for each time period sequentially, using the solution from the appropriate previous time period for $\hat{x}_{t-1,k}$. This first part of the algorithm, the Forward Pass, is illustrated in Figure 4.

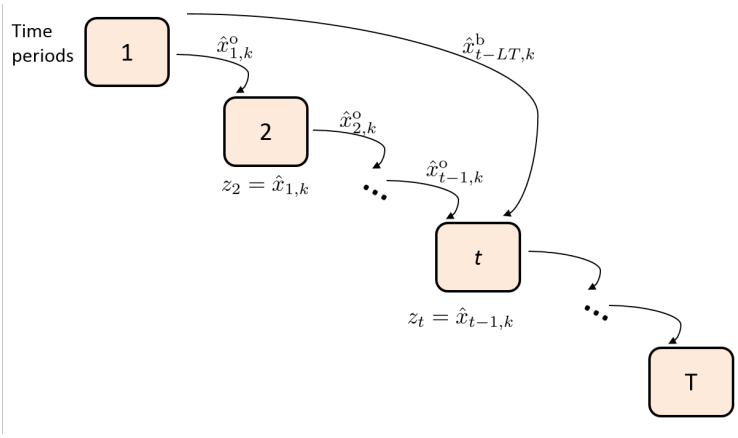

Figure 4: Forward Pass for iteration k generates a feasible solution to the full/original problem (over the full planning horizon)

The problem is assumed to have complete continuous recourse, which means that for any value of state variable (i.e., linking variable) and local integer variables, there are values for the continuous local variables such that the solution is feasible. This assumption is valid since feasibility can be achieved by adding nonnegative slack variables and penalizing them in the objective function.

The upper bound, UB_k, is calculated in the Forward Pass as follows in (42). It is easy to see that the sum of the optimal solutions of the Forward Pass subproblems at iteration k, $\hat{\Phi}_{t,k}$, minus the cost-to-go approximations, $\hat{\alpha}_t$, for all time periods at that iteration, yields a valid upper bound, since the sequential solution of the time periods in a myopic fashion, without relaxing any constraint or integrality, gives a feasible

solution to the full-space MILP problem.

$$UB_k = \sum_t \left(\hat{\Phi}_{t,k} - \hat{\alpha}_t \right) \qquad \forall\ k \tag{42}$$

5.2.2. Backward Pass

After solving the Forward subproblem for all the time periods, the next step is the Backward Pass, the purpose of which is to generate cuts. This step consists of solving the subproblems from the last to the first time period, so the solutions of future periods can be used to generate cuts to provide approximations to predict the cost-to-go functions within the planning horizon. These are cumulative cuts, but specific for each time period; i.e., they are added at each iteration k whenever a new Backward Pass subproblem for year t is solved, and they are kept in the formulation of the following Forward Pass. Note that the fixed variables stored in the Forward Pass, $\hat{x}_t$, are also used in the Backward Pass. The overall procedure of the Backward Pass is represented in Figure 5.

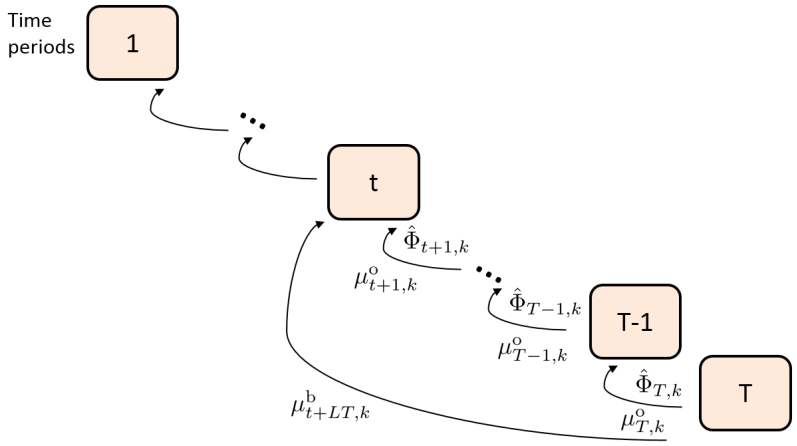

Figure 5: The backward pass generates cuts for the cost-to-go function approximations using the solutions from the forward pass

The lower bound, LB_k, is calculated in the Backward Pass as in (43). It is easy to see that the relaxed solution of the first time period $t = 1$ is a lower bound to the total cost since it only has a subset of the original constraints of the original problem.

$$LB_k = \hat{\Phi}_{1,k} \qquad \forall\ k \tag{43}$$

If our problem were convex and solved by standard Nested Benders Decomposition, the objective value and the Lagrange multiplier of the equality constraint (39b) would be enough to generate the Benders cut

(44). We assume here that t is fixed.

$$\alpha_t \geq \hat{\Phi}_{t+1,k} + \mu_{t+1,k}^{\intercal}(\hat{x}_{t,k} - x_t) \qquad \forall\ k \tag{44}$$

However, the cut cannot be directly generated because the subproblems have integer variables, and thus, are non-convex. In order to provide a valid cut, the subproblems have to be convexified, which can be done by considering the linear or the Lagrangean relaxation of the MILP. The cuts generated by the relaxed problems are the Benders and Lagrangean cuts, respectively (which follows the same notation as [61]). A third type of cut proposed by [61] as a Strengthened Benders cut is also valid for the Backward Pass subproblems.

The choice of cuts directly impacts the performance of the algorithm as some cuts are tighter and more/less computationally expensive to generate than the others. It is possible, and sometimes recommended, to combine different cuts in the algorithm if one does not strictly dominate the other. Following, we will define the different types of cuts that can be used in the Backward Pass, present their potential advantages and disadvantages, and in Section 5.3 we will demonstrate their validity.

Benders cut. The Benders cut is exactly like (44), but the coefficients, i.e., the optimal solution of the immediately after period, $\hat{\Phi}^{\text{LP}}_{t+1,k}$, and the multipliers for the complicating equalities, $\mu_{t+1,k}$, are obtained from the solution of the linear relaxation, as represented in equation (45). We assume here that t is fixed.

$$\alpha_t \geq \hat{\Phi}^{\text{LP}}_{t+1,k} + \mu_{t+1,k}^{\intercal}(\hat{x}_{t,k} - x_t) \qquad \forall\ k \tag{45}$$

This is the weakest of the possible cuts, but it has the advantage of being easily and quickly computed. It is expected to perform well if the formulation is tight and the solution of the linear relaxation is close to the actual solution of the MILP [51, 49]. For certain multistage capacity planning problems with integer recourse, there is evidence that Benders cuts alone are sufficient for driving the optimality gap to zero as the number of stages increases (see, e.g., [23], Corollaries 1 and 2). However, it is important to highlight that if this is the cut being used, the algorithm does not have guaranteed finite convergence since there can be a duality gap.

Lagrangean cut. The MILP subproblem ($\mathscr{P}_{t,k}$), given by (39), can also be convexified by considering its Lagrangean relaxation, which yields the convex hull of the noncomplicating constraints [20]. In our case this is done by dualizing the linking equalities (39b) and penalizing their violation in the objective function by the vector of Lagrange multipliers, $\mu_{t,k}$. Thus, the Lagrangean relaxation of ($\mathscr{P}_{t,k}$) is defined by ($\mathscr{L}_{t,k}$), in

(46).

$$\mathscr{L}_{t,k} : \Phi_t^{\text{LR}}(\mu_{t,k}, \hat{x}_{t-1,k}, \phi_{t,k}) = \begin{array}{ll} \min\limits_{x_t, y_t, z_t} & f_t(x_t, y_t) + \phi_{t,k}(\hat{x}_{t,k}) - \mu_{t,k}^{\intercal}(z_t - \hat{x}_{t-1,k}) \\ \text{s.t.} & (x_t, y_t, z_t) \in \mathcal{X}_t \end{array} \tag{46}$$

The closer the Lagrange multipliers are to their optimal value, the tighter the approximation is, and the stronger the cuts generated by theses multipliers are. The optimal Lagrange multipliers, $\bar{\mu}_{t,k}$, are obtained by the maximization problem (47).

$$\Phi_{t,k}^{\text{LD}}(\hat{x}_{t-1,k}, \phi_{t,k}) = \max_{\mu_{t,k}} \ \Phi_{t,k}^{\text{LR}}(\mu_{t,k}, \hat{x}_{t-1,k}, \phi_{t,k}) \tag{47}$$

The Lagrangean cut uses the coefficients obtained by the maximization problem (47), as represented in equation (48). We assume here that t is fixed.

$$\alpha_t \geq \hat{\Phi}_{t+1,k}^{\text{LD}} + \bar{\mu}_{t+1,k}^{\intercal}(\hat{x}_{t,k} - x_t) \qquad \forall\ k \tag{48}$$

The maximization problem in (47) can, however, be computationally expensive. Therefore, we adapted the Lagrange multiplier optimization algorithm for each of the subproblems of the Backward Pass based on [55] using the subgradient method [18]. The steps of the Backward Pass within the Nested Decomposition algorithm if only Lagrangean cuts are used is described below.

For a time period $t = T, ..., 1$ in iteration k:

1. Solve the original MILP subproblem in (39) to get the actual objective value, $\Phi_{t,k}^{\text{OP}}$;
2. Solve the linear relaxation of the MILP subproblem in (39), and store the dual variables, $\mu_{t,k}$;
3. Use the dual variables from the LP relaxation as an initial guess for the Lagrange multipliers;
4. Solve the Lagrangean subproblem in (46) to obtain the optimal value $\Phi_{t,k}^{\text{LR}}$;
5. Check stopping criteria:
   (a) If $\Phi_{t,k}^{\text{OP}} - \Phi_{t,k}^{\text{LR}} \leq \epsilon_2$, where ϵ_2 is a pre-specified tolerance, store the optimal value $\Phi_{t,k}^{\text{LR}}$, and multipliers $\mu_{t,k}$, and go to the next subproblem $t-1$, adding the appropriate future cost cut.
   (b) If no significant progress is achieved after re-solving the Lagrangean relaxation, i.e., if $|\Phi_{t,k}^{\text{LR,old}} - \Phi_{t,k}^{\text{LR}}| \leq \epsilon_3$, where ϵ_3 is a pre-specified tolerance and $\Phi_{t,k}^{\text{LR,old}}$ is the solution of the Lagrangean Relaxation in the previous step of the subgradient method, this means that no further effort should be made to decrease the duality gap of this subproblem in this iteration k. Store the optimal value $\Phi_{t,k}^{\text{LR}}$, and the multipliers $\mu_{t,k}$, and go to the next subproblem $t-1$, adding the appropriate future

cost cut.

6. If the stopping criteria are not met, update the set of multipliers using the subgradient method:
   $\mu_{t,k} = \mu_{t,k} - step_{t,k} \cdot (z_t - \hat{x}_{t-1,k})$

   where $step_{t,k} = \frac{\Phi^{\text{OP}}_{t,k} - \Phi^{\text{LR}}_{t,k}}{(z_t)^2}$, and go back to step 3.

Zou et al. [61] proved that if all the linking variables are binary, the Lagrangean cut is tight and the Nested Decomposition algorithm converges in a finite number of steps. However, in our case the state variables are integer and continuous, thus finite convergence is not guaranteed and there may be a duality gap associated with the solution [2].

Strengthened Benders cut. As mentioned in the previous sections, depending on the structure of the problem, the Benders cuts can be weak and the Lagrangean cuts can take a long time to compute. In order to mitigate potential performance issues, Zou et al. [61] proposed the Strengthened Benders cut, which is a compromise between Benders and Lagrangean cuts. Its generation is similar to the Lagrangean cut, but it does not use the subgradient method to improve the multipliers. Instead, it uses the coefficients from the first Lagrangean relaxation solved after the initialization of the multipliers using LP relaxation as shown in (49). We assume that t is fixed here.

$$\alpha_t \geq \hat{\Phi}^{\text{LR}}_{t+1,k} + \mu^{\intercal}_{t+1,k}(\hat{x}_{t,k} - x_t) \qquad \forall\ k \tag{49}$$

These cuts are at least as tight as the Benders cut [61], but usually less computationally expensive than the Lagrangean cuts.

5.3. Validity of the cuts

The Nested Decomposition algorithm is valid as long as the cuts generated in the Backwards Pass are valid according to the following definition [61].

Definition 1. Let $\{(\hat{\Phi}_{t,k}, \mu_{t,k})\}$ be the coefficients obtained from the Backward Pass in iteration k, and $\hat{x}_{t,k}$ be the value for x_t obtained in the Forward Pass of iteration k and fixed for the Backward Pass. Let $\Phi_t(\hat{x}_{t-1,k})$ be the optimal solution for subproblem t for a $\hat{x}_{t-1,k}$ assuming exact representation of the cost-to-go-function. A cut is valid for all periods $t \in \mathcal{T}$ and all iterations $k \in \mathcal{K}$ if

$$\Phi_t(\hat{x}_{t-1,k}) \geq \hat{\Phi}_{t+1,k} + (\mu_{t+1,k})^{\intercal}(\hat{x}_{t,k} - x_t) \tag{50}$$

The validity of the Benders, Lagrangean and Strengthened Benders cuts is proved in Theorems 1, 2, and 3, respectively.

Theorem 1. The Benders cut (45), generated by solving the linear relaxation of $(\mathscr{P}_{t,k})$, underestimates $\Phi_t(\hat{x}_{t-1,k})$.

Proof. It trivially follows that the linear relaxation of $(\mathscr{P}_{t,k})$, with optimal value $\Phi^{\text{LP}}_{t,k}$, underestimates the optimal value $\Phi_{t,k}$ of the original problem $(\mathscr{P}_{t,k})$, that is, $\Phi_{t,k} \geq \Phi^{\text{LP}}_{t,k}$. Therefore, since the Benders cut is valid for the LP problem, as proved by [7], it is also valid for the MILP problem. □

Theorem 2. The Lagrangean cut (48), generated by solving the Lagrangean dual of $(\mathscr{P}_{t,k})$, underestimates $\Phi_t(\hat{x}_{t-1,k})$.

Proof. This proof is an extension for mixed-integer state variables of the proof presented in [61], in which only binary state variables are considered. For the last period $t = T$, the cost-to-go function $\phi_{t,k} = 0$. If we relax the equality $z_t = \hat{x}_{t-1,k}$ in $(\mathscr{P}_{t,k})$ using the optimal multiplier of the Lagrangean problem (47), $\bar{\mu}_{k,t}$, we have for any $\hat{x}_{t-1,k} \in \mathbb{Z}^{n_1}_+ \times \mathbb{R}^{n_2}_+$, where $n_1 + n_2 = n$,

$$\begin{aligned} \Phi_t(\hat{x}_{t-1,k}) \geq & \min_{x_t,y_t,z_t} \left\{ f_t(x_t, y_t) - \bar{\mu}^{\intercal}_{t,k}(z_t - \hat{x}_{t-1,k}) : (x_t, y_t, z_t) \in \mathcal{X}_t \right\} \\ = & \quad \Phi^{\text{LD}}_{t,k} \end{aligned} \tag{51}$$

Thus, the Lagrangean cut is valid for $t = T$. Next, we prove by induction that the Lagrangean cut is also valid for the remaining timeperiods. Consider $t \leq T - 1$ and assume the Lagrangean cut defined by $\{(\hat{\Phi}_{t+2,k}, \bar{\mu}_{t+2,k}, \hat{x}_{t+1,k})\}$ is valid. Note that:

$$\Phi_t(\hat{x}_{t-1,k}) = \min_{x_t,y_t,z_t,\alpha_t} \left\{ f_t(x_t, y_t) + \alpha_t : (x_t, y_t, z_t) \in \mathcal{X}_t,\ z_t = \hat{x}_{t-1,k},\ \alpha_t \geq \Phi_{t+1} \right\} \tag{52}$$

Since the cut defined by $\{(\hat{\Phi}_{t+2,k}, \bar{\mu}_{t+2,k}, \hat{x}_{t+1,k})\}$ is valid, i.e. $\Phi_{t+1}(x_t) \geq \hat{\Phi}_{t+2,k} + (\bar{\mu}_{t+2,k})^{\intercal}(\hat{x}_{t+1,k} - x_{t+1})$ for any $x_{t+1} \in \mathbb{Z}^{n_1}_+ \times \mathbb{R}^{n_2}_+$, then the new feasible region $\mathcal{X}'_t$ that incorporates this cut is a relaxation of the original feasible region $\mathcal{X}_t$ of (52). Hence, by dualizing the equality constraint we have that

$$\begin{aligned} \Phi_t(\hat{x}_{t-1,k}) \geq & \min_{x_t,y_t,z_t,\alpha_t} \left\{ f_t(x_t, y_t) + \alpha_t : \quad (x_t, y_t, z_t) \in \mathcal{X}'_t, z_t = \hat{x}_{t-1,k} \right\} \\ \geq & \min_{x_t,y_t,z_t,\alpha_t} \left\{ f_t(x_t, y_t) + \alpha_t - \quad \bar{\mu}^{\intercal}_{t,k}(z_t - \hat{x}_{t-1,k}) : (x_t, y_t, z_t) \in \mathcal{X}'_t \right\} \\ = & \quad \Phi^{\text{LD}}_{t,k} \end{aligned} \tag{53}$$

Thus, the Lagrangean cut defined by $\{(\hat{\Phi}_{t+1,k}, \bar{\mu}_{t+1,k}, \hat{x}_{t,k})\}$ is valid for $t \in \mathcal{T}$, which completes the proof of Theorem 2. □

Theorem 3. The Strengthened Benders cut (49), generated by solving the Lagrangean relaxation of $(\mathscr{P}_{t,k})$ using the multipliers from the linear relaxation of $(\mathscr{P}_{t,k})$, underestimates $\Phi_t(\hat{x}_{t-1,k})$.

Proof. Since the Lagrangean dual is the tightest of the Lagrangean relaxations, the proof for Theorem 2 is also valid for Theorem 3. □

5.4. Accelerated Nested Decomposition Algorithm

The proposed Nested Decomposition algorithm (as described in Sections 5.2.1 and 5.2.2) performs very well for our type of problem as will be shown in the results. However, there is potential for improvement, particularly in the early iterations, since the initial planning years have little information of what happens in future years. It is well known that Benders decomposition and its variants can oscillate wildly during initial iterations when the cost-to-go approximations of future stages is poor [49]. Here, we propose an acceleration technique aimed at 'warm-starting' the initial cost-to-go approximations with information from an aggregated expansion model, which can potentially speed up the convergence.

The first step is to solve an aggregated version of the full-space MILP and use its solution to pre-generate cuts before entering in the first Forward Pass. The level of aggregation can be decided by the user. The key is to balance the trade-off between a highly aggregated model, which is fast to solve but generates weaker cuts, and a barely aggregated model, which will take almost as long to solve as the original MILP (especially if the solution of the LP relaxation is the main bottleneck) but generates stronger cuts. After some preliminary experiments, we opted to aggregate the model by having only one representative day per year with hourly-level information and relaxing the integrality of the unit commitment variables.

The aggregated model can provide multiple solutions for the linking variables x_t by using the solver's *solution pool* option [24]. The decision of how many solutions to use for cut generation is also user-defined (the larger the number of solutions used, the better the representation of the original model, but the longer it takes to compute). These solution values are fixed, $\hat{x}_{t,0}$, and used to generate cuts in a pre-Backward Pass, which is solved before entering the Nested Decomposition iterations. Thus, the algorithm uses the information from this aggregated model that has a view of the full planning horizon (but in an approximated fashion) to gather information about future periods and what are potential good solutions, so that in early iterations, the cost-to-go approximations possess stronger cuts and, consequently, more relevant information about the cost of future investment decisions. Note that any of the cuts presented before can be used in this step of the algorithm.

After finishing this pre-Backward Pass, the algorithm goes to the first Forward Pass keeping the pre-generated cuts. From this point on, the algorithm follows the same steps as in the standard Nested Decomposition. The main steps of the improved Nested Decomposition are shown in Figure 6.

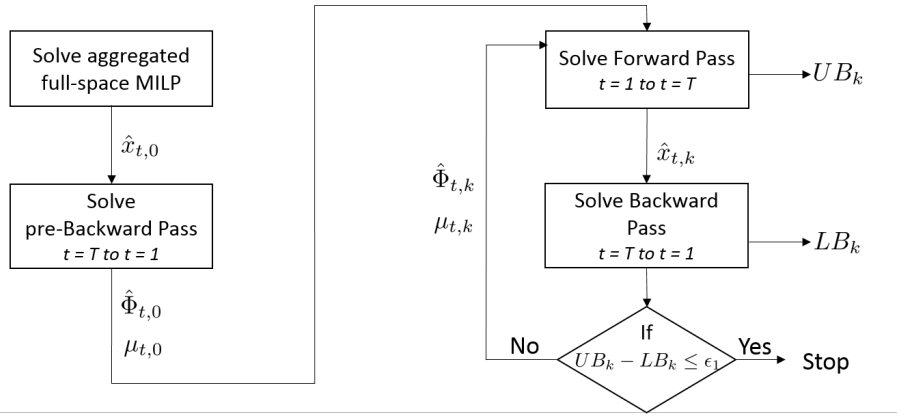

Figure 6: Accelerated Nested Decomposition algorithm using pre-generated cuts from an aggregated model to improve initial cost-to-go approximations.

Remark. Both the Nested Decomposition Algorithm and its accelerated version do not have guaranteed finite convergence with any of the presented cuts for the case of integer and continuous state variables, since there might be a duality gap associated with the solution.

6. Case study

In order to test the formulation proposed in Section 4 and the algorithm proposed in Section 5, we applied them to a case study approximating the Texas Interconnection, a power grid that covers most of the state of Texas. This system is managed by the Electric Reliability Council of Texas (ERCOT), which is an independent system operator responsible for the flow of electric power of about 90% of Texas' electric load. The choice for this interconnection is based on the fact that it is one of the three minor grids in the continental U.S. power transmission grid, so it is a fairly isolated system that has manageable size. Since our focus is on assessing our multi-scale model and decomposition algorithms, the results presented here should not be interpreted as a detailed analysis of the ERCOT system along the lines of other efforts in the literature [34].

Within the ERCOT covered area, we consider four geographical regions: Northeast, West, Coastal and South. We also include a fifth region, Panhandle, which is technically outside the ERCOT limits but due to its renewable generation potential, it supplies electricity to the ERCOT regions. Thus, Panhandle is considered a zone with zero load demand, i.e., it is only a supplier, not a consumer. The regions are shown in Figure 1. We assume the geographical midpoints for the Northeast, West, Coastal, South and Panhandle are Dallas, Glasscock County, Houston, San Antonio, and Amarillo, respectively.

For each of the regions, we use load and capacity factor profiles with an hourly resolution. Representative days are constructed using a k-means clustering algorithm. After normalizing 2004-2010 zonal load and renewables profiles into a list of 2555 vectors (7 years of data times 365 days per year – leap days were

excluded), d clusters (for $d = 1, ..., D = 12$) are constructed to find a most representative day, chosen as the day closest in Euclidean distance to the centroid of each resulting cluster. Each representative day is then assigned a weight proportional to the number of historical days in the corresponding cluster. Finally, the weighted load profile is normalized to equal 2015 aggregate ERCOT load of 347.5 TWh so that a fair comparison between runs with different numbers of clusters can be made. We assume a constant load growth of 1.4%/year to project load for the 30 years of the planning horizon being studied.

The investment cost, fixed and variable operating costs are derived from the National Renewable Energy Laboratory (NREL), available in the 2016 Annual Technology Baseline (ATB) Spreadsheet [37]. We consider a 30-year time horizon, in which the first year is 2015. The fuel price data for coal, natural gas and uranium correspond to the reference scenario of EIA Annual Energy Outlook 2016 [56]. A discount rate of 5.7% as chosen in [52] is used to distinguish between the costs incurred in various years of the planning horizon. All of our computation results assume no carbon tax or renewable generation quota is imposed.

We assume that the transmission loss is 1%/100 miles, which is the same fraction loss per miles used by [52]. The transmission line capacities were generated combining information from [42, 17, 1]. The operating data such as ramping limits, minimum operating reserve requirements, are from a range of sources [32, 25, 36, 59, 13]. The planning reserve requirement is assumed to be 13.75% [47]. For the nominal capacity of the clusters in each region, we use data from [13], which has a list of all the power plants in ERCOT for different categories. We compile this data into 7 existing cluster types and combine them with the data from [15] to have this information divided by the regions. Then, for each of the clusters in each region, we assume that the nominal capacity is the mean size of all generators within that cluster.

We first solve the proposed model for 1 to 12 representative days, with an optimality gap tolerance of 1%, to assess the impact of the number of representative days in the planning strategy. Then, we solve the proposed model as a full-space MILP, as well as using the Nested Decomposition with the cuts presented in Section 5.2.2 for the 4 representative days variant of the model, and applied the acceleration technique from Section 5.4 to the cut with the best performance. The two best versions of the algorithm are also tested for the 12 representative days.

Our computational tests were performed on a standard desktop computer with an Intel(R) Core(TM) i7-2600 CPU @ 3.40 GHz processor, with 8GB of RAM, running on Windows 7. We implement the monolithic formulation and all the versions of the Nested Decomposition algorithm in GAMS 24.7.1, and solve the LPs and MILPs using CPLEX version 12.6.3.

6.1. Impact of the number of representative days in the planning strategy

To evaluate the impact of including more representative days on the generation investment decisions, we ran the proposed model for 1 to 12 representative days per year. The fractional difference relative to the 12-day representation by generation source is shown in Figure 7. This plot does not include coal and nuclear

because their final year capacities are exactly the same irrespective of the number of representative days considered. The wind capacity suffers minor fluctuation and slowly converges to a capacity that is 3% higher than the 1-day model projected. On the other hand, a major difference occurs in projected PV and NG generation capacities. For 1 to 4 representative days, there is a clear trend of decreasing PV and increasing NG contributions as we increase the number of representative days. With the exception of the outlier in the "5 representative days" results, this trend can be seen for all the variants until the results reach a plateau with small fluctuations. This trend is in agreement with the results presented in [33]. It is important to highlight that the jump in the installed capacity of PV and natural gas with 5 representative days is due in part to the approach for selecting representative days. Namely, historical data are clustered according to the *joint* distribution of load and renewables capacity factors. This does not guarantee that the profile used to represent individual technologies will better match that technology's historical duration curves as the number of representative days is increased.

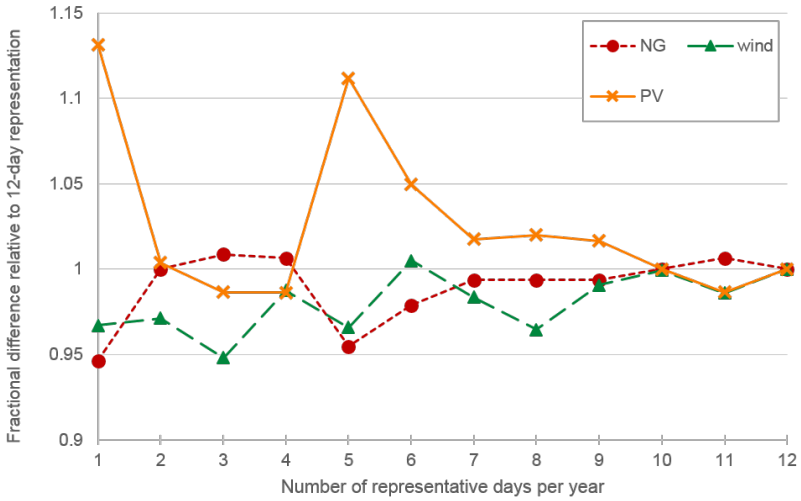

Figure 7: Capacity projections of natural gas, solar PV and wind at the end of the time horizon, varying with the number of representative days selected. The capacities are represented as a fraction of capacity projected by the 12-day model.

After this preliminary experiment, we opted to test the algorithm for 4 and 12 representative days per year. The 4 representative days variant was chosen because its results are similar to those of the 12-day variant, while requiring less computational time. We also opted to solve the 12 representative days variant because it exploits better the full potential of the Nested Decomposition algorithm, which may be needed for more complex models that require higher time resolution (e.g, models that consider storage)[33].

6.2. 4 representative days variant

The full-space MILP model using 4 representative days has 1,201,761 constraints, 413,644 discrete variables, and 594,147 continuous variables. After invoking CPLEX's presolver, the model was reduced to 800,755 constraints, 25,231 binary variables, 388,493 integer variables, and 225,367 continuous variables. We solved the problem using several methods: (1) we solved the monolithic formulation directly in CPLEX; (2) we applied the Nested Decomposition algorithm with each of the cut options; (3) we employed the acceleration technique to the Nested Decomposition variant with the best performance in (2). For all the versions of the algorithm, we set a maximum of 20 iterations and an optimality gap tolerance of 0.01% for any MILP subproblem, and for the full-space MILP we set a maximum time of 4 hours. The comparison between the performance of the full-space MILP and the various algorithm versions is shown in Figure 8.

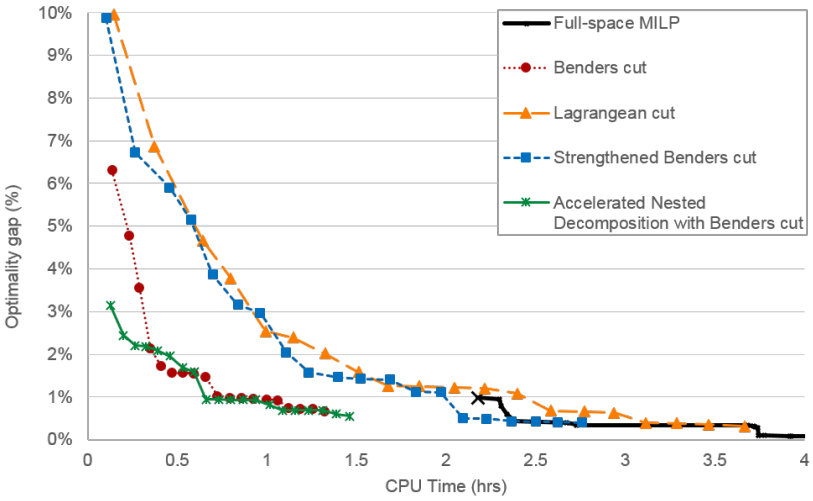

Figure 8: Algorithm performance in the 4-representative day model. The results show that the Benders cut and its accelerated version are the fastest in finding a solution within 1% optimality gap.

All of the versions of the algorithm found solutions within 10% gap in less than 10 minutes, while the first integer solution by the monolithic MILP took 2.04 hours. The Benders cuts was the most efficient among the possible cuts, finding a solution within 1% gap in 47.5 minutes. This was already expected due to the tightness of the linear relaxation, and to how quickly the Benders cuts are generated. Further improvements can be gained by employing the accelerated version (see Section 5.4) in which warm-start cuts are generated before invoking the Nested Benders approach. The accelerated Nested Decomposition with Benders cuts greatly reduces the optimality gap in the initial iterations and is the fastest (39.6 minutes) at finding a solution within 1% optimality gap.

6.3. 12 representative days variant

The full-space MILP model for the 12 representative days variant has 3,626,721 constraints, 1,243,084 discrete variables, and 1,774,947 continuous variables. After invoking CPLEX's presolver, the model was reduced to 2,428,687 constraints, 77,071 binary variables, 1,183,373 integer variables, and 673,825 continuous variables. Despite these reductions, CPLEX terminated with a 100% optimality gap after solving the monolithic formulation for 24 hours, our chosen time limit. We solved the problem using the Nested Decomposition algorithm with the two versions that performed the best using 4 representative days (Benders cuts, and warm-start cuts + Benders cuts). For those, we set a maximum of 20 iterations and an optimality gap tolerance of 0.01% for any MILP subproblem. The comparison between the performance of the two algorithms is shown in Figure 9.

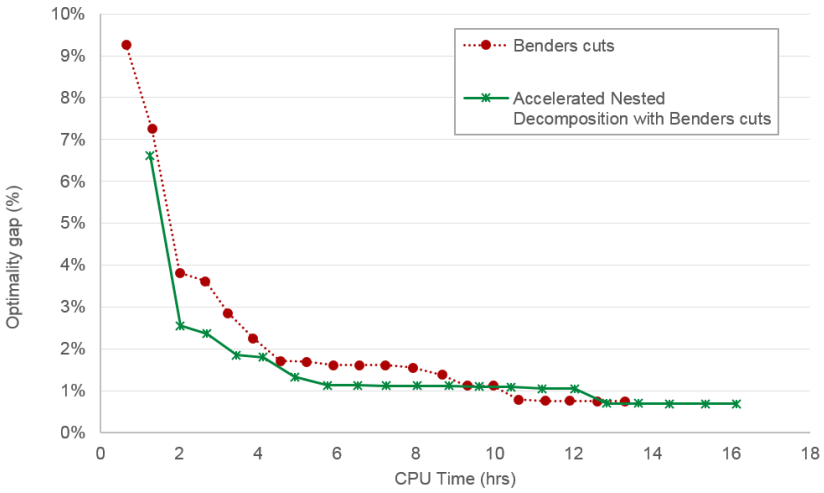

Figure 9: Algorithm performance in the 12-representative day model. The results show that the Accelerated Nested Decomposition algorithm provides smaller optimality gaps in the initial iterations.

Both versions of the algorithm obtained solutions within 2% optimality gap in less than 5 hours. We notice again that by "warm-starting" the algorithm we were able to reduce the optimality gap in the initial iterations. The accelerated version found a solution with a 2% gap in its 4^{th} iteration and 3.5 hours, while the normal Benders cut version took 4.6 hours and 7 iterations to reach the same optimality gap.

6.4. Cost breakdown comparison

We also compare the impact of the having more representative days in the cost breakdown, which is shown in Figure 10. The results indicate that for both cases the cost is driven by the fuel consumption (coal, natural-gas and uranium), which accounts for about 60% of the total cost. The major difference between these representations is in the startup contribution, which goes from 0.3% to 1.5% of the total cost when considering 4 and 12 representative days. These results show that by having a better representation of the

dynamics of the systems, i.e., higher number of representative days with hourly load and capacity factor, the startup cost becomes more relevant.

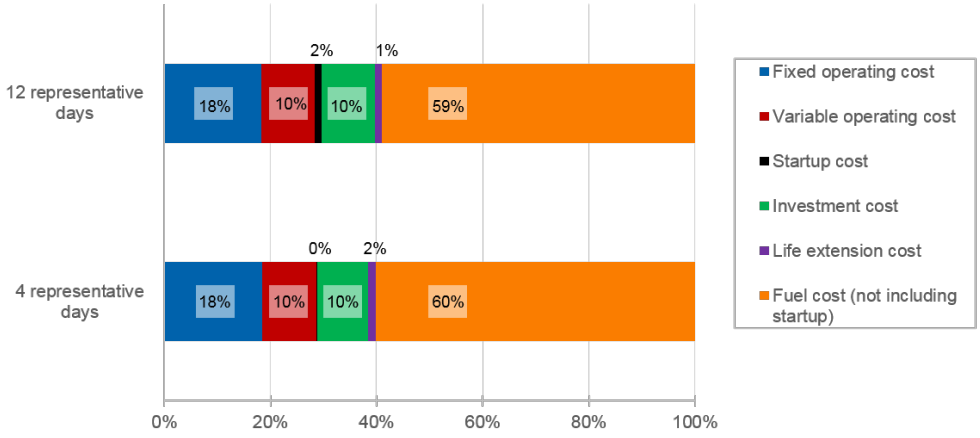

Figure 10: Breakdown of system costs using 4 and 12 representative days.

7. Conclusion

In this paper we propose an MILP model to solve power systems planning problem considering increasing share of renewables. We adopt clustering and time scale strategies in order to reduce the size of the model without greatly impacting the quality of the results. Although there are novelties in the problem formulation, especially in the description of retirement and handling of multi-scale aspects, the major contribution of this paper is in its solution strategy. We develop a decomposition algorithm based on Nested Benders Decomposition for this multi-period deterministic problem with integer and continuous state variables. We also present several options of valid cuts for the Backward Pass, and make use of an acceleration technique to speed up the conversion of the algorithm.

The formulation and solution framework are tested for a case study in the ERCOT region. The results show that the algorithm can provide substantial speed-up and allow the solution of larger instances. This improvement in solution time is important because it allows one to perform several sensitivity analysis and better understand the drivers for a variety of scenarios. Additionally, we conclude that for the proposed model and case study, it is sufficient to have 4 representative days per year in order to adequately represent the variability in load and capacity factor for the different regions. In other instances, such as when studying the potential role of energy storage technologies, it may be necessary to consider more than 4 representative days to model grid operations. For those cases, the developed algorithms are shown to be critical to finding an optimal solution in a reasonable amount of time.

There are several future research direction worth investigating. Regarding the model formulation, it would be interesting to see the impacts of including storage, transmission expansion, and having a more detailed

representation of transmission. Regarding the algorithm, there is potential for parallelization, especially for the Lagrangean cut version of the algorithm, which could improve its performance and make it more competitive with other versions.

Nomenclature

Indices and Sets

$r \in \mathcal{R}$	set of regions within the area considered
$i \in \mathcal{I}$	set of generator clusters
$i \in \mathcal{I}_r$	set of generator clusters in region r
$i \in \mathcal{I}_r^{\text{old}}$	set of existing generator clusters in region r at the beginning of the time horizon, $\mathcal{I}_r^{\text{old}} \subseteq \mathcal{I}_r$
$i \in \mathcal{I}_r^{\text{new}}$	set of potential generator clusters in region r, $\mathcal{I}_r^{\text{new}} \subseteq \mathcal{I}_r$
$i \in \mathcal{I}_r^{\text{TH}}$	set of thermal generator clusters in region r, $\mathcal{I}_r^{\text{TH}} \subseteq \mathcal{I}_r$
$i \in \mathcal{I}_r^{\text{RN}}$	set of renewable generator clusters in region r, $\mathcal{I}_r^{\text{RN}} \subseteq \mathcal{I}_r$
$i \in \mathcal{I}_r^{\text{Told}}$	set of existing thermal generator clusters in region r, $\mathcal{I}_r^{\text{Told}} \subseteq \mathcal{I}_r^{\text{TH}}$
$i \in \mathcal{I}_r^{\text{Tnew}}$	set of potential thermal generator clusters in region r, $\mathcal{I}_r^{\text{Tnew}} \subseteq \mathcal{I}_r^{\text{TH}}$
$i \in \mathcal{I}_r^{\text{Rold}}$	set of existing renewable generator clusters in region r, $\mathcal{I}_r^{\text{Rold}} \subseteq \mathcal{I}_r^{\text{RN}}$
$i \in \mathcal{I}_r^{\text{Rnew}}$	set of potential renewable generator clusters in region r, $\mathcal{I}_r^{\text{Rnew}} \subseteq \mathcal{I}_r^{\text{RN}}$
$t \in \mathcal{T}$	set of time periods (years) within the planning horizon
$d \in \mathcal{D}$	set of representative days in each year t
$s \in \mathcal{S}$	set of sub-periods of time per representative day d in year t
$k \in \mathcal{K}$	set of iterations in the Nested Decomposition algorithm

Deterministic Parameters

$L_{r,t,d,s}$	load demand in region r in sub-period s of representative day d of year t (MW)
$L_t^{\max}$	peak load in year t (MW)
W_d	weight of the representative day d
Hs	duration of sub-period s (hours)
$Qg_{i,r}^{\text{np}}$	nameplate (nominal) capacity of a generator in cluster i in region r (MW)
$Ng_{i,r}^{\text{old}}$	number of existing generators in each cluster, $i \in \mathcal{I}_r^{\text{old}}$, per region r at the beginning of the time horizon
$Ng_i^{\max}$	maximum number of generators in the potential clusters $i \in \mathcal{I}_r^{\text{new}}$
$Q_{i,t}^{\text{inst,UB}}$	upper bound on yearly capacity installations based on generation technology ($^{\text{MW}}/_{\text{year}}$)
$R_t^{\min}$	system's minimum reserve margin for year t (fraction of the peak load)
ED_t	energy demand during year t (MWh)
LT_i	expected lifetime of generation cluster i (years)
T_t^{remain}	remaining time until the end of the time horizon at year t (years)
$Ng_{i,r,t}^{\text{r}}$	number of generators in cluster i of region r that achieved their expected lifetime
Q_i^{v}	capacity value of generation cluster i (fraction of the nameplate capacity)
$Cf_{i,r,t,d,s}$	capacity factor of generation cluster $i \in \mathcal{I}_r^{\text{RN}}$ in region r at sub-period s, of representative day d of year t (fraction of the nameplate capacity)
$Pg_i^{\min}$	minimum operating output of a generator in cluster $i \in \mathcal{I}_r^{\text{TH}}$ (fraction of the nameplate capacity)
$Ru_i^{\max}$	maximum ramp-up rate for cluster $i \in \mathcal{I}_r^{\text{TH}}$ (fraction of nameplate capacity)
$Rd_i^{\max}$	maximum ramp-down rate for cluster $i \in \mathcal{I}_r^{\text{TH}}$ (fraction of nameplate capacity)
F_i^{start}	fuel usage at startup (MMbtu/MW)
$Frac_i^{\text{spin}}$	maximum fraction of nameplate capacity of each generator that can contribute to spinning reserves (fraction of nameplate capacity)
$Frac_i^{\text{Qstart}}$	maximum fraction of nameplate capacity of each generator that can contribute to quick-start reserves (fraction of nameplate capacity)
$Op^{\min}$	minimum total operating reserve (fraction of the load demand)
$Spin^{\min}$	minimum spinning operating reserve (fraction of the load demand)
$T_{r,r'}^{\text{loss}}$	transmission loss factor between region r and region $r' \neq r$ ($^{\%}/_{\text{miles}}$)
$D_{r,r'}$	distance between region r and region $r' \neq r$ (miles)

Ir	nominal interest rate
If_t	discount factor for year t
$OCC_{i,t}$	overnight capital cost of generator cluster i in year t ($/MW)
$ACC_{i,t}$	annualized capital cost of generator cluster i in year t ($/MW)
$DIC_{i,t}$	discounted investment cost of generator cluster i in year t ($/MW) [2]
CC_i^{m}	capital cost multiplier of generator cluster i (unitless)
LE_i	life extension cost for generator cluster i (fraction of the investment cost of corresponding new generator)
$FOC_{i,t}$	fixed operating cost of generator cluster i ($/MW)
$P_{i,t}^{\text{fuel}}$	price of fuel for generator cluster i in year t ($/MMBtu)
HR_i	heat rate of generator cluster i (MMBtu/MWh)
$Tx_t^{\text{CO}_2}$	carbon tax in year t ($/kg CO_2)
$EF_i^{\text{CO}_2}$	full lifecycle CO_2 emission factor for generator cluster i ($kgCO_2$/MMBtu)
$VOC_{i,t}$	variable O&M cost of generator cluster i ($/MWh)
$RN_t^{\min}$	minimum renewable energy production requirement during year t (fraction of annual energy demand)
PEN_t^{rn}	penalty for not meeting renewable energy quota target during year t ($/MWh)
PEN_t^c	penalty for curtailment during year t ($/MWh)
C_i^{start}	fixed startup cost for generator cluster i ($/MW)
$\hat{ngo}_{i,r,t,k}^{\text{rn}}$	solution from the Forward Pass, iteration k, of the number of operational renewable generators in cluster i of region r at year t, $ngo_{i,r,t}^{\text{rn}}$, which is a fixed parameter for the year $t+1$ (unitless)
$\hat{ngo}_{i,r,t,k}^{\text{th}}$	solution from the Forward Pass, iteration k, of the number of operational thermal generators in cluster i of region r at year t, $ngo_{i,r,t}^{\text{th}}$, which is a fixed parameter for the year $t+1$ (unitless)
$\hat{ngb}_{i,r,t,k}^{\text{rn}}$	solution from the Forward Pass, iteration k, of the number of renewable generators built in cluster i of region r at year t, $ngb_{i,r,t}^{\text{rn}}$, which is a fixed parameter for the year $t+LT_i$ (unitless)
$\hat{ngb}_{i,r,t,k}^{\text{th}}$	solution from the Forward Pass, iteration k, of the number of thermal generators built in cluster i of region r at year t, $ngb_{i,r,t}^{\text{th}}$, which is a fixed parameter for the year $t+LT_i$ (unitless)
$\mu_{i,r,t,k}^{\text{o,rn}}$	multiplier of the linking equality related to the number of operational renewable generators in cluster i of region r at year $t-1$ in iteration k (unitless)
$\mu_{i,r,t,k}^{\text{o,th}}$	multiplier of the linking equality related to the number of operational thermal generators in cluster i of region r at year $t-1$ in iteration k (unitless)
$\mu_{i,r,t,k}^{\text{b,rn}}$	multiplier of the linking equality related to the number of renewable generators built in cluster i of region r at year $t-LT_i$ in iteration k (unitless)
$\mu_{i,r,t,k}^{\text{b,th}}$	multiplier of the linking equality related to the number of thermal generators built in cluster i of region r at year $t-LT_i$ in iteration k (unitless)
$\hat{x}_{t,k}$	fixed solution of x_t in iteration k (concise notation)
$\hat{x}_{t,k}^{\text{o}}$	fixed solution of x_t in iteration k corresponding to $\hat{ngo}_{i,r,t,k}^{\text{rn}}$ and $\hat{ngo}_{i,r,t,k}^{\text{th}}$ (concise notation)
$\hat{x}_{t,k}^{\text{b}}$	fixed solution of x_t in iteration k corresponding to $\hat{ngb}_{i,r,t,k}^{\text{rn}}$ and $\hat{ngb}_{i,r,t,k}^{\text{th}}$ (concise notation)
$\hat{\Phi}_{t,k}$	fixed optimal value for subproblem corresponding to period t in iteration k
$\mu_{t,k}$	Lagrange multiplier (concise notation)
$\mu_{t,k}^{\text{o}}$	Lagrange multiplier corresponding to $\mu_{i,r,t,k}^{\text{o,rn}}$ and $\mu_{i,r,t,k}^{\text{o,th}}$ (concise notation)
$\mu_{t,k}^{\text{b}}$	Lagrange multiplier corresponding to $\mu_{i,r,t,k}^{\text{b,rn}}$ and $\mu_{i,r,t,k}^{\text{b,th}}$ (concise notation)
$\bar{\mu}_{t,k}$	optimal Lagrange multiplier (concise notation)
$step_{t,k}$	stepsize used in the subgradient method (unitless)
ϵ_1, ϵ_2, ϵ_3	tolerances for the decomposition algorithm

Continuous variables

Φ	net present cost[3] throughout the time horizon, including amortized investment cost, operational and environmental cost ($)
Φ_t^{opex}	amortized operating costs in year t ($)

[2] $DIC_{i,t}$ is used in the calculation for the life extension investment cost, which is in terms of a fraction LE_i of the capital cost. Therefore the investment cost for the existing cluster is approximated as being the same as for the potential clusters that have the same or similar generation technology.

[3] All the costs are in 2015 USD.

Φ_t^{capex}	amortized investment costs in year t (\$)
Φ_t^{PEN}	amortized penalty costs in year t (\$)
$p_{i,r,t,d,s}$	power output of generation cluster i in region r during sub-period s of representative day d of year t (MW)
def_t^{rn}	deficit from renewable energy quota target during year t (MWh)
$cu_{r,t,ss,s}$	curtailment slack generation in region r during sub-period s of representative day d of year t (MW)
$p_{r,r',t,d,s}^{\text{flow}}$	power transfer from region r to region $r' \neq r$ during sub-period s of representative day d of year t (MW)
$q_{i,r,t,d,s}^{\text{spin}}$	spinning reserve capacity of generation cluster i in region r during sub-period s of representative day d of year t (MW)
$q_{i,r,t,d,s}^{\text{Qstart}}$	quick-start capacity reserve of generation cluster i in region r during sub-period s of representative day d of year t (MW)
$ngo_{i,r,t}^{\text{rn}}$	number of generators that are operational in cluster $i \in \mathcal{I}_r^{\text{RN}}$ of region r in year t (continuous relaxation)
$ngb_{i,r,t}^{\text{rn}}$	number of generators that are built in cluster $i \in \mathcal{I}_r^{\text{RN}}$ of region r in year t (continuous relaxation)
$ngr_{i,r,t}^{\text{rn}}$	number of generators that retire in cluster $i \in \mathcal{I}_r^{\text{RN}}$ of region r in year t (continuous relaxation)
$nge_{i,r,t}^{\text{rn}}$	number of generators that had their life extended in cluster $i \in \mathcal{I}_r^{\text{RN}}$ of region r in year t (continuous relaxation)
Φ_t	objective function value for subproblem t assuming exact representation of the cost-to-go function (\$)
$\Phi_{t,k}$	objective function value for subproblem t in iteration k (\$)
$\phi_{t,k}$	cost-to-go function (\$)
α_t	expected future year cost, when calculating the cost for year t (\$)
$\Phi_{t,k}^{\text{LP}}$	net present cost of the linear relaxation of the subproblem for year t in iteration k (\$)
$\Phi_{t,k}^{\text{LR}}$	net present cost of the Lagrangean relaxation of the subproblem for year t in iteration k (\$)
$\Phi_{t,k}^{\text{LD}}$	net present cost of the Lagrangean dual of the subproblem for year t in iteration k (\$)
$\Phi_{t,k}^{\text{OP}}$	net present cost of the original MILP subproblem for year t in iteration k (\$)
$ngo_{i,r,t}^{\text{rn,prev}}$	number of generators that are operational in cluster $i \in \mathcal{I}_r^{\text{RN}}$ of region r in year $t-1$ (continuous relaxation)
$ngb_{i,r,t}^{\text{rn,LT}}$	number of generators that are built in cluster $i \in \mathcal{I}_r^{\text{RN}}$ of region r in year $t - LT_i$ (continuous relaxation)
$ngo_{i,r,t}^{\text{th,prev}}$	number of generators that are operational in cluster $i \in \mathcal{I}_r^{\text{TH}}$ of region r in year $t-1$ (continuous relaxation)
$ngb_{i,r,t}^{\text{th,prev}}$	number of generators that are built in cluster $i \in \mathcal{I}_r^{\text{TH}}$ of region r in year $t - LT_i$ (continuous relaxation)
x_t	state (linking) variables in the concise notation
z_t	duplicated state variables in the concise notation
y_t	local variables in the concise notation

Discrete variables

$ngo_{i,r,t}^{\text{th}}$	number of generators that are operational in cluster $i \in \mathcal{I}_r^{\text{TH}}$ of region r in year t (integer variable)
$ngb_{i,r,t}^{\text{th}}$	number of generators that are built in cluster $i \in \mathcal{I}_r^{\text{TH}}$ of region r in year t (integer variable)
$ngr_{i,r,t}^{\text{th}}$	number of generators that retire in cluster $i \in \mathcal{I}_r^{\text{TH}}$ of region r in year t (integer variable)
$nge_{i,r,t}^{\text{th}}$	number of generators that had their life extended in cluster $i \in \mathcal{I}_r^{\text{TH}}$ of region r in year t (integer variable)
$u_{i,r,t,d,s}$	number of thermal generators ON in cluster $i \in \mathcal{I}_r$ of region r during sub-period s of representative day d of year t (integer variable)
$su_{i,r,t,d,s}$	number of generators starting up in cluster i during sub-period s of representative day d in year t (integer variable)
$sd_{i,r,t,d,s}$	number of generators shutting down in cluster i during sub-period s of representative day d in year t (integer variable)

Acronyms

NG	natural gas

ST	steam turbine
CT	gas-fired combustion turbine
CC	combined cycle
CCS	carbon capture and storage
PV	solar photo-voltaic
CSP	concentrated solar panel

Acknowledgments

Authors gratefully acknowledge financial support from ExxonMobil Research and Engineering Company, CAPES Foundation - Ministry of Education of Brazil (Scholarship n° 13241-13-3), and the Center for Advanced Process Decision-making at Carnegie Mellon University.

Appendix A. Calculated parameters

Regarding the parameters used in equations (26)-(28), the discount factor in year t, If_t, is calculated from the interest rate, Ir:

$$If_t = \frac{1}{(1+Ir)^t}$$

and the discounted investment cost $DIC_{i,t}$ is given by:

$$DIC_{i,t} = ACC_{i,t} \cdot \left(\sum_{t' \leq \min(LT_i, T^{\text{remain}})} DF_{t'} \right)$$

where the annualized capital cost $ACC_{i,t}$ is given by:

$$ACC_{i,t} = \frac{OCC_{i,t} \cdot Ir}{1 - \frac{1}{(1+Ir)^{LT_i}}}$$

and the remaining time in the horizon T_t^{remain} is defined by $T_t^{\text{remain}} = T - t + 1$.

References

[1] ABB (2010). *CREZ Reactive Power Compensation Study*. Technical Report. doi:10.1109/PES.2010.5588066.

[2] Ahmed, S. (2016). Blessing of binary in stochastic integer programming. In *XIV International Conference on Stochastic Programming*.

[3] Albadi, M., & El-Saadany, E. (2010). Overview of wind power intermittency impacts on power systems. *Electric Power Systems Research*, *80*, 627 – 632. URL: http://www.sciencedirect.com/science/article/pii/S0378779609002764. doi:http://dx.doi.org/10.1016/j.epsr.2009.10.035.

[4] Alguacil, N., Motto, A. L., & Conejo, A. J. (2003). Transmission expansion planning: a mixed-integer lp approach. *IEEE Transactions on Power Systems*, *18*, 1070–1077. doi:10.1109/TPWRS.2003.814891.

[5] Bahiense, L., Oliveira, G. C., Pereira, M., & Granville, S. (2001). A mixed integer disjunctive model for transmission network expansion. *IEEE Transactions on Power Systems*, *16*, 560–565. doi:10.1109/59.932295.

[6] Bakirtzis, G. A., Biskas, P. N., & Chatziathanasiou, V. (2012). Generation expansion planning by MILP considering mid-term scheduling decisions. *Electric Power Systems Research*, *86*, 98 – 112. URL: http://www.sciencedirect.com/science/article/pii/S0378779611003178. doi:http://dx.doi.org/10.1016/j.epsr.2011.12.008.

[7] Birge, J. R. (1985). Decomposition and partitioning methods for multistage stochastic linear programs. *Operations Research*, *33*, 989–1007. URL: http://dx.doi.org/10.1287/opre.33.5.989. doi:10.1287/opre.33.5.989. arXiv:http://dx.doi.org/10.1287/opre.33.5.989.

[8] Bouffard, F., & Galiana, F. D. (2008). Stochastic security for operations planning with significant wind power generation. *IEEE Transactions on Power Systems*, *23*, 306–316. doi:10.1109/TPWRS.2008.919318.

[9] Cadre, H. L., Papavasiliou, A., & Smeers, Y. (2015). Wind farm portfolio optimization under network capacity constraints. *European Journal of Operational Research*, *247*, 560 – 574. URL: http://www.sciencedirect.com/science/article/pii/S0377221715004920. doi:http://dx.doi.org/10.1016/j.ejor.2015.05.080.

[10] Cerisola, S., Álvaro Baíllo, Fernández-López, J. M., Ramos, A., & Gollmer, R. (2009). Stochastic power generation unit commitment in electricity markets: A novel formulation and a comparison of solution methods. *Operations Research*, *57*, 32–46. URL: http://dx.doi.org/10.1287/opre.1080.0593. doi:10.1287/opre.1080.0593. arXiv:http://dx.doi.org/10.1287/opre.1080.0593.

[11] Conejo, A. J., Baringo, L., Kazempour, S. J., & Sissiqui, A. S. (2016). *Investment in Electricity Generation and Transmission - Decision Making under Uncertainty*. Springer International Publishing. doi:10.1007/978-3-319-29501-5.

[12] Ding, J., & Somani, A. (2010). A long-term investment planning model for mixed energy infrastructure integrated with renewable energy. In *2010 IEEE Green Technologies Conference* (pp. 1–10). doi:10.1109/GREEN.2010.5453785.

[13] Electric Reliability Council of Texas (2011). Long Term Study Task Force Meeting - Generic Database Characteristics REV 1. URL: http://www.ercot.com/calendar/2011/5/3/34168-LTS.

[14] EPRI (2013). *PRISM 2.0: Regional Energy and Economic Model Development and Initial Application; US- REGEN Model Documentation*. Technical Report. URL: https://www.epri.com/#/pages/product/000000003002000128/.

[15] ERCOT (2016). *Report on the Capacity, Demand, and Reserves (CDR) in the ERCOT Region, 2017-2026*. Technical Report. URL: http://www.ercot.com/content/wcm/lists/96607/CapacityDemandandReserveReport-Dec2016.pdf.

[16] ERCOT (2016). Weather. URL: http://www.ercot.com/about/weather.

[17] ERCOT Systems Planning (2014). *Panhandle Renewable Energy Zone (PREZ) Study Report*. Technical Report April. URL: http://www.ercot.com/content/news/presentations/2014/Panhandle%20Renewable%20Energy%20Zone%20Study%20Report.pdf.

[18] Fisher, M. L. (2004). The lagrangian relaxation method for solving integer programming problems. *Manage. Sci.*, *50*, 1861–1871. URL: http://dx.doi.org/10.1287/mnsc.1040.0263. doi:10.1287/mnsc.1040.0263.

[19] Flores-Quiroz, A., Palma-Behnke, R., Zakeri, G., & Moreno, R. (2016). A column generation approach for solving generation expansion planning problems with high renewable energy penetration. *Electric Power Systems Research*, *136*, 232 – 241. URL: http://www.sciencedirect.com/science/article/pii/S0378779616300177. doi:http://dx.doi.org/10.1016/j.epsr.2016.02.011.

[20] Frangioni, A. (2005). About lagrangian methods in integer optimization. *Annals of Operations Research*, *139*, 163–193. URL: http://dx.doi.org/10.1007/s10479-005-3447-9. doi:10.1007/s10479-005-3447-9.

[21] Frank, S., Steponavice, I., & Rebennack, S. (2012). Optimal power flow: a bibliographic survey i. *Energy Systems*, *3*, 221–258. URL: http://dx.doi.org/10.1007/s12667-012-0056-y. doi:10.1007/s12667-012-0056-y.

[22] Frank, S., Steponavice, I., & Rebennack, S. (2012). Optimal power flow: a bibliographic survey ii. *Energy Systems*, *3*, 259–289. URL: http://dx.doi.org/10.1007/s12667-012-0057-x. doi:10.1007/s12667-012-0057-x.

[23] Huang, K., & Ahmed, S. (2009). The value of multistage stochastic programming in capacity planning under uncertainty. *Operations Research*, *57*, 893–904. URL: http://www.jstor.org/stable/25614804.

[24] IBM (2015). *IBM ILOG CPLEX Optimization Studio CPLEX User's Manual*. Technical Report. URL: http://pic.dhe.ibm.com/infocenter/cosinfoc/v12r6/index.jsp?topic=/ilog.odms.ide.help/OPL_Studio/opllangref/topics/opl_langref_scheduling_sequence.html.

[25] Kerl, P. Y., Zhang, W., Moreno-Cruz, J. B., Nenes, A., Realff, M. J., Russell, A. G., Sokol, J., & Thomas, V. M. (2015). New approach for optimal electricity planning and dispatching with hourly time-scale air quality and health considerations. *Proceedings of the National Academy of Sciences*, *112*, 10884–10889. URL: http://www.pnas.org/content/112/35/10884.abstract. doi:10.1073/pnas.1413143112.

[26] Koltsaklis, N. E., & Georgiadis, M. C. (2015). A multi-period, multi-regional generation expansion planning model incorporating unit commitment constraints. *Applied Energy*, *158*, 310 – 331. URL: http://www.sciencedirect.com/science/article/pii/S0306261915009873. doi:http://dx.doi.org/10.1016/j.apenergy.2015.08.054.

[27] Krishnan, V., Ho, J., Hobbs, B. F., Liu, A. L., McCalley, J. D., Shahidehpour, M., & Zheng, Q. P. (2016). Co-optimization of electricity transmission and generation resources for planning and policy analysis: review of concepts and modeling approaches. *Energy Systems*, *7*, 297–332. doi:10.1007/s12667-015-0158-4.

[28] Lannoye, E., Flynn, D., & O'Malley, M. (2011). The role of power system flexibility in generation planning. In *2011 IEEE Power and Energy Society General Meeting* (pp. 1–6). doi:10.1109/PES.2011.6039009.

[29] Latorre, G., Cruz, R. D., Areiza, J. M., & Villegas, A. (2003). Classification of publications and models on transmission expansion planning. *IEEE Transactions on Power Systems*, *18*, 938–946. doi:10.1109/TPWRS.2003.811168.

[30] Liu, Y., Sioshansi, R., & Conejo, A. (2017). Multistage stochastic investment planning with multiscale representation of uncertainties and decisions. *IEEE Transactions on Power Systems*, *PP*, 1–1. doi:10.1109/TPWRS.2017.2694612.

[31] Macdonald, A. E., Clack, C. T. M., Alexander, A., Dunbar, A., Wilczak, J., & Xie, Y. (2016). Future cost-competitive electricity systems and their impact on us co2 emissions. *Nature Climate Change*, *6*, 526–531. doi:10.1038/nclimate2921.

[32] Mai, T., Drury, E., Eurek, K., Bodington, N., Lopez, A., & Perry, A. (2013). *Resource Planning Model: An Integrated Resource Planning and Dispatch Tool for Regional Electric Systems*. Technical Report January. URL: https://www.nrel.gov/docs/fy13osti/56723.pdf.

[33] Nahmmacher, P., Schmid, E., Hirth, L., & Knopf, B. (2016). Carpe diem: A novel approach to select representative days for long-term power system modeling. *Energy*, *112*, 430 – 442. URL: http://www.sciencedirect.com/science/article/pii/S0360544216308556. doi:http://doi.org/10.1016/j.energy.2016.06.081.

[34] Newell, S. A., Spees, K., Pfeifenberger, J. P., Karkatsou, I., Wintermantel, N., & Carden, K. (2014). *Estimating the economically optimal reserve margin in ERCOT*. Technical Report. URL: http://www.brattle.com/system/news/pdfs/000/000/613/original/Estimating_the_Economically_Optimal_Reserve_Margin_in_ERCOT.pdf?1391445083 the Brattle Group, Cambridge MA.

[35] North American Electric Reliability Corporation (2009). *Accommodating High Levels of Variable Generation*. Technical Report April. URL: http://www.nerc.com/files/ivgtf_report_041609.pdf.

[36] Northwest Power and Conservation Council (2010). *Sixth Northwest Conservation and Electric Power Plan*. Technical Report February. URL: https://www.nwcouncil.org/energy/powerplan/6/plan/.

[37] NREL (National Renewable Energy Laboratory) (2016). 2016 Annual Technology Baseline (ATB) Spreadsheet. URL: http://www.nrel.gov/analysis/data_tech_baseline_legacy.html.

[38] O'Neill, R. P., Krall, E. A., Hedman, K. W., & Oren, S. S. (2013). A model and approach to the challenge posed by optimal power systems planning. *Mathematical Programming*, *140*, 239–266. URL: http://dx.doi.org/10.1007/s10107-013-0695-3. doi:10.1007/s10107-013-0695-3.

[39] Palmintier, B., & Webster, M. (2011). Impact of unit commitment constraints on generation expansion planning with renewables. In *2011 IEEE Power and Energy Society General Meeting* (pp. 1–7). doi:10.1109/PES.2011.6038963.

[40] Palmintier, B., & Webster, M. (2014). Heterogeneous unit clustering for efficient operational flexibility modeling. In *2014 IEEE PES General Meeting | Conference Exposition* (pp. 1–1). doi:10.1109/PESGM.2014.6939001.

[41] Papavasiliou, A., Oren, S. S., & O'Neill, R. P. (2011). Reserve requirements for wind power integration: A scenario-based stochastic programming framework. *IEEE Transactions on Power Systems*, *26*, 2197–2206. doi:10.1109/TPWRS.2011.2121095.

[42] Park, H., & Baldick, R. (2013). Transmission planning under uncertainties of wind and load: Sequential approximation approach. *IEEE Transactions on Power Systems*, *28*, 2395–2402. doi:10.1109/TPWRS.2013.2251481.

[43] Pereira, M. V. F., & Pinto, L. M. V. G. (1991). Multi-stage stochastic optimization applied to energy planning. *Mathematical Programming*, *52*, 359–375. URL: https://doi.org/10.1007/BF01582895. doi:10.1007/BF01582895.

[44] Pina, A., Silva, C., & Ferrão, P. (2011). Modeling hourly electricity dynamics for policy making in long-term scenarios. *Energy Policy*, *39*, 4692 – 4702. URL: http://www.sciencedirect.com/science/article/pii/S0301421511005180. doi:http://doi.org/10.1016/j.enpol.2011.06.062.

[45] Pina, A., Silva, C. A., & Ferrão, P. (2013). High-resolution modeling framework for planning electricity systems with high penetration of renewables. *Applied Energy*, *112*, 215 – 223. URL: http://www.sciencedirect.com/science/article/pii/S030626191300487X. doi:http://dx.doi.org/10.1016/j.apenergy.2013.05.074.

[46] Poncelet, K., Delarue, E., Duerinck, J., Six, D., & D'haeseleer, W. (2014). The importance of integrating the variability of renewables in long-term energy planning models. *TME Working Paper - Energy and Environment*, (pp. 1–18). URL: https://www.mech.kuleuven.be/en/tme/research/energy_environment/Pdf/wp-importance.pdf.

[47] Potomac Economics (2016). *2015 State of the Market Report for the ERCOT Whole Electricity Markets*. Technical Report June. URL: http://www.puc.texas.gov/industry/electric/reports/ERCOT_annual_reports/2015annualreport.pdf.

[48] Pozo, D., Sauma, E. E., & Contreras, J. (2013). A three-level static milp model for generation and transmission expansion planning. *IEEE Transactions on Power Systems*, *28*, 202–210. doi:10.1109/TPWRS.2012.2204073.

[49] Rahmaniani, R., Crainic, T. G., Gendreau, M., & Rei, W. (2016). The Benders decomposition algorithm: A literature review. *European Journal of Operational Research*, *259*, 801–817. URL: http://www.sciencedirect.com/science/article/pii/S0377221716310244. doi:http://dx.doi.org/10.1016/j.ejor.2016.12.005.

[50] Roh, J. H., Shahidehpour, M., & Fu, Y. (2007). Market-based coordination of transmission and generation capacity planning. *IEEE Transactions on Power Systems*, *22*, 1406–1419. doi:10.1109/TPWRS.2007.907894.

[51] Sahinidis, N., & Grossmann, I. (1991). Convergence properties of generalized benders decomposition. *Computers and Chemical Engineering*, *15*, 481 – 491. URL: http://www.sciencedirect.com/science/article/pii/009813549185027R. doi:http://dx.doi.org/10.1016/0098-1354(91)85027-R.

[52] Short, W., Sullivan, P., Mai, T., Mowers, M., Uriarte, C., Blair, N., Heimiller, D., & Martinez, A. (2011). *Regional Energy Deployment System (ReEDS)*. Technical Report December. URL: https://www.nrel.gov/analysis/reeds/pdfs/reeds_documentation.pdf. doi:NREL/TP-6A20-46534.

[53] Shortt, A., & O'Malley, M. (2010). Impact of variable generation in generation resource planning models. In *IEEE PES General Meeting* (pp. 1–6). doi:10.1109/PES.2010.5589461.

[54] Steeger, G., & Rebennack, S. (2017). Dynamic convexification within nested benders decomposition using lagrangian relaxation: An application to the strategic bidding problem. *European Journal of Operational Research*, *257*, 669 – 686. URL: http://www.sciencedirect.com/science/article/pii/S037722171630621X. doi:http://dx.doi.org/10.1016/j.ejor.2016.08.006.

[55] Thome, F., Pereira, M., Granville, S., & Fampa, M. (2013). Non-Convexities Representation on Hydrothermal Operation Planning using SDDP, . URL: http://www.psr-inc.com/publications.

[56] U.S. Energy Information Administration (EIA) (2016). *Annual Energy Outlook 2016*. Technical Report. URL: https://www.eia.gov/outlooks/aeo/pdf/0383(2016).pdf.

[57] U.S. Environmental Protection Agency Clean Air Markets Division (2013). *Documentation for EPA Base Case v.5 .13 Using the Integrated Planning Model*. Technical Report. URL: https://www.epa.gov/sites/production/files/2015-07/documents/documentation_for_epa_base_case_v.5.13_using_the_integrated_planning_model.pdf.

[58] Velásquez Bermúdez, J. M. (2002). Gddp: Generalized dual dynamic programming theory. *Annals of Operations Research*, *117*, 21–31. URL: https://doi.org/10.1023/A:1021557003554. doi:10.1023/A:1021557003554.

[59] Westinghouse Electric Corporation (1984). *The Westinghouse Pressurized Water Reactor Nuclear Power Plant*. Technical Report. URL: http://www4.ncsu.edu/~doster/NE405/Manuals/PWR_Manual.pdf.

[60] Zhu, J., & yuen Chow, M. (1997). A review of emerging techniques on generation expansion planning. *IEEE Transactions on Power Systems*, *12*, 1722–1728. doi:10.1109/59.627882.

[61] Zou, J., Ahmed, S., & Sun, X. A. (2016). Stochastic Dual Dynamic Integer Programming. *Submitted for publication*, . URL: http://www.optimization-online.org/DB{_}HTML/2016/05/5436.html.

[62] Zou, J., Ahmed, S., & Sun, X. A. (2017). Multistage Stochastic Unit Commitment Using Stochastic Dual Dynamic Integer Programming. *Submitted for publication*, . URL: http://www.optimization-online.org/DB_HTML/2017/05/6003.html.