

Prescribing net demand for two-stage electricity generation scheduling

J.M. Morales^{*}, M.A. Muñoz, S. Pineda

OASYS Group, University of Malaga, Ada Byron Research Building, Arquitecto Francisco Peñalosa St., 18, 29010, Malaga, Spain

ARTICLE INFO

Keywords:

Smart predict
Net demand prescription
Two-stage power generation scheduling
Data-driven optimization

ABSTRACT

We consider a two-stage generation scheduling problem comprising a forward dispatch and a real-time re-dispatch. The former must be conducted facing an uncertain net demand that includes non-dispatchable electricity consumption and renewable power generation. The latter copes with the plausible deviations with respect to the forward schedule by making use of balancing power during the actual operation of the system. Standard industry practice deals with the uncertain net demand in the forward stage by replacing it with a good estimate of its conditional expectation (usually referred to as a *point forecast*), so as to minimize the need for balancing power in real time. However, it is well known that the cost structure of a power system is highly asymmetric and dependent on its operating point, with the result that minimizing the amount of power imbalances is not necessarily aligned with minimizing operating costs. In this paper, we propose a bilevel program to construct, from the available historical data, a *prescription* of the net demand that does account for the power system's cost asymmetry. Furthermore, to accommodate the strong dependence of this cost on the power system's operating point, we use clustering to tailor the proposed prescription to the foreseen net-demand regime. By way of an illustrative example and a more realistic case study based on the European power system, we show that our approach leads to substantial cost savings compared to the customary way of doing.

1. Introduction

Many decision-making processes under uncertainty can be modeled by optimization problems where some of the input parameters are not perfectly known. The field of Optimization under Uncertainty focuses on developing tools to tackle these problems depending on the knowledge of those parameters that the decision maker actually has. For example, if these parameters can be modeled reasonably well as random variables following certain probability distributions, then the decision maker should probably resort to stochastic programming techniques [1]. In contrast, if all the decision maker knows about said parameters is their range of variation or support, then she should rather opt for robust optimization methods instead [2].

In the realm of power system operations, there is a vast literature on operations research methods, models, and algorithms for power dispatch that rely on stochastic programming or robust optimization or hybrids of both. The richness of this literature makes it materially impossible and pointless to embrace it all in this paper. Instead, we refer the reader to monograph Morales et al. [3] and references therein for examples of power systems operating problems based on stochastic programming, to the seminal work Bertsimas et al. [4] on the application of robust optimization for unit commitment and generation scheduling,

and to the recent contribution Dvorkin [5] on a distributionally robust chance-constrained electricity market.

Despite the firm and promising advances in Optimization under Uncertainty, still one of the most widely extended practices in decision making is to replace the unknown parameter with a sensible value or estimate, some sort of “the most likely value” that the parameter can take on. A natural candidate to play that role is the expected value of the parameter. Thus, the decision maker can, in addition, exploit all the powerful tools that the disciplines of Statistics, Forecasting and Machine Learning have developed for decades to estimate that expected value *conditional* on all the information the decision maker has available at the moment the decision must be made. The adherence of the power sector to this strategy is particularly notorious, essentially because it is argued to be simpler, more transparent, computationally cheaper and easily accepted by the different stakeholders (see, e.g., Morales et al. [6], Wang and Hobbs [7], Morales and Pineda [8], Kazempour et al. [9] for further details on this issue). However, operations researchers have repeatedly shown that this strategy results in suboptimal decisions in general, because the conditional expected value of the parameter ignores the impact of the parameter's uncertainty on the decision's value [1].

^{*} Corresponding author.

E-mail addresses: juan.morales@uma.es (J.M. Morales), miguelangeljmd@uma.es (M.A. Muñoz), spineda@uma.es (S. Pineda).

<https://doi.org/10.1016/j.orp.2023.100268>

Received 5 September 2022; Received in revised form 9 December 2022; Accepted 23 January 2023

Available online 25 January 2023

2214-7160/© 2023 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Against this background, research efforts have been recently placed on finding a compromise solution. Intuitively, the idea is still to replace the uncertain parameter but with a point estimate (generally different from the parameter's conditional expectation) that is purposely computed to result in the optimal or a nearly optimal decision in view of the parameter's uncertainty. This alternative point estimate is usually called a *prescription* and the term *smart predict* has been recently coined to refer to this midway solution strategy. In this line, we find a number of research works, e.g., Donti et al. [10], Muñoz et al. [11], Elmachtoub and Grigas [12], El Balghithi et al. [13]. In particular, the authors in Donti et al. [10] propose a heuristic gradient-based procedure to produce estimates of uncertain parameters in optimization problems based on the objective of the task for which these estimates will be used. In Muñoz et al. [11], instead, they introduce a bilevel programming framework to the same end. In Elmachtoub and Grigas [12], they deal with linear programs with an uncertain cost vector, for which they develop a tailored convex loss function to compute an estimate of the cost vector that accounts for the underlying linear optimization problem. Finally, the work in El Balghithi et al. [13] is a continuation of that in Elmachtoub and Grigas [12] (first released as an arXiv preprint in 2017), where the authors provide bounds on how well a certain method to predict the cost vector from training data generalizes out of sample.

In the field of power systems, it has also been shown that, by smartly tuning the input parameters of current operational and procedures, these can mimic the performance of their stochastic-programming-based counterparts to a large extent. For instance, in Morales et al. [6], they propose a bilevel programming model to compute the amount of (uncertain) renewable power generation that must be considered in a forward electricity market to maximize the short-run market efficiency. In the same vein, the authors in Dvorkin et al. [14] show that, by means of bilevel programming too, the reserve requirements in an European-style two-stage electricity market can be set so that the market can be almost as cost-efficient as the ideal two-stage electricity market run by a full stochastic programming approach. Our proposal is in the same spirit, but we make use of bilevel programming not to directly prescribe renewable production or reserve capacity, but to determine straightforward *rules* to improve the value of the (net-demand) forecasts currently employed by system operators. In any case, the works of Morales et al. [6], Dvorkin et al. [14] reveal that the power sector can highly benefit from the aforementioned smart-predict strategy. Actually, in Donti et al. [10], they apply it to power generation and grid-scale electricity battery operation, and in Muñoz et al. [11] to the offering problem of a thermal power producer competing strategically in an electricity market. In Carriere and Kariniotakis [15], Muñoz et al. [16], they focus instead on renewable energy producers, for which they propose different smart-predict strategies for energy trading. Lastly, the authors in Garcia et al. [17] use a bilevel programming framework similar to that proposed in Muñoz et al. [11] whereby they train several autoregressive models to estimate the uncertain demand and the size of the energy reserves in a joint reserve allocation and energy dispatch problem.

Within this context, and given a two-stage power scheduling problem, our contributions are the following:

- We propose a bilevel program to learn the value of the net demand (i.e., that obtained by subtracting the weather-driven renewable power production from the non-dispatchable power load) that the scheduling problem must consider so that the power system operating costs are minimized in expectation. This value (that is, the prescription) is, in general, different from the conditional mean of the net demand that is currently employed in standard industry practice.
- We show that our method can be easily used to upgrade *any* given point forecast of the net demand into a prescription that accounts for the power system's cost asymmetry.

- Motivated by the fact that a power system usually features distinct net-demand regimes, we introduce a data clustering and partitioning strategy that, on the one hand, increases the performance of the prescription of the net demand that our approach produces (by making it dependent on the foreseen net-demand regime), and, on the other, substantially decreases the computational effort to solve the associated bilevel estimation problem.
- We evaluate the economic benefits that our approach achieves through an out-of-sample test on a stylized version of the European power system that makes use of real data, in particular, of actual and day-ahead predicted net-demand values downloaded from the ENTSO-e Transparency Platform [18].

The rest of this paper is organized as follows. Section 2 describes the two-stage power generation scheduling problem we consider throughout our paper, motivates the ultimate goal of our work, and formulates the generic bilevel program we use to construct, from the available historical data, prescriptions of the net demand intended to minimize the expected system operation costs. In Section 3, we resort to various simplifying assumptions to guarantee that the globally optimal solution of the bilevel program can be obtained by solving a mixed-integer linear program (MILP). The potential of the net-demand prescriptions provided by this MILP is then illustrated, discussed and justified using a small power system in Section 4. In Section 5, we introduce a procedure based on data clustering and partitioning to enhance the value of our prescription and to speed up the solution of the mixed-integer prescription problem. A case study based on the European power system is used in Section 6 to investigate the benefits of our approach on real data. Finally, Section 7 concludes the paper with some final remarks.

2. Problem description

We consider a two-stage generation scheduling problem consisting of a forward power dispatch and a real-time balancing phase. The forward dispatch is decided some time prior to the actual delivery of energy, for instance, from 15 min to 36 h in advance. The forward dispatch is required to decide the output levels of inflexible generating units (e.g. nuclear-based power plants) that need advance planning due to technical limits such as ramping constraints or minimum times. The real-time balancing phase processes the energy imbalances with respect to the forward production schedule. The real-time balancing rescheduling aims at adjusting the output of flexible generating units (such as gas-based power plants) that can quickly deviate from the forward schedule to ensure the equilibrium between electricity production and consumption. More details about the two-stage generation scheduling problem used in this paper can be found in Morales et al. [3]

In our framework, the forward stage first determines the power dispatch that minimizes the anticipated electricity production costs by solving

$$\min_{\mathbf{z}^F} f^{\text{obj}}(\mathbf{z}^F, \hat{\mathbf{y}}) \quad (1a)$$

$$\text{s.t. } f^{\text{gen}}(\mathbf{z}^F, \hat{\mathbf{y}}) \leq 0 \quad (1b)$$

$$f^{\text{net}}(\mathbf{z}^F, \hat{\mathbf{y}}) \leq 0, \quad (1c)$$

where \mathbf{z}^F denotes the scheduling decisions (such as the dispatch of thermal generating units and the consumption of dispatchable loads) and $\hat{\mathbf{y}}$ represents a point or single-value estimate of the uncertain parameters (in our case, the net demand, given as the difference between the non-dispatchable load and the weather-driven renewable power generation). Objective function (1a) determines the forward system operating costs, while the generic constraints (1b) and (1c) enforce the technical limitations of the generating units and the network, respectively.

Since the net demand is uncertain, the scheduling decisions that results from the forward problem (1) are to be adjusted during the

real-time operation of the power system to accommodate the *eventual* net-demand value that is realized. This adjustment is conducted through the following generic real-time balancing problem:

$$\min_{\mathbf{z}^B} g^{\text{obj}}(\mathbf{z}^F, \mathbf{z}^B, \mathbf{y}) \quad (2a)$$

$$\text{s.t. } g^{\text{gen}}(\mathbf{z}^F, \mathbf{z}^B, \mathbf{y}) \leq 0 \quad (2b)$$

$$g^{\text{net}}(\mathbf{z}^F, \mathbf{z}^B, \mathbf{y}) \leq 0, \quad (2c)$$

where \mathbf{z}^B symbolizes the adjustment variables in relation to the forward decisions \mathbf{z}^F —here treated as a known and fixed input vector coming from (1)—and \mathbf{y} represents the eventually realized value of the uncertain parameters. The objective function (2a) includes both the forward dispatch cost (which, at this point, is known and constant and, therefore, could be removed from the minimization) and the imbalance cost caused by unexpected net-load variations that occur very close to the real-time operation of the system. Similarly to problem (1), the constraints (2b) and (2c) enforce the technical limitations of the generating units and the network, respectively. As customary, the objective function and constraints of models (1) and (2) are well known and properly defined.

Morales et al. [6] show that the value $\hat{\mathbf{y}}$ of the net demand that is used in the forward scheduling problem (1) may have a major impact on the subsequent balancing costs (2a). Current practice, however, is content with a simple and direct solution, which is to take $\hat{\mathbf{y}}$ as an estimate of the expectation of the net demand \mathbf{y} conditional on all the information at the forecaster's disposal. This information is generally known as *context*. Yet, this expectation is oblivious to the minimization of the balancing costs that drives the real-time balancing problem (2) and therefore, may turn out to be highly suboptimal. In other words, the conditional expectation overlooks the typical asymmetries affecting the minimization of the balancing costs, such as the fact that the distribution of the renewable generation is often skewed and that the costs of supplying upward and downward balancing energy are usually different.

Instead of employing the conditional expectation of the uncertain parameters \mathbf{y} as the estimate $\hat{\mathbf{y}}$ used in the forward scheduling problem (1), we propose a regression procedure that provides an *alternative value* for $\hat{\mathbf{y}}$ that explicitly accounts for the potential impact of the uncertain parameters on the subsequent real-time balancing problem (2). This alternative value is what is called a *prescription* and a procedure to obtain it runs as follows. Suppose we have a sample of N data points expressed in the form $\{(\mathbf{x}_i, \mathbf{y}_i)\}_{i \in \mathcal{N}} := \{(\mathbf{x}_1, \mathbf{y}_1), \dots, (\mathbf{x}_i, \mathbf{y}_i), \dots, (\mathbf{x}_N, \mathbf{y}_N)\}$, where $\mathbf{x} \in \mathbb{R}^p$ is a vector of features or covariates making up the *context* and $\mathbf{y} \in \mathbb{R}^m$ is a vector of uncertain parameters. Our objective is to utilize said sample to infer a functional relation $\hat{\mathbf{y}} = h(\mathbf{x})$ with $h: \mathbb{R}^p \rightarrow \mathbb{R}^m$ such that, given the context \mathbf{x} , the provided prescription $\hat{\mathbf{y}}$ is trained to deliver the minimum total system costs in expectation when inserted into the forward scheduling problem (1).

Now suppose that the prescriptive function h is parameterized on a coefficient vector \mathbf{q} of appropriate dimension, we propose to train $h_{\mathbf{q}}(\cdot)$, i.e., to estimate \mathbf{q} , by way of the following bilevel optimization problem:

$$\min_{\mathbf{q}, \mathbf{z}_i^F, \mathbf{z}_i^B} \frac{1}{|\mathcal{N}|} \sum_{i \in \mathcal{N}} g^{\text{obj}}(\mathbf{z}_i^F, \mathbf{z}_i^B, \mathbf{y}_i) \quad (3a)$$

$$\text{s.t. } g^{\text{gen}}(\mathbf{z}_i^F, \mathbf{z}_i^B, \mathbf{y}_i) \leq 0, \quad \forall i \in \mathcal{N} \quad (3b)$$

$$g^{\text{net}}(\mathbf{z}_i^F, \mathbf{z}_i^B, \mathbf{y}_i) \leq 0, \quad \forall i \in \mathcal{N} \quad (3c)$$

$$\mathbf{z}_i^F \in \{\arg \min_{\mathbf{z}} f^{\text{obj}}(\mathbf{z}, h_{\mathbf{q}}(\mathbf{x}_i))\} \quad (3d)$$

$$\text{s.t. } f^{\text{gen}}(\mathbf{z}, h_{\mathbf{q}}(\mathbf{x}_i)) \leq 0 \quad (3e)$$

$$f^{\text{net}}(\mathbf{z}, h_{\mathbf{q}}(\mathbf{x}_i)) \leq 0, \quad \forall i \in \mathcal{N}. \quad (3f)$$

Essentially, the bilevel program (3) seeks the \mathbf{q} -parameterized prescriptive function $h_{\mathbf{q}}(\cdot)$ that minimizes the empirical expectation of the total dispatch and re-dispatch costs over the data set $\{(\mathbf{x}_i, \mathbf{y}_i)\}_{i \in \mathcal{N}}$.

For this purpose, it replicates, per data point $\{(\mathbf{x}_i, \mathbf{y}_i)\}$, the sequential two-stage scheduling process consisting of the forward power dispatch (3d)–(3f) and the subsequent balancing re-dispatch (3a)–(3c). From a statistical point of view, problem (3) takes the form of an *empirical risk minimization problem* and as such, is amenable to *regularization* and *robustification* [19], admitting strategies for *feature selection* too.

Depending on the nature of variables $\mathbf{z}^F, \mathbf{z}^B$ and functions $f^{\text{obj}}, f^{\text{gen}}, f^{\text{net}}, g^{\text{obj}}, g^{\text{gen}}, g^{\text{net}}$, and $h_{\mathbf{q}}$, the difficulty of solving the bilevel optimization problem (3) can vary significantly. For instance, if the forward scheduling decisions \mathbf{z}^F include binary variables representing the on/off status of thermal generating units, then computing the global optimal solution of the bilevel problem is tremendously challenging as discussed in Fanghanel and Dempe [20]. Likewise, even in the case that all scheduling decisions \mathbf{z}^F are continuous, the lower-level problem (3d)–(3f) can be non-convex if constraints g^{net} account for the nonlinear AC power flow equations. In that case too, computing the global optimal solution of (3) is also a very challenging task. If all variables are continuous and all functions affine, then the bilevel optimization problem (3) can be solved by replacing the lower-level problem with its equivalent KKT optimality conditions [11]. Said strategy, however, requires additional binary variables and large enough constants that increase the numerical instability and the computational burden of the resulting single-level optimization problem. Besides, as discussed in Pineda and Morales [21], Kleinert et al. [22], existing methods to validate these upper bounds may fail and lead to suboptimal solutions.

Given that the development of methods to solve a generic bilevel program like (3) is outside the scope of this paper, in the next section we make use of some simplifying assumptions that allow us to efficiently solve the bilevel problem (3) to global optimality using commercially available software for mixed-integer programming. In practice, we may not need to solve the estimation problem (3) to global optimality. In effect, finding a good feasible solution to this problem may be enough to reap the bulk of the benefits of our approach. This enables the use of heuristics to solve problem (3).

We also remark that some system operators are indeed aware of the importance of scheduling generation taking into account the subsequent real-time operation of the power system. For example, the California Independent System Operator (CAISO) *manually modifies the load forecast*, anticipating the ramping capacity needs based on the operator's experience [23]. This is explicitly explained in California Independent System Operator [24] as follows: "Operators in the ISO and energy imbalance market can manually modify load forecasts used in the market through load adjustments (...) to increase ramping capacity within the ISO by (...) committing additional units". Our approach is not only consistent with this *modus operandi*, but also supports it with a scientifically grounded methodology to modify the load forecast.

Once the bilevel problem (3) is solved and the prescriptive function $h_{\mathbf{q}}$ is obtained, the forward dispatch for an unseen value of the context \mathbf{x} can be easily determined by solving the following deterministic optimization problem:

$$\min_{\mathbf{z}^F} f^{\text{obj}}(\mathbf{z}^F, h_{\mathbf{q}}(\mathbf{x})) \quad (4a)$$

$$\text{s.t. } f^{\text{gen}}(\mathbf{z}^F, h_{\mathbf{q}}(\mathbf{x})) \leq 0 \quad (4b)$$

$$f^{\text{net}}(\mathbf{z}^F, h_{\mathbf{q}}(\mathbf{x})) \leq 0. \quad (4c)$$

It is important to realize that the strategy we propose is completely in contrast with those other approaches that make use of stochastic programming to *simultaneously* decide the forward scheduling decisions and the real-time adjustments, see, e.g., Aghaei et al. [25], Pritchard et al. [26], Zavala et al. [27]. Indeed, our approach builds upon the fact that, in practice, the forward dispatch and the subsequent balancing re-dispatch are *sequentially* optimized. The next section elaborates on the simplifying hypotheses we make to transform the bilevel program (3) into a single-level mixed-integer optimization problem.

3. Contextual merit-order dispatch

For the remainder of this article, we assume the following.

- Forward dispatch decisions and real-time balancing actions are all modeled by continuous variables, that is, on/off binary variables are not considered.
- The inter-temporal constraints of power production portfolios, such as ramping limits and minimum-up and -down times, are not explicitly accounted for in the generation scheduling problem.
- Network constraints are only taken care of in the real-time stage using a pipeline representation of the transmission network. That is, constraints (3f) are removed from the bilevel optimization problem (3).
- All power loads are assumed to be non-dispatchable and therefore, the objective function of the two-stage scheduling problem boils down to the minimization of the generating cost of dispatchable units.

With these assumptions in place, the forward scheduling problem (1) is simplified to a merit-order dispatch whereby generators are scheduled based on their marginal production costs, as explained later. Furthermore, with the pipeline representation of the transmission network, our setup becomes more aligned with European practices, in keeping with the case study we present in Section 6. Likewise, we consider non-dispatchable loads only just to simplify the subsequent exposition of our approach. Finally, by way of the rest of the assumptions, the bilevel model (3) becomes simpler and computationally manageable.

Nevertheless, any of these assumptions could also be dropped at the expense of increasing the complexity of this model as discussed at the end of Section 2. For instance, if on/off commitment decisions of the generating units are considered, solving the bilevel model (3) requires specific solution methodologies depending on the application. If all decisions are continuous but additional intertemporal or network constraints are included in the day-ahead stage, then the single-level reformulation of model (3) requires dual variables of the lower-level constraints and large enough constants, which increases its computational burden and endangers optimality guarantees. Finally, the use of a linearized (e.g., DC) power flow model in the real-time operational stage (2) of the generation scheduling problem would not have any qualitative impact on the computational complexity of the estimation problem (3).

All in all, the particular forward generation scheduling problem we consider in this paper is given as follows:

$$\min_{p_g, g \in G} \sum_{g \in G} C_g p_g \quad (5a)$$

$$\text{s.t.} \quad \sum_{g \in G} p_g = \hat{L} \quad (5b)$$

$$0 \leq p_g \leq \bar{P}_g, \quad \forall g \in G, \quad (5c)$$

where $p_g, \bar{P}_g, C_g \in \mathbb{R}^+$ and $G \subseteq \mathbb{N}$ is the set of generation units (it can also represent generation blocks with different cost). Each unit g has associated a production level p_g and a marginal cost C_g , Eq. (5b) enforces the aggregate power balance, with parameter $\hat{L} \in \mathbb{R}^+$ representing a point or single-value estimate of the total net demand $L \in \mathbb{R}^+$ in the system, which is unknown at the moment the forward scheduling is determined and thus, is to be treated as a random variable. Lastly, Eq. (5c) sets the capacity of each generation unit.

The linear program (5) stands for an economic dispatch problem whereby generation units are dispatched following a cost-merit order, meaning that units g with a lower cost C_g are dispatched first. This is illustrated in Fig. 1 for four units with capacities $\bar{P}_1, \bar{P}_2, \bar{P}_3, \bar{P}_4$ and marginal costs C_1, C_2, C_3, C_4 , respectively. The vertical solid line represents the total net demand \hat{L} and the shadow area indicates the optimal dispatch computed by model (5). Since units 1 and 2 are the

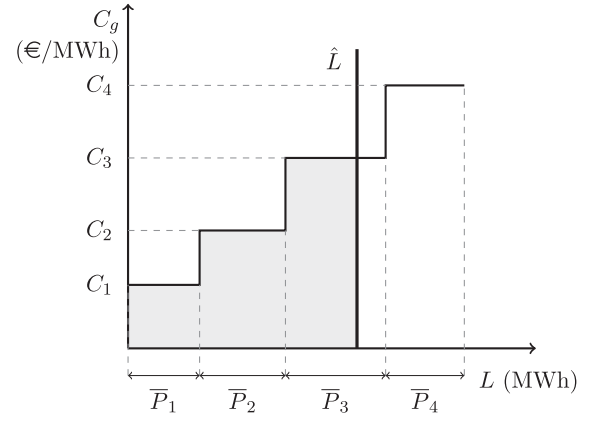


Fig. 1. Illustration of a cost-merit order dispatch. The vertical solid line represents the total net demand \hat{L} and the shadow area represents the optimal dispatch computed by model (5).

cheapest ones, their dispatch is set at their maximum capacity. Unit 3 is partially dispatched until the demand is fulfilled. Unit 4 is the most expensive one and therefore, is left out of the demand supply. To ease the discussion that follows, hereinafter we consider that the units in the set G are ordered such that $g < g'$ if and only if $C_g < C_{g'}$. Hence, if we denote the optimal solution to (5) as $\{p_g^*\}_{g \in G}$, it holds that $p_g^* > 0$ implies that $p_g^* = \bar{P}_g$, whenever $g < g'$.

Since the system net demand is uncertain, the power dispatch that results from the forward problem (5) is to be adjusted during the real-time operation of the power system to satisfy the actual net demand. The aim of the real-time balancing problem is to correct the imbalance of the system in a cost-efficient manner. To this end, we consider a pipeline model where $G(b)$ and $D(b)$ represent the set of generation units and loads that are connected to node b , in that order. With some abuse of notation, let $(L_{di} \in \mathbb{R})_{d \in D(b)}$ be a certain realization $i \in \mathcal{N}$ of the net load $d \in D(b)$ connected to node b of the system. We also define $o(l)$ and $e(l)$ as the origin and ending nodes of line l , respectively. Thus, $\{l : o(l) = b\}$ and $\{l : e(l) = b\}$ represent the subset of lines that start or end at node b , in that order. Once introduced this notation, the real-time balancing problem under consideration renders

$$\min_{\Xi} \sum_{g=1}^G (C_g^u r_g^u - C_g^d r_g^d) \quad (6a)$$

$$\text{s.t.} \quad 0 \leq p_g^* + r_g^u - r_g^d \leq \bar{P}_g, \quad \forall g \in G \quad (6b)$$

$$0 \leq r_g^u \leq R_g^u, \quad \forall g \in G \quad (6c)$$

$$0 \leq r_g^d \leq R_g^d, \quad \forall g \in G \quad (6d)$$

$$\sum_{g \in G(b)} (p_g^* + r_g^u - r_g^d) = \sum_{d \in D(b)} L_{di} + \sum_{l: o(l)=b} f_l - \sum_{l: e(l)=b} f_l, \quad \forall b \in B \quad (6e)$$

$$|f_l| \leq \bar{F}_l, \quad \forall l \in A, \quad (6f)$$

where $\Xi := \{r_g^u, r_g^d \in \mathbb{R}^+, g \in G, f_l \in \mathbb{R}, l \in A\}$ is the set of decision variables and $p_g^*, R_g^u, R_g^d, C_g^u, C_g^d, \bar{P}_g, \bar{F}_l \in \mathbb{R}^+$ and $L_{di} \in \mathbb{R}$ are known parameters.

The power output of each flexible unit g may be increased by an amount r_g^u , based on the marginal cost for upward balancing energy C_g^u , or decreased by an amount r_g^d of downward balancing energy, which entails a marginal benefit (linked to fuel-cost savings) of C_g^d . These actions are driven by the nodal power balance equation (6e) and the minimization of the total balancing costs (6a). Naturally, the amount of balancing energy provided from each generation unit g , either upward or downward, must be such that the eventual power output from that unit (taking into account the forward optimal schedule p_g^*) is positive and lower than its capacity \bar{P}_g , as stated in Eq. (6b).

Moreover, constraints (6c) and (6d) limit the amount of up- and down-balancing power that can be deployed from each generation unit to R_g^u and R_g^d , which are indicative of how flexible the underlying asset actually is. Finally, line capacity limits are imposed by (6f), with f_l being the power flow through line l and \bar{F}_l the capacity of the line.

Now suppose we have a sample of N data points expressed in the form $\{(\mathbf{x}_i, L_i)\}_{i \in \mathcal{N}} := \{(\mathbf{x}_1, L_1), \dots, (\mathbf{x}_i, L_i), \dots, (\mathbf{x}_N, L_N)\}$, where $\mathbf{x} \in \mathbb{R}^p$ is a vector of features or covariates making up the context and $L \in \mathbb{R}^+$ is the random net system demand. As stated in the previous section, our objective is to utilize said sample to infer a functional relation $\hat{L} = h(\mathbf{x})$, $h : \mathbb{R}^p \rightarrow \mathbb{R}^+$, such that, given the context \mathbf{x} , the provided prescription \hat{L} is trained to deliver the minimum total system costs in expectation when inserted into the power balance equation (5b). For simplicity, and because it proves to perform very satisfactorily in the numerical experiments of Section 6, we restrict h to the family of affine linear functions, i.e., $h(\mathbf{x}) = \mathbf{q}^\top \mathbf{x}$, with $\mathbf{q} \in \mathbb{R}^p$ and with one of the features, say x_1 , fixed to one. This selection, together with the particular structure of problems (5) and (6) circumvent the need for applying the conventional procedures to solve the bilevel problem (3). Instead, to determine \mathbf{q} , we solve the following empirical risk minimization problem:

$$\min_{\mathbf{q}, Y} \frac{1}{N} \sum_{i \in \mathcal{N}} \sum_{g \in G} (C_g p_{gi} + C_{g_i}^u r_{gi}^u - C_{g_i}^d r_{gi}^d) \quad (7a)$$

$$\text{s.t. } 0 \leq p_{gi} + r_{gi}^u - r_{gi}^d \leq \bar{P}_g, \quad \forall i \in \mathcal{N}, \quad \forall g \in G \quad (7b)$$

$$0 \leq r_{gi}^u \leq R_g^u, \quad \forall i \in \mathcal{N}, \quad \forall g \in G \quad (7c)$$

$$0 \leq r_{gi}^d \leq R_g^d, \quad \forall i \in \mathcal{N}, \quad \forall g \in G \quad (7d)$$

$$\sum_{g \in G(b)} (p_{gi} + r_{gi}^u - r_{gi}^d) = \sum_{d \in D(b)} L_{di} + \sum_{l: o(l)=b} f_{li} - \sum_{l: e(l)=b} f_{li}, \quad \forall i \in \mathcal{N}, \quad \forall b \in B \quad (7e)$$

$$|f_{li}| \leq \bar{F}_l, \quad \forall i \in \mathcal{N}, \quad \forall l \in A \quad (7f)$$

$$\sum_{g \in G} p_{gi} = \hat{L}_i, \quad \forall i \in \mathcal{N} \quad (7g)$$

$$\hat{L}_i = \sum_{j=1}^p q_j x_{ji}, \quad \forall i \in \mathcal{N} \quad (7h)$$

$$u_{gi} \bar{P}_g \leq p_{gi} \leq u_{(g-1)i} \bar{P}_g, \quad \forall i \in \mathcal{N}, \quad \forall g \in G : g > 1 \quad (7i)$$

$$u_{gi} \bar{P}_g \leq p_{gi} \leq \bar{P}_g, \quad \forall i \in \mathcal{N}, \quad g = 1 \quad (7j)$$

$$u_{gi} \leq u_{(g-1)i}, \quad \forall i \in \mathcal{N}, \quad \forall g \in G : g > 1 \quad (7k)$$

$$u_{gi} \in \{0, 1\}, \quad \forall i \in \mathcal{N}, \quad \forall g \in G, \quad (7l)$$

where $Y := \{p_{gi}, r_{gi}^u, r_{gi}^d, u_{gi}, f_{li}\}_{(i,g,l)}$. Intuitively, the estimation problem (7) computes the coefficient vector $\mathbf{q} = (q_1, \dots, q_p)$ such that the total system cost averaged over the sample is minimized. This is the reason why all the decision variables related to the power dispatch and the provision of regulating power, i.e., $p_{gi}, r_{gi}^u, r_{gi}^d$, appear augmented with the sample index i in (7). Constraints (7b)–(7g) serve exactly the same purpose as their analogs in (5) and (6). Eq. (7h) expresses the prescription \hat{L}_i of the net system demand L under context \mathbf{x}_i as an affine function of the features, whose coefficients are to be computed by solving (7). Finally, the set of constraints (7i)–(7l) guarantee that the power dispatch $\{p_{gi}\}_{g \in G}$ coming from (7) for each sample i is optimal in the forward problem (5), that is, these constraints constitute the optimality conditions of the forward dispatch problem (5). Accordingly, these constraints enforce, for each sample i , the merit-order dispatch of the generation units $\{p_{gi}\}_{g \in G}$, forcing that $p_{g'i} > 0 \implies p_{gi} = \bar{P}_g$, for all $g \in G : g < g'$. Importantly, this formulation of the optimality conditions of problem (5) neither necessitates dual variables, nor large enough constants, both of which are frequently used to solve lineal bilevel programs [21].

Problem (7) is, therefore, a mixed-integer linear program due to the binary character of variables u_{gi} , which are used to impose the cost-merit order. As such, this problem can be solved using commercially available solvers such as CPLEX [28]. Once we obtain the optimal coefficient vector \mathbf{q}^* , we can produce the net-demand prescription $\hat{L} =$

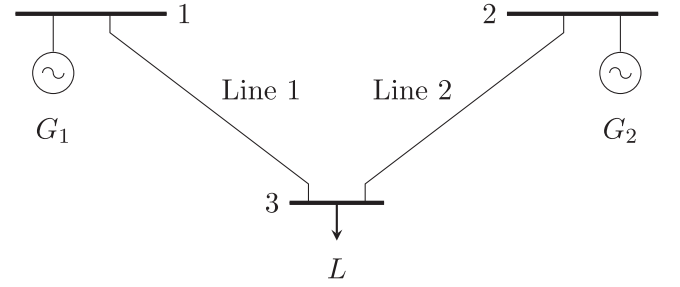


Fig. 2. Three-bus power system with one random demand and two thermal generators.

$(\mathbf{q}^*)^\top \mathbf{x}$, which is to be fed into (5b) under the context \mathbf{x} to readily obtain the forward dispatch decisions.

Now, let L^F denote the expected net-demand value (which is typically known as a point prediction of the net demand). Even though this expected value has been and is normally used as \hat{L} in the forward problem (5), it is not consistent with the plausible asymmetry in the cost of dealing with the subsequent prediction errors through the real-time problem (6). Indeed, it is most often the case that the cost of increasing the electricity production in real time is different from that of diminishing it. In this line, problem (7) offers a handy way to construct a new estimate \hat{L} to be used in (5) that takes into account the referred cost asymmetry. Indeed, the training problem (7) can be used in practice to upgrade the point prediction L^F to a prescription for \hat{L} , which does account for the asymmetry of the power system's balancing costs. For this, it suffices to include L^F as one of the features that are part of the context \mathbf{x} . This is what we do in the following example and in the case study of Section 6. That said, the context \mathbf{x} could include any other information that could be deemed useful to improve the prescription \hat{L} , for instance, the estimated standard deviation or quantiles of the conditional net demand, if these were available through a probabilistic forecast.

We finish this discussion with an important remark. Despite the fact that constraints (7i)–(7l) turn our training problem (7) into a mixed-integer program, enforcing the cost-merit order through these constraints is critical to train an affine model $h(\mathbf{x}) = \mathbf{q}^\top \mathbf{x}$ that renders economic benefits within the two-stage scheduling problem described in Section 2. Furthermore, even though the training problem (7) may require some computational effort to be solved, the affine model it delivers is intended to remain effective for a period of time (e.g., weeks or months), and hence, the task of solving the mixed-integer program (7) only has to be undertaken once in a while. That said, the parameter vector \mathbf{q} is to be re-estimated offline on a regular basis to capture changes in the dynamics of the electricity market and the power system (for example, to account for seasonal variations).

4. Example

Consider the small three-bus system depicted in Fig. 2, which is composed of one random demand L at bus 3, two thermal generators, G_1 and G_2 , at buses 1 and 2, respectively; and two lines, Line 1 and Line 2, connecting nodes 1 and 3, and buses 2 and 3, in that order.

The technical and economic characteristics of generating units G_1 and G_2 are collated in Table 1. Note that, in comparison, unit G_1 is smaller and cheaper than G_2 . In contrast, the latter is significantly more flexible as it features re-dispatch costs, i.e., C_g^u and C_g^d , that are much more competitive. We remark that $C_1^d = -20 \text{ €/MWh}$ implies that this power unit must be paid 20 € for each MWh its production is decreased in the real-time stage. Unless stated otherwise, line capacities are assumed infinite.

The only demand in the system, namely, L , is random. Suppose we have a sample $\{(\mathbf{x}_i, L_i)\}_{i=1}^N$, where the feature $\mathbf{x}_i = (1, L_i^F)^\top$. Again, L_i^F represents a classical point prediction of the demand L built out of

Table 1
Technical and economic specifications of power plants. Marginal costs are given in €/MWh and capacities in MW.

| | C_g | C_g^u | C_g^d | \bar{P}_g | R_g^u | R_g^d |
|-------|-------|---------|---------|-------------|---------|---------|
| G_1 | 5 | 30 | -20 | 60 | 60 | 60 |
| G_2 | 15 | 20 | 10 | 150 | 150 | 150 |

whichever available information the forecaster had at her disposal to produce it by way of whatever machine learning or forecasting technique she could have developed to that end. We stress that this setup is very common in reality, where power system operators often count on specialized software to produce good point predictions L_i^F . Our objective is to use our methodology and training model (7) described in Section 3 to recycle this standard point prediction with the aim of fabricating a better value for \hat{L} in Eq. (5b).

For this small example, we generate samples in the form $\{(x_i, L_i)\}_{i=1}^N$ as follows. We consider that the per-unit (p.u.) point forecast of the net demand L follows a uniform distribution between a and b . Therefore, $L^F \sim \bar{L} \cdot U(a, b)$, where \bar{L} is a factor representing the maximum power load at bus 3. We further assume that the per-unit net demand itself L/\bar{L} follows a Beta distribution with mean equal to L^F/\bar{L} and standard deviation σ . Hence, $L \sim \bar{L} \cdot \text{Beta}(\alpha, \beta)$, where the scale and shape parameters α and β are related to the mean L^F/\bar{L} and the standard deviation σ as follows:

$$\frac{L^F}{\bar{L}} = \frac{\alpha}{\alpha + \beta}, \tag{8a}$$

$$\sigma^2 = \frac{\alpha \cdot \beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}. \tag{8b}$$

Eqs. (8) guarantee that the point prediction L^F is an unbiased estimator of the conditional net demand L . In this illustrative example we fix $\sigma = 0.075$ p.u. and generate 20 samples $\{(x_i, L_i)\}_{i=1}^N$ with $N = 750$. Each L_i^F in x_i is randomly drawn from $\bar{L} \cdot U(a, b)$. Given L_i^F/\bar{L} and σ , and provided that the system of nonlinear equations (8) has a solution (notice that $\alpha, \beta > 0$), parameters α_i and β_i can be computed as

$$\alpha_i = -\frac{1}{\sigma^2} \left(\left(\frac{L_i^F}{\bar{L}} \right)^2 - \frac{L_i^F}{\bar{L}} + \sigma^2 \right) \frac{L_i^F}{\bar{L}}, \tag{9a}$$

$$\beta_i = \frac{1}{\sigma^2} \left(\left(\frac{L_i^F}{\bar{L}} \right)^2 - \frac{L_i^F}{\bar{L}} + \sigma^2 \right) \left(\frac{L_i^F}{\bar{L}} - 1 \right). \tag{9b}$$

Each L_i is then randomly taken from $\bar{L} \cdot \text{Beta}(\alpha_i, \beta_i)$. We take the first 500 data points of each sample as the training set and the last 250 as the test set.

We postulate the affine model $\hat{L} = q_0 + q_1 L^F$ and solve problem (7) on the training set to determine coefficients q_0 and q_1 . Finally, to evaluate the performance of the affine model, for each data point (x_i, L_i) in the test set, we simulate the *sequential* forward and real-time problems (5) and (6), with $\hat{L} = q_0 + q_1 L_i^F$ in (5b), and L_i in (6e). We then compute the sum of the forward and real-time production costs averaged over the 250 data points in the test set. This mean sum is further averaged over the 20 samples we generate. Our approach, which uses a *prescription* of the system net demand for scheduling generation, is referred to as P-SC (from prescriptive scheduling). We compare it with the customary practice of directly using the point forecast L_i^F as \hat{L} in (5b), which is referred to as F-SC (from forecast scheduling). Notice that our approach boils down to the conventional one if $q_0 = 0$ and $q_1 = 1$. Finally, the relative cost difference between these approaches is denoted as Δcost .

In the results we discuss next, we set a base case with $a = 0.03$, $b = 0.97$, $\bar{L} = 100$ MW, and the technical and economic parameters of the three-bus system described above.¹ We then define variants of this case by changing one or some of those parameters.

4.1. Impact of power regulation costs

Table 2 provides the cost savings that our approach achieves with respect to the conventional one under different G_2 's power regulation costs. For completeness, this table also includes the average cost of these two approaches for the test set and the values of the intercept q_0 and the linear coefficient q_1 of the affine model for \hat{L} our approach utilizes. These values represent expectations over the test data points of the 20 samples generated as indicated above. The first row in the table corresponds to the base case.

Interestingly, our approach systematically corrects the point forecast of the net demand L downwards, with a linear coefficient q_1 which is, on average, lower than or equal to 1, and a negative intercept q_0 in expectation. This is so because it is economically advantageous for the system to cope with positive net demand errors (i.e., eventual demand increases) by deploying upward balancing energy from unit G_2 . Indeed, the alternative would be to deal with negative demand errors by down-regulating (i.e., by providing downward balancing energy) with unit G_1 , a recourse that is clearly much more expensive.

To further elaborate on this phenomenon, the second row in Table 2 provides results for a variant of the base case in which C_2^u has been decreased from 20 to 15 €/MWh. Now that up-regulating (i.e., providing upward balancing energy) through unit G_2 is even cheaper, the downward correction of our approach to the net demand point forecast is more pronounced and the associated cost savings due to said phenomenon become larger. On the contrary, if it is the provision of downward balancing energy by G_2 what becomes 5 €/MWh cheaper and, hence, free (see third row of Table 2), the net demand point forecast is barely corrected and the costs savings brought by our approach (with regard to F-SC) become smaller as a result. Note that correcting the point forecast upwards in this case (in an attempt to profit from the free downward balancing energy provided by G_2) would be counterproductive in reality, as the system may risk having to resort to the high-cost downward balancing energy of unit G_1 in those likely scenarios in which the net demand ends up being lower than the capacity of this unit.

At this point, it may be instructive to see what happens when we drop constraints (7i)–(7l) from the mixed-integer program through which we train the affine model $\hat{L} = q_0 + q_1 L^F$. This is indeed very tempting, because, if these constraints are removed, the training model (7) becomes a very pleasant linear program, similar to the stochastic-programming-based formulation advocated, for instance, in Pritchard et al. [26] (with the data points in (7) playing the role of the “scenarios” in Pritchard et al. [26]). However, these constraints ensure the optimality of the lower-level problem (3d)–(3f) and guarantee that the above affine model is learned following a cost-merit-order principle. Therefore, if these constraints are dropped from (7), the affine model $\hat{L} = q_0 + q_1 L^F$ is not trained for the target task. This is exactly what Table 3 shows. This table is analogous to Table 2, but for a linear training model made up of constraints (7a)–(7h) only. We denote this approach as L-SC from “Linear”. The training model L-SC ignores the merit order and hence, takes for granted that the system can benefit from the cheap downward balancing energy of unit G_2 by allocating a non-zero production to this unit in the forward problem regardless of whether unit G_1 has been fully dispatched or not. This is, however, an strategy forbidden by the problem, which explains the poor actual performance of L-SC. This phenomenon is especially notorious for the case $C_2^d = 15$ €/MWh, in which the demand is heavily overestimated (the mean of the estimated demand is increased by 45%) and only the free downward regulation from unit G_2 is used.

Therefore, due to the catastrophic impact that removing the merit-order constraints (7i)–(7l) from the training model (7) may have on the actual performance of the obtained affine model, the strategy L-SC is no longer considered in the rest of our analysis.

¹ We take $a = 0.03$ and $b = 0.97$ pu to ensure that (8) has a real solution.

Table 2

Cost savings in percentage under different values of G_2 's balancing energy costs C_2^u and C_2^d (both given in €/MWh).

| C_2 | C_2^u | C_2^d | F-SC cost | P-SC cost | Δcost | q_0 | q_1 |
|-------|---------|---------|-----------|-----------|---------------------|--------|-------|
| 15 | 20 | 10 | 418.59€ | 416.91€ | 0.40% | -0.277 | 0.982 |
| 15 | 15 | 10 | 404.40€ | 391.88€ | 3.10% | -0.253 | 0.899 |
| 15 | 20 | 15 | 413.65€ | 412.93€ | 0.17% | -0.285 | 1.009 |

Table 3

Cost savings in percentage under different values of G_2 's balancing energy costs C_2^u and C_2^d (both given in €/MWh). Constraints (7i)–(7l) have been dropped from the training model (7).

| C_2 | C_2^u | C_2^d | F-SC cost | L-SC cost | Δcost | q_0 | q_1 |
|-------|---------|---------|-----------|-----------|---------------------|--------|-------|
| 15 | 20 | 10 | 418.59€ | 444.05€ | -6.08% | 4.967 | 0.964 |
| 15 | 15 | 10 | 404.40€ | 418.84€ | -3.57% | 5.490 | 0.883 |
| 15 | 20 | 15 | 413.65€ | 680.92€ | -64.61% | 36.807 | 0.722 |

Table 4

Impact of grid congestion on cost savings.

| \bar{F}_1 (MW) | F-SC cost | P-SC cost | Δcost | q_0 | q_1 |
|------------------|-----------|-----------|---------------------|--------|-------|
| ∞ | 418.59€ | 416.91€ | 0.40% | -0.277 | 0.982 |
| 30 | 1034.70€ | 724.46€ | 29.98% | 15.725 | 0.175 |

4.2. Impact of grid congestion

Here we introduce a variant of the base case in which the capacity of Line 1 has been set to 30 MW. Recall that the capacity of this line in the base case is unlimited, which we denote by symbol “ ∞ ” in Table 4. The results collated in this new table are analogous to those in Table 2.

Recall that the estimation problem (7), whereby we determine the affine function $\hat{L} = q_0 + q_1 L^F$, explicitly accounts for network constraints. In contrast, the computation of the net-demand point forecast L^F is typically based on statistical criteria alone and, consequently, ignores any possible limiting effect of the grid.

When the capacity of Line 1 is limited to 30 MW, our approach strongly corrects the point prediction L^F downwards, so that \hat{L} is kept in between 16 and 32 MW approximately. Thus, unit G_1 is dispatched well below the expected demand. This is clever because, in doing so, no (expensive) downward regulation from this unit has to be deployed in real time to comply with the limiting capacity of Line 1. In this way, the eventual realized demand at bus 3 can be satisfied, instead, with cheaper up-regulation from unit G_2 through Line 2. The ultimate result is that using \hat{L} , given by our approach, in the forward problem (5) is way more profitable than using the raw point forecast L^F .

Based on this analysis, if network constraints were accounted for at the day-ahead stage, that is, in the optimization problem (5), the cost savings achieved by our approach would decrease, at the expense of significantly complicating the resolution of the bilevel problem (3).

4.3. Impact of the peak demand

Now we change the peak demand and consider two variants of the base case in which we take $\bar{L} = 50$ MW and $\bar{L} = 150$ MW (in the base case, $\bar{L} = 100$ MW). The results of this new analysis are compiled in Table 5.

Again, as in the analysis of the impact of G_2 's balancing energy costs in Section 4.1, our approach systematically corrects the net-demand point forecast downwards to reduce the usage of down-regulation from G_1 in favor of the up-regulation from G_2 . However, the cost savings achieved by our approach get diluted as the peak demand is augmented. The reason for this is twofold. First, the probability of events where the net demand takes on a value below the capacity of unit G_1 diminishes with growing \bar{L} . For instance, when $\bar{L} = 50$ MW, the probability that the net demand is smaller than the capacity of G_1 is equal to one, which explains why our method delivers the highest

Table 5

Impact of peak demand.

| \bar{L} | F-SC cost | P-SC cost | Δcost | q_0 | q_1 |
|-----------|-----------|-----------|---------------------|--------|-------|
| 50 | 182.92€ | 181.55€ | 0.75% | -0.138 | 0.982 |
| 100 | 418.59€ | 416.91€ | 0.40% | -0.277 | 0.982 |
| 150 | 751.73€ | 750.54€ | 0.16% | -0.421 | 0.997 |

Table 6

Impact of the net-demand regime.

| U(a, b) | F-SC cost | P-SC cost | Δcost | q_0 | q_1 |
|---------------|-----------|-----------|---------------------|--------|-------|
| U(0.03, 0.97) | 418.59€ | 416.91€ | 0.40% | -0.277 | 0.982 |
| U(0.03, 0.50) | 239.60€ | 234.54€ | 2.11% | -0.102 | 0.917 |
| U(0.50, 0.97) | 587.82€ | 586.42€ | 0.24% | -6.646 | 1.088 |

cost savings in this variant (from among the three cases considered in this analysis). In contrast, as \bar{L} grows, that probability diminishes and the cheaper downward balancing generation from G_2 becomes more available. Second, the balancing costs account for a lower percentage of the total costs as the peak demand \bar{L} increases.

4.4. Impact of the net demand regime

We conclude this small example by studying how the net demand regime affects the prescriptive power of the affine function $\hat{L} = q_0 + q_1 L^F$ that we determine by way of problem (7). To this end, we modify the support of the uniform distribution from which the per-unit net-demand point prediction is randomly drawn. Thus, we distinguish a *low-demand* regime, with $L^F \sim \bar{L} \cdot U(0.03, 0.5)$, and a *high-demand* regime, with $L^F \sim \bar{L} \cdot U(0.5, 0.97)$. We also consider the base case, where $L^F \sim \bar{L} \cdot U(0.03, 0.97)$ and therefore, no demand regime is differentiated. The corresponding results are provided in Table 6.

In line with the observations in the previous analysis of the impact of the peak demand, under a low-demand regime, the expensive, but flexible unit G_2 is not dispatched in the forward problem. The downward correction to the net-demand point forecast our approach prescribes is then intended to benefit from the up-regulation provided by G_2 , which is clearly more competitive than the down-regulation offered by G_1 . The system features, therefore, a distinct cost asymmetry given by the expensive downward balancing generation of G_1 versus the cheap upward balancing generation of G_2 . Our approach sees this asymmetry and corrects the net-demand point forecast downwards accordingly. In addition, since the beta distribution modeling the point forecast error is right-skewed for low levels of demand, said correction leads to substantial cost savings. In contrast, under a high-demand regime, G_2 is very likely to participate in the forward dispatch, whereas there is a lower probability that G_1 be needed to provide downward balancing energy, since the distribution of the point forecast error is left-skewed. Consequently, the cost structure of the system looks very different under a high-demand regime, which prompts a quite different affine function and reduces the cost savings obtained from our method.

Most importantly, in the base case, when no net-demand regime is distinguished, most of the benefits our approach can potentially bring for low values of net demand are lost. This motivates us to cluster net-demand observations into different regimes and use optimization problem (7) to compute a possibly different affine model in the form $\hat{L} = q_0 + q_1 L^F$ for each demand regime, similarly to segmented regression in classical statistics. This is formalized in the next section.

5. Data clustering and partitioning

Take $\mathcal{N} := \{1, \dots, i, \dots, N\}$, that is, the index set of the data sample $\{(\mathbf{x}_1, L_1), \dots, (\mathbf{x}_i, L_i), \dots, (\mathbf{x}_N, L_N)\}$ with $\mathbf{x}_i \in \mathbb{R}^p$ and $L_i \in \mathbb{R}^+$, $\forall i \in \mathcal{N}$. We partition \mathcal{N} into a collection $\{\mathcal{N}_k\}_{k=1}^K$ of K subsets that are

Table 7
Base and peak generation capacity (GW) installed per node of the European network.

| Country | AT | BE | BG | CH | CZ | DE | DK | EE | ES | FI | FR | GB | GR | HR |
|---------|-----|-----|------|-----|------|------|------|-----|------|-----|------|------|-----|-----|
| Base | 0.4 | 6.1 | 6.7 | 3.4 | 14.4 | 46 | 2.4 | 2 | 16.3 | 6.5 | 68.3 | 20.6 | 3.9 | 1.3 |
| Peak | 5.9 | 6.8 | 1 | 0.6 | 1.3 | 27.9 | 1.7 | 0.2 | 29.6 | 3.6 | 11.9 | 31.2 | 4.9 | 0.7 |
| Country | HU | IE | IT | LT | LU | LV | NL | NO | PL | PT | RO | SE | SI | SK |
| Base | 3.3 | 1.8 | 8.7 | 0 | 0 | 0 | 4.5 | 0 | 27.8 | 1.8 | 5.4 | 11.1 | 1.8 | 2.7 |
| Peak | 4.1 | 4.2 | 46.2 | 1.8 | 0.1 | 1.2 | 19.3 | 0 | 3.5 | 4.6 | 2.9 | 1.1 | 0.7 | 1.5 |

pairwise disjoint and whose union is equal to \mathcal{N} . Consider the one-to-one mapping $\phi : \mathcal{N} \rightarrow \{1, 2, \dots, K\}$, such that $\phi(i) = k$ if data point $(\mathbf{x}_i, L_i) \in \mathcal{N}_k$. Therefore, $\mathcal{N}_k := \{i \in \mathcal{N} : \phi(i) = k\}$.

We compute K affine models of the form $\hat{L} = \mathbf{q}_k^T \mathbf{x}$, $k \leq K$, by solving the estimation problem (7) for each subset sample \mathcal{N}_k . In practice, this means replacing N and \mathcal{N} in (7) with $|\mathcal{N}_k|$ and \mathcal{N}_k , respectively.

To construct a meaningful mapping ϕ , we employ the K -means algorithm that is implemented in the Python package *scikit-learn* [29], using the Euclidean distance. We note that, to construct ϕ , this algorithm receives the feature sample $\{\mathbf{x}_i\}_{i \in \mathcal{N}}$ as input. Once the mapping ϕ is computed, it is used to determine the cluster to which a new feature vector \mathbf{x} belongs. That cluster is the one that minimizes the distance between its centroid and the feature vector \mathbf{x} . That is, given a new observation of \mathbf{x} , say \mathbf{x}_{N+1} , $\phi(\mathbf{x}_{N+1}) = k$ means that \mathbf{x}_{N+1} is predicted to belong to partition \mathcal{N}_k , and therefore, $\hat{L} = \mathbf{q}_k^T \mathbf{x}_{N+1}$ is to be used in the forward problem (5).

On a different issue, the estimation problem (7) is a MIP program and, as such, computationally expensive in general. Actually, the size of (7) grows linearly with the sample size. To keep the time to solve (7) reasonably low, we reduce the cardinality of subsets $\{\mathcal{N}_k\}_{k=1}^K$ by means of the PAM K-medoids algorithm [30] through the Python package implementation *scikit-learn-extra*. This algorithm selects the most representative data points within each subset \mathcal{N}_k , the so-called *medoids*, by minimizing the sum of distances between each point in \mathcal{N}_k and said medoids. We remark that this reduction process results in data points (the medoids) with *unequal* probability masses, so extra care should be taken when formulating objective function (7a) for each subset \mathcal{N}_k considering the medoids only. More specifically, the uniform weight $\frac{1}{N}$ appearing in the objective function (7a) should be replaced with a medoid-dependent weight representing the probability mass assigned to each medoid as a result of the reduction process.

6. Case study

In this section we assess the performance of our approach in a realistic case study that is based on the stylized model for the European power system that is described in Nahmmacher et al. [31]. Accordingly, we consider a pipeline network model with 28 nodes, each representing an European country. The capacities of the lines are also obtained from Nahmmacher et al. [31], in particular, we take the values from “Table 14. Transmission capacities between model regions (GW)” that correspond to the year 2020. We assume that each node in the network (i.e., each European country) includes two types of power plants technologies, which we denote as *base* and *peak*, respectively. Again, the available capacity of both technologies has been assigned based on the data in Nahmmacher et al. [31] corresponding to 2020 for each country. More specifically, the base power-plant capacities have been obtained by adding up the installed capacities of the technologies “Nuclear”, “Hard coal”, “Oil” and “Lignite” and the peak power-plant capacities from the technologies “Natural Gas”, “Waste” and “Other gases”. The nodes of the system and the resulting generation capacities of each type are listed in Table 7.

To build a data sample of the form $\{(\mathbf{x}_i, L_i)\}_{i \in \mathcal{N}}$, we have collected the actual aggregate hourly demand, wind, solar and hydro energy production for each country (node of the system) in 2020 from the ENTSO-e Transparency Platform [18]. We have also retrieved the day-ahead forecast of the hourly demand and the produced wind and solar

Table 8
Uniform distributions from which the marginal production, up- and down-balancing costs (in €/MWh) of the units in the European system have been sampled.

| | C | C^u | C^d |
|------|-----------|-----------|-------------|
| Base | U(8, 12) | U(60, 70) | U(-40, -50) |
| Peak | U(36, 44) | U(45, 50) | U(30, 35) |

energy from this platform. To get series of net demand values (both forecast and actual), we have subtracted the respective wind, solar and hydro power data series from the aggregate day-ahead forecast/actual demand series. We clarify that no day-ahead forecast for the hydro power production is available in ENTSO-E Transparency Platform [18], so the series of real hydro power production has been used (instead of the missing day-ahead hydro forecast) for the computation of day-ahead forecasts of the nodal net demands. Some minor gaps in the data extracted from ENTSO-E Transparency Platform [18] have been filled through linear interpolation.

The marginal costs of energy generation and up- and downward balancing energy provision of each unit are randomly sampled from the uniform distributions specified in Table 8. The so-obtained values for these costs have remained fixed throughout the experiments performed in this section. We point out that in the uniform distributions of Table 8, we have considered that base power units are cheap but inflexible, and thus, with costly balancing energy. In contrast, peak power plants are expensive, but flexible, and hence, with more competitive balancing energy costs.

We conduct a rolling simulation on the data of 2020, in which we gradually select non-overlapping windows of 150 points each. From each window, we randomly sub-sample (without replacement) the indexes corresponding to the training and test sets, which are eventually made up of 100 and 50 samples, respectively. We take ten windows over which we average the results that follow. First, we compute the average cost of the baseline method F-SC, i.e., the method that uses the point forecast L^F of the net demand as \hat{L} in (5), which amounts to 5092.8k€. For comparison, we also compute the average cost of a perfect information benchmark in which \hat{L} is equal to the actual realization of the net demand L . This perfect forecast approach is unrealistic but can be used to assess how much could be gained from improving the prescribed net demand. For this case study, the perfect information cost is 3380.9k€, i.e, 33.6% lower than that delivered by method F-SC.

As in the example of Section 4, we consider a feature vector \mathbf{x} made up of the day-ahead forecast of the system net demand, L^F , (measured in MWh), enlarged with an additional feature fixed to one to accommodate the intercept of the affine models $\hat{L} = \mathbf{q}_k^T \mathbf{x} = q_{0k} + q_{1k} L^F$, $k \leq K$.

In the analysis we conduct next, we consider various values for K (number of partitions and hence of affine models) and several percentage reductions of the number of data in each partition $\mathcal{N}_k, k \leq K$. The results of this analysis are summarized in Table 9, where “ $r\%$ ” in the first column means that only that percentage of medoids in the partition \mathcal{N}_k (more precisely $\lceil \frac{r}{100} |\mathcal{N}_k| \rceil$, where $\lceil \cdot \rceil$ denotes the ceiling operator) have been used to estimate the affine function $\hat{L} = \mathbf{q}_k^T \mathbf{x}$ through (7). This table shows the average cost achieved by our approach for different values of r and K , the cost savings with respect

Table 9

Average cost savings and average time to solve the estimation problem (7) for a number K of partitions and various levels r of reduction in the size of the original training sets (in percentage).

| r | K | P-SC cost | Δcost | Time (s) |
|------|-----|-----------|---------------------|----------|
| 100% | 1 | 4948.7k€ | 2.83% | 2127.7 |
| 100% | 2 | 4874.4k€ | 4.29% | 283.7 |
| 100% | 5 | 4851.6k€ | 4.74% | 75.9 |
| 100% | 7 | 4851.1k€ | 4.75% | 28.0 |
| 50% | 1 | 4956.9k€ | 2.67% | 180.0 |
| 50% | 2 | 4877.3k€ | 4.23% | 27.2 |
| 50% | 5 | 4869.4k€ | 4.39% | 7.4 |
| 50% | 7 | 4886.1k€ | 4.06% | 5.5 |
| 20% | 1 | 4971.6k€ | 2.38% | 8.3 |
| 20% | 2 | 4883.2k€ | 4.12% | 3.2 |
| 20% | 5 | 4882.8k€ | 4.12% | 1.1 |
| 20% | 7 | 4890.9k€ | 3.97% | 1.4 |

to the cost of the baseline method F-SC (5092.8k€), and the average time the solution to the K estimation problems (7) takes. The reported cost savings have been computed out of sample, that is, on the test sets. Beyond the fact that these savings are significant in general, it is clear that our prescriptive approach benefits from exploiting different affine models under different net-demand regimes, which confirms the preliminary conclusion we draw in this regard through the small example of Section 4. Nevertheless, it is also true that the added benefit rapidly plateaus as K grows. Actually, the bulk of the economic gains we get through the partitioning of the data sample is already reaped with $K = 2$. On the other hand, increasing K has a positive side effect: It remarkably reduces the time to solve the MIP problem (7). In addition, this time can be shortened even further, with a tolerable reduction in cost savings, by using only the medoids of the partitions $\mathcal{N}_k, k \leq K$, when estimating the affine models through (7). Notwithstanding this, K cannot be made arbitrarily big since it can lead to *overfitting*. Therefore, the choice of K should be guided by a training-validation scheme similar to the one we show in Table 9 to ensure that increasing K does not lead to a reduction in the benefits of the affine rules out of sample.

To comprehend where those cost savings our approach yields come from, in Fig. 3 we plot the predicted aggregate net demand L^F against the one prescribed by our method, i.e., \hat{L} . The plot corresponds to one window of 150 data points taken at random out of the ten we have considered in the rolling-window simulation. Furthermore, the figure depicts results from the case with five partitions ($K = 5$). It can be seen that, when the system net demand is predicted low, our method prescribes to overestimate it. This prescription is motivated by two facts. On the one hand, the overestimation of the net demand in the forward problem is covered by cheap power plants, whereas it reduces the need for upward balancing. On the other, even though it slightly increases the demand for downward balancing, the group of units that down-regulate remains the same in any case, i.e., with and without the overestimation, due to the limitations of the network. As a result, the cost savings linked to the reduction in up-balancing outweigh the extra costs incurred by the increase in down-balancing. It is interesting to note that, as mentioned at the end of Section 2, system operators, based on their accumulated experience, often introduce an upward bias into the net demand forecast [24], precisely as our model here suggests.

As the level of net demand grows, the overestimation of the system net load that our method prescribes diminishes to a point where the prescribed amount flattens (see partitions \mathcal{N}_4 and \mathcal{N}_5). Again, this phenomenon is caused by the network and the limitations it imposes. Indeed, our method avoids dispatching power plants in the forward problem which, despite being their turn in the cost-merit order, would have been irretrievably down-regulated in real time because of network bottlenecks. For instance, in partition \mathcal{N}_4 , F-SC consistently dispatches the DE base generator, with its massive 46 GW, to maximum capacity. However, due to grid constraints, this unit is subsequently

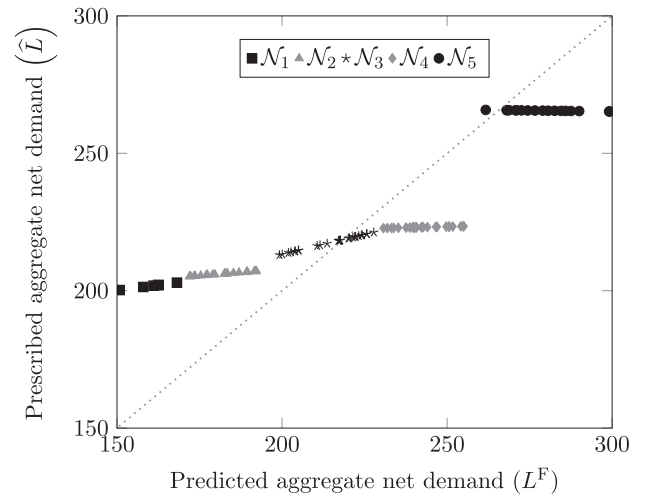


Fig. 3. Prescribed affine transformation of the day-ahead net-demand forecast (aggregated system-wise). Demand is given in GW.

down-redispached to around 30 GW. On the contrary, P-SC takes into consideration that this power plant is one of the latest to be scheduled in this partition and foresees the grid limitation on the power flow, thus constraining the aggregated energy production and systematically dispatching such a unit to the previously mentioned 30 GW.

Finally, Fig. 3 also illustrates that, even though the prescriptive functions $h_q(\cdot)$ we employ are affine, the use of several clusters leads to a *piece-wise* affine function that is able to approximate the non-linear relationships between the features and the prescribed parameters.

7. Conclusions

In this paper, we have proposed a data-driven method to prescribe the value of net demand that the forward stage in a two-stage generation scheduling problem should use in order to minimize the expected total cost of operating the underlying power system. For this purpose, we have formulated a mixed-integer linear program that trains an affine function to map the *predicted* net demand into the *prescribed* one.

Numerical experiments conducted out of sample on a stylized model of the European electricity grid reveal that the cost savings implied by the estimated affine mappings are substantial, well above 2%. Furthermore, on the grounds that the cost structure of a power system is highly dependent on its operating point, and hence, on the level of net demand, we have devised a K -means-based partition strategy of the data sample to train different affine mappings for different net-demand regimes. The utilization of this strategy is shown to have a positive twofold effect in the form of substantially increased costs savings and a remarkable drop in the computational burden of the proposed MIP training model. Finally, we have further complemented the partitioning of the data sample with a medoid-based reduction in the size of the partitions, achieving additional speedups in solution times. All this together opens up the possibility to leverage our prescriptive approach in larger instances.

Future work will include attempts to optimize the partitioning of the data sample by embedding it into the MIP training model and to devise a procedure to efficiently update the parameters of the affine rules as new data become available without having to solve a new instance of the estimation problem (3). Another direction for future research is the extension of the proposed methodology to prescribe net demand trajectories (of 24 h, for example) for generation scheduling problems that include inter-temporal constraints such as ramping limits or minimum times.

CRedit authorship contribution statement

J.M. Morales: Conceptualization, Methodology, Investigation, Formal analysis, Validation, Writing – original draft, Writing – review & editing, Supervision, Funding acquisition. **M.A. Muñoz:** Data curation, Software, Methodology, Investigation, Formal analysis, Validation, Writing – review & editing. **S. Pineda:** Conceptualization, Methodology, Investigation, Formal analysis, Validation, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work was supported in part by the European Research Council (ERC) under the EU Horizon 2020 research and innovation program (grant agreement No. 755705), in part by the Spanish Ministry of Science and Innovation (AEI/10.13039/501100011033) through project PID2020-115460GB-I00, and in part by the Junta de Andalucía (JA) and the European Regional Development Fund (FEDER) through the research project P20_00153. M. Á. Muñoz is also funded by the Spanish Ministry of Science and Innovation through the State Training Subprogram 2018 of the State Program for the Promotion of Talent and its Employability in R&D&I, within the framework of the State Plan for Scientific and Technical Research and Innovation 2017–2020 and by the European Social Fund (reference PRE2018-083722 and ENE2017-83775-P). Finally, the authors thankfully acknowledge the computer resources, technical expertise, and assistance provided by the SCBI (Supercomputing and Bioinformatics) center of the University of Málaga. Partial funding for open access charge: Universidad de Málaga / CBUA.

References

- [1] Birge JR, Louveaux F. Introduction to stochastic programming. Springer Science & Business Media; 2011.
- [2] Ben-Tal A, El Ghaoui L, Nemirovski A. Robust optimization. Princeton University Press; 2009.
- [3] Morales JM, Conejo AJ, Madsen H, Pinson P, Zugno M. Integrating renewables in electricity markets: Operational problems, vol. 205. Springer Science & Business Media; 2013.
- [4] Bertsimas D, Litvinov E, Sun XA, Zhao J, Zheng T. Adaptive robust optimization for the security constrained unit commitment problem. *IEEE Trans Power Syst* 2012;28(1):52–63.
- [5] Dvorkin Y. A chance-constrained stochastic electricity market. *IEEE Trans Power Syst* 2019;35(4):2993–3003.
- [6] Morales JM, Zugno M, Pineda S, Pinson P. Electricity market clearing with improved scheduling of stochastic production. *European J Oper Res* 2014;235(3):765–74.
- [7] Wang B, Hobbs BF. Real-time markets for flexiramp: A stochastic unit commitment-based analysis. *IEEE Trans Power Syst* 2015;31(2):846–60.
- [8] Morales JM, Pineda S. On the inefficiency of the merit order in forward electricity markets with uncertain supply. *European J Oper Res* 2017;261(2):789–99.
- [9] Kazempour J, Pinson P, Hobbs BF. A stochastic market design with revenue adequacy and cost recovery by scenario: Benefits and costs. *IEEE Trans Power Syst* 2018;33(4):3531–45.
- [10] Donti P, Amos B, Kolter JZ. Task-based end-to-end model learning in stochastic optimization. In: *Advances in neural information processing systems*. 2017, p. 5484–94.
- [11] Muñoz M, Pineda S, Morales J. A bilevel framework for decision-making under uncertainty with contextual information. *Omega* 2022;108:102575.
- [12] Elmachtoub AN, Grigas P. Smart “Predict, then Optimize”. *Manage Sci* 2022;68(1):9–26.
- [13] El Balghiti O, Elmachtoub AN, Grigas P, Tewari A. Generalization bounds in the predict-then-optimize framework. In: *Advances in neural information processing systems*, vol. 32. Curran Associates, Inc.; 2019.
- [14] Dvorkin V, Delikaraoglou S, Morales JM. Setting reserve requirements to approximate the efficiency of the stochastic dispatch. *IEEE Trans Power Syst* 2018;34(2):1524–36.
- [15] Carriere T, Kariniotakis G. An integrated approach for value-oriented energy forecasting and data-driven decision-making application to renewable energy trading. *IEEE Trans Smart Grid* 2019;10(6):6933–44.
- [16] Muñoz M, Morales JM, Pineda S. Feature-driven improvement of renewable energy forecasting and trading. *IEEE Trans Power Syst* 2020;35(5):3753–63.
- [17] Garcia JD, Street A, Homem-de-Mello T, Muñoz FD. Application-driven learning via joint estimation and optimization for demand and reserves requirement forecast. 2021, preprint [arXiv:2102.13273](https://arxiv.org/abs/2102.13273).
- [18] ENTSO-E Transparency Platform. 2020, Available (online): <https://transparency.entsoe.eu/>. [Last Accessed 23 December 2021].
- [19] Shafieezadeh-Abadeh S, Kuhn D, Esfahani PM. Regularization via mass transportation. *J Mach Learn Res* 2019;20(103):1–68.
- [20] Fanghänel D, Dempe S. Bilevel programming with discrete lower level problems. *Optimization* 2009;58(8):1029–47.
- [21] Pineda S, Morales JM. Solving linear bilevel problems using big-Ms: not all that glitters is gold. *IEEE Trans Power Syst* 2019;34(3):2469–71.
- [22] Kleinert T, Labbé M, Plein F, Schmidt M. There’s no free lunch: on the hardness of choosing a correct big-M in bilevel optimization. *Oper Res* 2020;68(6):1716–21.
- [23] Yurdakul O, Meyer A, Sivrikaya F, Albayrak S. Forecasting daily primary three-hour net load ramps in the CAISO system. In: *2020 52nd North American power symposium*. IEEE; 2021, p. 1–6.
- [24] California Independent System Operator. 2019 annual report on market issues and performance. Technical report, Department of Market Monitoring; 2020, URL: <http://www.caiso.com/Documents/2019AnnualReportonMarketIssuesandPerformance.pdf>.
- [25] Aghaei J, Shayanfar H, Amjadi N. Joint market clearing in a stochastic framework considering power system security. *Appl Energy* 2009;86(9):1675–82.
- [26] Pritchard G, Zakeri G, Philpott A. A single-settlement, energy-only electric power market for unpredictable and intermittent participants. *Oper Res* 2010;58(4-part-2):1210–9.
- [27] Zavala VM, Kim K, Anitescu M, Birge J. A stochastic electricity market clearing formulation with consistent pricing properties. *Oper Res* 2017;65(3):557–76.
- [28] IBM ILOG CPLEX Optimization Studio. 2020, Available (online): <https://www.ibm.com/analytics/cplex-optimizer>.
- [29] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. *J Mach Learn Res* 2011;12:2825–30.
- [30] Kaufman L, Rousseeuw PJ. Finding groups in data: An introduction to cluster analysis, vol. 344. John Wiley & Sons; 2009.
- [31] Nahmmacher P, Schmid E, Knopf B. Documentation of LIMES-EU – A long-term electricity system model for Europe. Potsdam Institute of Climate Impact Research (PIK); 2014, URL: https://www.pik-potsdam.de/en/institute/departments/transformation-pathways/models/limes/DocumentationLIMES_EU_2014.pdf.