ARTICLE IN PRESS

Electrical Power and Energy Systems xxx (xxxx) xxx

Contents lists available at ScienceDirect

International Journal of Electrical Power and Energy Systems

journal homepage: www.elsevier.com/locate/ijepes

Modeling storage systems in electricity markets with high shares of renewable generation: A daily clustering approach

Alberto Orgaz *, Antonio Bello, Javier Reneses

Institute for Research in Technology (IIT), ICAI School of Engineering, Comillas Pontifical University, Madrid, 28015, Spain

ARTICLE INFO

Keywords:
Clustering
Energy storage systems
Optimization
Power system modeling
Temporal aggregation

ABSTRACT

In the energy transition towards a carbon-free society, the continuous changes that energy systems are experiencing, the increasing penetration of renewable generation, and the incorporation of short-term storage technologies such as batteries have increased the effort to model and predict its development and operation. In this context, power system models play a relevant role. However, to be suitable for the decision-making process, these tools should not require a massive computational effort. To overcome this challenge, this paper proposes a new methodology that reduces the temporal dimension of the problem while maintaining accurate results. This methodology is specially designed for medium-term operation of real-size power systems with a significant presence of renewable generation and storage systems. By means of a two-stage clustering algorithm, the proposed approach transforms the temporal structure of the model's input parameters into different levels of time aggregation. This arrangement makes the problem manageable and computationally tractable. In addition, together with the newly incorporated model formulation, this methodology allows capturing at the same time the short- and medium-term variability present in power systems.

1. Introduction

In the last years, and as a result of the goals proposed in the Paris Agreement of 2015 [1], numerous regulatory policies have originated with the aspiration of reducing the emission of greenhouse gasses to prevent global warming. In the electricity sector, the way to achieve these climate change goals is by lowering the environmental impact of energy generation, which generally implies decarbonization of the electricity systems. This process is enabled by the integration of renewable technologies such as wind and solar. However, renewable technologies lead to an increase in variability and uncertainty in the system. Therefore, the more penetration of this kind of generation, the bigger the need for flexibility in the system [2]. Flexibility on the generation side is provided by different technologies, such as gas power and energy storage systems, for instance, pumped hydro or battery energy storage systems (BESS) [3]. During sunny and windy days, storage systems are helpful to store the surplus of renewable generation. On the contrary, over long dark and windless periods, gas power plants act as a backup solution to compensate for the shortage in generation by rapidly turning on or off. However, because gas-power plants are carbon-emitting, energy storage systems are recently gathering momentum as the proposed solution for this issue [4]. Furthermore, BESS serves other purposes, such as peak-load shaving, frequency regulation, and energy management in centralized charging stations, which makes it an even more interesting alternative [5,6]. Since flexibility is often associated with the short-term dynamics of the system, the challenge resides in how to include all the effects of short-term variability in medium- to long-term operational studies with an appropriate level of detail.

In power system modeling, a key component to achieve an adequate representation of the outcome of energy systems is the temporal resolution chosen to model it. In this context, the temporal resolution becomes crucial to capture the variability of the wind and solar resources, as well as the cycles of the complementing technologies that provide the necessary flexibility in the system [7]. In electricity systems, time steps can vary from minutes or hours in the short-term operation, to years or decades, in the long-term planning. Indeed, the best results are obtained when setting this resolution to the highest degree, typically on an hourly basis. However, hourly resolution significantly increases the complexity of models, raising potential computational difficulties in real-size power systems. The most common alternative to face this issue is temporal aggregation. Following this approach often results in computationally tractable models that are, therefore, easier to solve. However, it also results in a loss of accuracy.

* Corresponding author.
E-mail addresses: alberto.orgaz@iit.comillas.edu (A. Orgaz), antonio.bello@iit.comillas.edu (A. Bello), javier.reneses@iit.comillas.edu (J. Reneses).

https://doi.org/10.1016/j.ijepes.2021.107706
Received 26 April 2021; Received in revised form 23 September 2021; Accepted 14 October 2021

Nomenclature

Indices and Sets

$b \in B$	Battery energy storage systems
$bl \in BL$	Time blocks (determined by changes in st)
$e \in E$	Market agents
$h \in H$	Hydro generation units
$hr \in HR$	Hours
$p \in P$	Periods (e.g., weeks, months)
$st, st' \in ST$	System states
$t \in T$	Thermal generation units
$td \in TD$	Types of days

Parameters

ρc_b	Charging efficiency [p.u.]
ρd_b	Discharging efficiency [p.u.]
$\theta_{e,p}$	Conjectured-price response [€/MWh^2]
cap_b	Maximum energy capacity of battery b [MWh]
D_p	Demand in period p [MWh]
dr_b	Maximum charging/discharging power [MW]
dur_{st}	Duration of system state st [h]
$Ndays_{p,td}$	Number of days of type of day td in period p
o_b	Ownership of battery b [p.u.]
o_h	Ownership of hydro unit h [p.u.]
o_t	Ownership of thermal unit t [p.u.]

Variables

λ_p	Electricity price in period p [€/MWh]
$be_{h,p,st}$	Pumped power [MW]
$C_{e,p}$	Cost function for agent e [€]
$pc_{b,p,st}$	Charging power of battery b [MW]
$pd_{b,p,st}$	Discharging power of battery b [MW]
$ph_{h,p,st}$	Hydro production [MW]
$pt_{t,p,st}$	Thermal production [MW]
$Q_{e,p}$	Production of agent e in period p [MWh]
$SOC_{b,p,td,bl}$	State of charge of battery b [MWh]

1.1. Literature review on temporal aggregation methods

Simplifying the temporal framework of the model significantly reduces the size of the problem and its complexity. In this regard, many works in the literature address temporal aggregation techniques. In the context of medium-term power system models, these techniques are generally classified into two different types of time-period aggregation: representative periods and system states [8].

1.1.1. Representative periods

They can be understood as groups of consecutive time steps that characterize operational cycles, such as hours, days, or even weeks. Following this approach, the whole horizon of the problem gets approximated by several repetitions of these representative periods. One of the advantages of reducing the time frame into representative periods is that the chronology is preserved within each period. In systems with a high share of renewable generation and flexible resources, chronology allows integrating the intra-period dynamics into the model to better reflect ramping or storage constraints, for instance.

The use of representative periods differs in terms of the duration of the period itself. Some works use typical or representative days to estimate the whole operation of the system. For instance, Ref. [9] proposes an original approach to represent a full year of data in six to ten typical days, depending on the share of renewable considered. It uses k-means clustering to extract representative profiles from the demand that, integrated into a model of the electricity system of Great Britain, generating accurate results. Aimed at the same power system, Ref. [10] introduces the use of several time series in the clustering process to find representative days that preserve the correlations present between system variables such as demand and wind and solar production. That helps to represent the fluctuations of renewable resources. In contrast, other works use a longer duration for the representative periods, such as weeks [11] or months [12]. This representation favors a better measure of the seasonal cycles present in the system operation.

However, none of the abovementioned works consider the continuity between the representative periods in their selection procedure. Preserving inter-period chronology is essential to be able to model long-term operations. In this sense, a novel formulation is introduced in [13] to address the coupling of typical days. This approach allows for capturing seasonal variability. However, this approach is applied at a regional level, raising concerns about its extensibility to large-scale, more complex problems. Furthermore, Ref. [8] introduces an enhanced version of the representative periods' methodology to account for the chronology between days. The results show an overall better representation of short- and long-term dynamics.

A new challenge arises when selecting these representative periods. On the one hand, the shape of representative periods in terms of period length and resolution determines the level of detail and representativeness of the model outcomes to a great extent. On the other hand, the number of representative periods is directly related to computational complexity. Therefore, its selection is critical for achieving both a tractable and accurate model. Many works in the literature attempt to establish methodologies to select representative periods and validate their effectiveness in energy system modeling. Ref. [14] provides an analysis where different time aggregation methods are compared in a case study of the electricity system of Great Britain. Although the authors conclude there is not a definitive method to accurately represent long-term operation with simplified time resolutions, they suggest heuristic-based methods offer superior performance. However, heuristics have limitations in establishing clear and general criteria to select representative periods for every situation. Another meaningful comparison of different methods is carried out in [15]. After analyzing several clustering methods, the authors conclude that the impact on the results is not significantly linked with the aggregation method used but with the system's characteristics in which it is applied. Furthermore, the authors in [16] provide a classification of time series aggregation distinguishing between two types of algorithms: optimization- and clustering-based procedures. Although optimization-based methods perform at least as well as clustering-based methods, their computational limitations often lead to using the latter instead.

On the whole, clustering methods are adequate when finding characteristic operational patterns. In this category, k-means clustering is probably the most common method for time-series aggregation [13,17,18]. K-means is an unsupervised learning method that forms groups of data points by minimizing the Euclidean distance between them. By doing so, each cluster is defined by a center (referred to as centroid), calculated as the arithmetic mean of the data points belonging to the cluster. Alternatively, k-medoids clustering shares the same properties of k-means, with the difference that instead of centroids, the centers of each cluster are located among the data points of the time series [8,19]. Another systematic approach for this purpose is hierarchical clustering [10,15]. This method initially represents each data point as individual clusters. Then, it iteratively merges similar clusters according to the chosen similarity measure until the targeted clusters are reached. Additionally, in [20], a framework is proposed that highlights the main features of aggregation techniques. After examined and compared different clustering methods to choose representative periods, the authors concluded that centroid-based algorithms achieve a better representation of the operational part of the system. Finally, a detailed classification of these works is provided in [21].

Another critical component of temporal aggregation methods is the representation of peak and off-peak hours behavior. Peak periods are hours with extreme conditions. The consideration of extreme events is relevant because the dispatching of the units during these periods may have a significant impact on wholesale market prices, among other implications. Therefore, to account for the extreme variability of the operational processes in power systems, extreme periods are usually removed from the targeted time series before clustering so that they can be later added independently as new clusters [19,22]. The authors in Ref. [15] compare several variants to integrate extreme periods: adding extreme periods as new cluster centroids and replacing clusters with peak periods. The observed results in a residential supply system across the different integrations show no significant difference over the inclusion of an extreme event as in [14,19]. Alternatively, the works in Refs. [17,23] include extreme events within the constraints of the optimization model, so they ensure feasibility in the sense that the results meet the unique requirements during those periods. More recently, a new iterative method was proposed in Ref. [24], where the authors use slack variables from energy system optimization problems to select extreme periods. According to this method, the optimization problem is iteratively run until convergence is met. Throughout the iterative process, the value of slack variables helps to identify new extreme periods that are successively integrated into the problem's constraints.

In short, the selection of representative periods is highly linked to the specific characteristics of the targeted system, making it difficult to establish a general methodology that works well in every case it is applied.

1.1.2. System states

Load levels or load blocks comprise one of the simplest ways to aggregate time in energy modeling. Traditionally, load duration curve blocks have been used to approximate the demand time series, simplifying the temporal framework of the model. Refs. [25], and [26] constitute examples of the implementation of this approach in power system models. However, the lack of chronology in this methodology makes it unsuitable for the new challenges of energy modeling.

Instead, system states constitute an enhanced framework of the traditional load levels. System states are created by grouping hours with similar characteristics within a period. This framework, introduced in [27], allows for efficient computation without a substantial accuracy loss since it incorporates the chronological information between time slots. Although system states are not temporally coupled, the transitions between them reflect this chronology. This framework helps the representation of the unit commitment in the model.

Several works in the literature have chosen similar approaches to maintain chronology when reducing the time dimension in electricity models. In [28], the authors incorporate the operation of storage units into medium-term models within the system states framework. This formulation is extended to multiple-node systems in [18]. Continuing with this work, Ref. [8] proposes an improved representation of the system states by dividing the formulation into short- and long-term storage constraints. This separation attempts to better handle each temporal scope's dynamics, although it adds further complexity to the problem. In addition, the short-term operation is not properly represented. Furthermore, in [29], the authors propose a novel approach using a hierarchical clustering procedure to group consecutive hours and determine the optimal capacity investment plan of a power system with a significant presence of renewable generation and storage units. Although this approach shows better performance than other methods, it cannot capture short-term dynamics well.

Recently, a comparison of the two main approaches for time aggregation was carried out in [30]. The authors conclude that the representative periods' approach in [8] achieve a better representation of the short-term storage. However, this approach is insufficient for overcoming the chronological clusters methodology in [29] in the handling of long-term dynamics. In short, none of these methods achieves a definitive superior outcome under every condition.

1.2. Objectives and contributions

Medium- and long-term analysis of real-size electricity systems requires a high level of detail of many outputs of the system operation, such as monthly contracts, the commitment of marginal units, hourly electricity prices, and weekly hydro management signals. All these outputs play a relevant role in the decision-making process. This degree of detail, though, is not accurately achieved applying classical aggregation approaches. In this work, a new approach is proposed to address this problem.

This paper proposes a comprehensive methodology to obtain a temporal representation suitable for medium- and long-term optimization models of power systems operation. It is especially applicable to systems with a significant presence of flexible units and a high share of renewable generation. This methodology consists of a hybrid approach, in which representative days are used in combination with system states. A novel two-stage clustering technique is implemented to obtain accurate daily representatives of the operation within each period (weeks or months). Unlike other works that select representative days from the whole time horizon of the problem [8,13], the methodology proposed in this paper divides the time horizon into periods, analyzes each day individually, and groups them according to its similarities. Finally, it selects a representative for each of those groups. These representatives are subsequently split into system states or groups of hours. This process allows preserving a daily representation of the system, which aids to accurately incorporate short-term variability in planning models and correctly model particular aspects of the system such as short-term storage and gas resources management. At the same time, monthly or weekly periods maintain the necessary chronology to properly characterize long-term storage operation and contract decisions, such as take or pay clauses. In this context, the concept of "short-term storage" mainly refers to BESS management, typically carried out hourly or daily. Therefore, it is distinguished from the modeling of medium-term storage operation, such as hydro reservoirs and units with pumping capabilities, mainly done weekly or monthly. Short- and medium-term storage operations are very relevant to obtaining a correct and accurate representation of market outcomes in the medium and long term.

The main contributions of this paper are outlined below:

- The definition of an innovative temporal aggregation methodology to model both the short- and medium-term operation in electricity systems. This methodology combines representative days and system states in a configurable fashion to suit the desired accuracy and available computational resources.
- The incorporation of short-term storage constraints into a medium-term market equilibrium model. This formulation can significantly reduce the computational intensity of the problem while providing valuable results. In this sense, although framed around a market equilibrium model, the proposed methodology is flexible and generic and could be implemented in any optimization model aimed at power market planning based upon fundamentals.
- An extensive sensitivity analysis of the results in a real-size electricity system case study, where the outcomes of a vast variety of model realizations are compared against a benchmarked hourly execution.

In short, the methodology proposed in this paper is specifically designed for its implementation in fundamental optimization models for the planning and operation of electricity markets in the medium and long term. Overall, this methodology may be helpful for market agents, regulators, and researchers that need practical tools to study and analyze the evolution of electricity systems under different regulatory policies or energy scenarios.

The remainder of this paper is structured as follows: Section 2 introduces the proposed methodology and describes the formulation of the market equilibrium model. Meanwhile, Section 3 contains the case study in which this methodology is applied for storage solutions management optimization. Finally, the conclusions of this work are provided in Section 4.

2. Methodology

This section covers the temporal aggregation methodology proposed in this paper and explains the main aspects of consideration when building it.

2.1. Temporal aggregation

The temporal dimension of the model can be divided into two features, the temporal resolution, also known as time steps, time slices or system states, and the temporal horizon, i.e., the temporal distance covered by the model. The time horizon typically ranges between a few months and two or three years in the medium term. As for temporal resolution, it will be defined by the proposed temporal aggregation methodology.

As explained above, time aggregation is implemented to simplify energy system models and make them easier to solve. In this sense, the primary motivation for carrying out temporal aggregation is to make the model suitable to be integrated into the decision-making process of a market agent in the medium to long term. Depending on how to carry out this aggregation, the results obtained from the models will realistically or not represent such systems.

In general, electricity systems are characterized by seasonalities at different scales. These patterns, generated by consumer behavior or climate conditions, among other factors, determine to a great extent the system's operation. Evidence of this behavior is the seasonal operation found in energy storage systems: large hydro units reservoirs, for instance, are often guided by weekly, monthly, or yearly instructions that adapt to these variations. Furthermore, BESS usually keeps daily cycles (a cycle in this context is understood as the time that the capacity of a storage system takes to completely charge and discharge). If the objective is to represent the medium-term operation correctly, the model must adequately consider the different idiosyncrasies present in electricity systems, such as seasonality at multiple scales. To achieve that, the methodology proposed in this paper starts from selecting a control variable that holds the information about all these patterns of the fundamentals. The hourly observations of this variable are then clustered into periods (typically weeks or months), types of days (clusters of days within a period), and system states (clusters of hours within a day). This new temporal structure is subsequently applied to the rest of the parameters of the fundamental model (Fig. 1).

The methodology is broken down into the following steps:

Step 1: Define the control variable

Initially, set the time series $X = \{x_1, \ldots, x_{hr}, \ldots, x_{HR}\}$ as the control variable in which the clustering process is applied, where x_{hr} is the observation at hour hr, and HR is the total number of hours throughout the whole horizon of the problem. This selection is crucial to have a good view of the operation of the system.

Traditionally, the demand dynamics, which presents daily, weekly and annual seasonality, drove the market outcome. In this situation, demand would usually be selected as the control variable. However, this changed the moment non-dispatchable sources of electricity started to appear. For instance, market operation in hours with low demand and high wind generation is very different from the operation in hours with low demand and no presence of wind. Therefore, it is necessary to select a control variable that better represents the operation of the market when these sources of generation are increasingly predominant. The solution to this problem is to define the so-called net demand as the control variable. The net demand is calculated as the part of the demand that is covered by dispatchable generation, so its value depends on the main factors that drive the operational variations present in the system. This is obviously different for each electricity system. Nevertheless, the design of the proposed methodology makes it highly configurable, so this term can be defined in every system depending on the usual operation of the existing technologies. Because net demand contains all the idiosyncrasies that guide the operation of the existing technologies in electricity systems, it is reasonable to use this variable as the basis to reduce the set of periods.

Step 2: Divide the time series into periods

Once the control variable has been selected, the whole horizon of the problem is divided into consecutive time periods. Periods are predefined depending on the nature of the system and the specific aspects that regulate it. Weeks or months are typically chosen as these periods since hydro operation cycles tend to follow a similar behavior. In the following, periods will be particularized to months for the sake of clarity.

For each month m, let Y_m be the m-partition of X with cardinality M_m, containing the time series observations $y_n \in X,\ n = 1, \ldots, M_m$ corresponding to the hours of month m.

Step 3: Normalize the monthly time series

Normalize the monthly time series Y_m into $\hat{Y}_m$, such that each data point of the original time series y_n is transformed in $\hat{y}_n$ according to $\hat{y}_n = \frac{y_n}{max(y_j : j=1,\ldots,M_m)}$.

This step is critical when dealing with data from multiple areas. In such a case, normalization ensures that data from larger areas are not over-represented.

Step 4: Arrange the monthly time series into daily vectors

Arrange the normalized monthly time series $\hat{Y}_m$ in daily vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_D\}$, where D is the number of days in month m, and each vector $\mathbf{v}_i \in \mathbb{R}^{24}$ corresponds to a day.

After the normalization, the monthly time series is organized by days as preparation for the next step.

Step 5: Aggregate days into types of days

Find K clusters $TD = \{TD_1, \ldots, TD_K\}$ and its corresponding cluster centers $\{\mathbf{td}_1, \ldots, \mathbf{td}_K\} \in \mathbb{R}^{24}$ (hereinafter referred as centroids or "types of days") by means of a k-means clustering algorithm, so that the Euclidean distance from each vector $\mathbf{v}_i$ to its respective cluster centroid $\mathbf{td}_j$ is minimized.

This stage consists of grouping the days belonging to each month according to their similarity. For this step, a vector-based k-means clustering method will be implemented. Selecting an appropriate value of clusters is a difficult task. The lower the number of types of days, the lower the complexity of the model but the worse the accuracy of the results. Therefore, it is required to establish reasonable criteria to propose an optimal number of types of days, depending on our needs. More details about the measures used are found in Section 3.

Step 6: Assign each day to a type of day

Assign each vector $\mathbf{v}_i$ to the closest cluster TD_j, such that the membership of each vector $\pi(\mathbf{v}_i)$ can be stated as: $\pi(\mathbf{v}_i) = \mathrm{argmin}_{\mathbf{td}_1 \ldots \mathbf{td}_K} \sum_{j=1}^{K} \|\mathbf{v}_i - \mathbf{td}_j\|^2$.

Once all types of days have been calculated, there is a direct correspondence between days and types of day, in the sense that every day is assigned to a type of day.

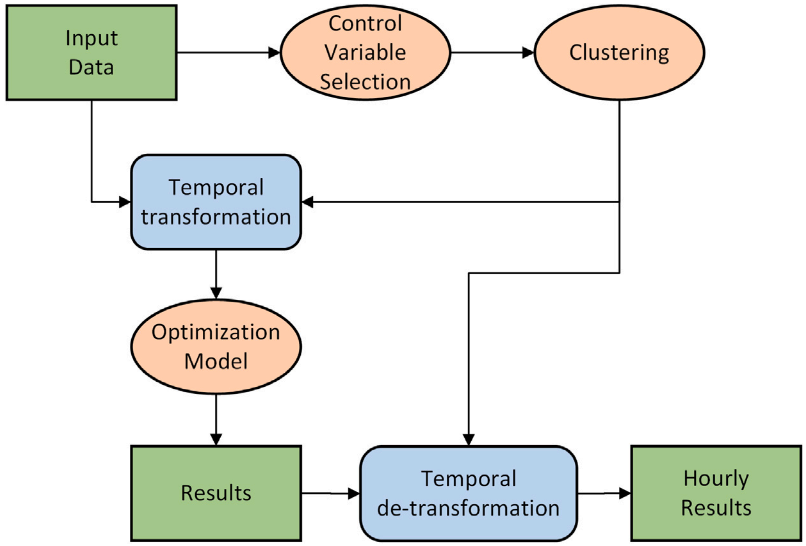

Fig. 1. Diagram of the proposed methodology.

Step 7: Aggregate types of days into system states

For each type of day $\mathbf{td}_j = \{td_1, \ldots, td_{24}\}$, apply again a clustering process to find N clusters of hours $ST = \{ST_1, \ldots, ST_N\}$, also known as "system states", and its corresponding cluster centers $\{st_1, \ldots.., st_N\} \in \mathbb{R}$.

Likewise, a second round of clustering will be carried out to find system states within the types of days. According to this procedure, the hours belonging to the centroids of the types of days will be entered in a one-dimensional k-means clustering algorithm to extract the system states. This ensures that all days represented by every type of day are considered equal in the modeling structure. As a consequence of this clustering process, each hour of the model is assigned to a system state. As abovementioned, which number of system states should be selected depends on the specific characteristics of the energy system model as well as the intended accuracy of the results and the available computational resources.

Note that although k-means clustering has been selected as the most suitable grouping method to be used in this paper, the contribution of this work is not centered on k-means, which is widely used in the literature, but on the complete clustering methodology that has been developed around this procedure, and how it is able to obtain temporal aggregation configurations adequate for medium-term market modeling.

2.2. Model description

The proposed methodology is implemented using a specific formulation in a market equilibrium model. This section is intended for providing a detailed description of this formulation and how it can be integrated with the clustering structure presented in Section 2.1.

The proposed formulation represents an extension of the conjectural variation (CV) based model developed in [25], in such a way that it incorporates the designed temporal aggregation into the objective function as well as long- and short-term storage. Because market equilibrium models are able to accurately capture market behavior, this approach represents the most realistic way to model operational planning in liberalized electricity systems. In this context, each market agent e aims to maximize its profit with its production $Q_{e,p}$ at every period. As demonstrated in [25], the market equilibrium problem can be stated as the following equivalent quadratic minimization problem:

$$\min_{Q_{e,p}} \sum_{e,p} \left(C_{e,p}\left(Q_{e,p}\right) + \frac{\theta_{e,p}}{2} Q_{e,p}^2 \right) \tag{1}$$

subject to:

$$\sum_{e} Q_{e,p} = D_p : \lambda_p \qquad \forall p \tag{2}$$

$$\mathcal{H}\left(Q_{e,p}\right) \geq 0 \qquad \forall e,p \tag{3}$$

$C_{e,p}\left(Q_{e,p}\right)$ represents the cost function for agent e and includes the total operational and maintenance costs. The conjectured-price response $\theta_{e,p}$ measures the change in price λ_p with respect to the production $Q_{e,p}$, as in (4):

$$\theta_{e,p} = -\frac{\partial \lambda_p}{\partial Q_{e,p}} \qquad \forall e,p \tag{4}$$

Eq. (2) is the demand balance constraint, and the rest of the technical and economic constraints are represented in (3). Henceforth, a specific formulation will be introduced to model the behavior of storage units.

Concerning short-term storage operation, constraints that model the operational behavior of BESS must be incorporated into the model.

$$\begin{aligned} SOC_{b,p,td,bl} &= SOC_{b,p,td,bl-1} \\ &+ \sum_{(hr,st)\in bl} (pc_{b,p,st} \cdot \rho c_b - \frac{pd_{b,p,st}}{\rho d_b}) \ \forall b,p,td \in p, bl \in td \end{aligned} \tag{5}$$

Eq. (5) describes the development of the intra-day state of charge of battery b (SOC_b), to maintain a feasible energy balance: the stored energy at the end of period p equals the stored energy at the end of the previous period plus all the charging and discharging (limited by their corresponding efficiencies) of the battery occurred during this period.

The representation of days by a set of system states introduced in this paper allows us to define (5) at just specific hours when there is a switch of state. Under this scheme, it is possible to divide the day into time blocks (formulated as bl), grouping consecutive hours assigned to the same system state. For instance, given the example structure in Fig. 2, the approximated hourly curve of Feb-11 can be divided into 7 time blocks: $[hr_1, hr_2 - hr_6, hr_7, hr_8 - hr_{13}, hr_{14} - hr_{18}, hr_{19} - hr_{21}, hr_{22} - hr_{24}]$. In general, the number of blocks will depend on the number of system states as well as how they transition throughout the day. In turn, these transitions result from the non-supervised clustering process that assigns every hour with a unique state. This approach reduces the number of equations considerably, and therefore, the computational complexity of the model, while maintaining an exact solution considering the time-defined aggregation.

Additionally, Eq. (6) establishes the daily cyclic condition of the BESS, assuming their state of charge is the same at the beginning of

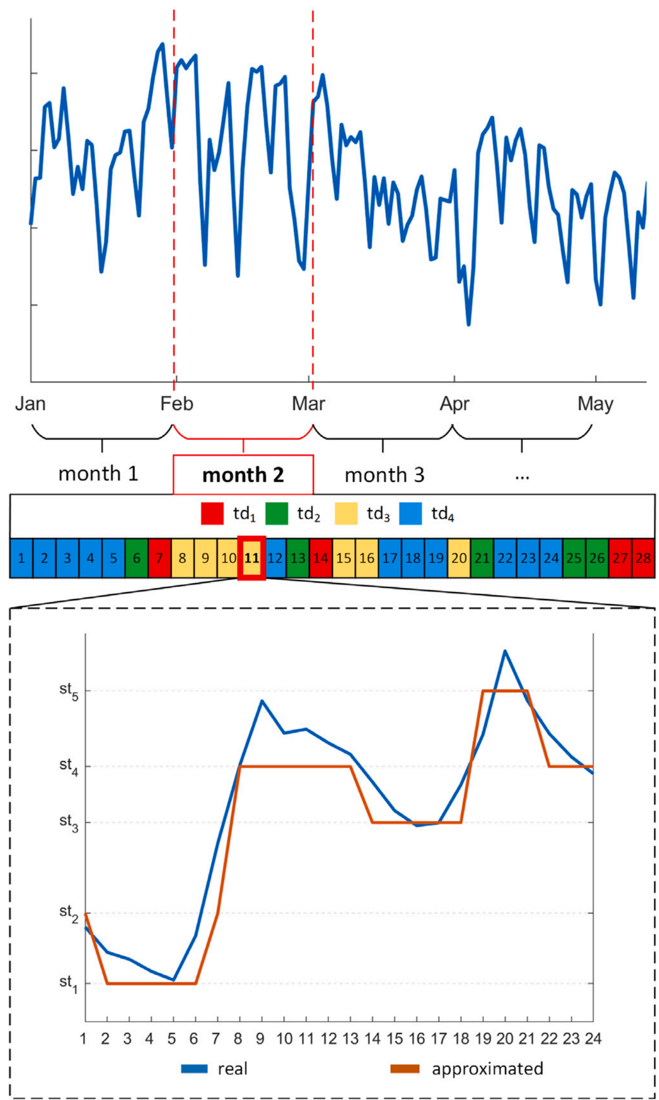

Fig. 2. Example of the clustering process following the temporal aggregation structure proposed in Section 2.1.

each day, and they are initially empty. This is a common assumption in batteries formulation and is often similarly represented in other works, such as Ref. [4].

$$\frac{\sum_{st \in td}((pc_{b,p,st} \cdot \rho c_b - pd_{b,p,st}/\rho d_b)dur_{st})}{Ndays_{p,td}} = 0 \ \forall b, p, td \in p \tag{6}$$

On this basis, BESS operation is not temporally coupled between different types of days. This assumption in line with the natural operation of BESS. According to the normal development of electricity prices, lower prices are found in the first hours of the day, corresponding with the night time. Therefore, it is precisely in this time frame when it makes the most sense to charge the batteries. This means that batteries should be mostly discharged at the end of the day in almost every real situation. Nevertheless, this methodology could be extended to accommodate another form of supraday operation. Additionally, regarding the technical representation of BESS, this model does not consider the battery lifespan. BESS degradation is mainly dependent upon frequency and depth of discharge. As this model is intended for medium-term power system planning, a simplified representation of BESS operation is adequate for this purpose. Anyhow, the formulation could be adapted to include battery degradation, as it is done in [5], for instance.

Finally, Eqs. (7)–(9) establish the lower and upper bounds of the variables that define the battery's operational design.

$$pc_{b,p,st} \le dr_b \qquad \forall b, p, st \tag{7}$$

$$pd_{b,p,st} \le dr_b \qquad \forall b, p, st \tag{8}$$

$$SOC_{b,p,td,bl} \le cap_b \qquad \forall b, p, td, bl \tag{9}$$

Moving on to the rest of the model's formulation, the total production of each market agent $Q_{e,p}$ can be separated into hydro production and thermal production. If the pumped water from reservoirs and the energy charged and discharged from the batteries are also considered, the following statement holds:

$$\begin{aligned} Q_{e,p} = \sum_{st} \Bigg(& \sum_{t:o_t=e} pt_{t,p,st} \\ & + \sum_{h:o_h=e} (ph_{h,p,st} - be_{h,p,st}) \\ & + \sum_{b:o_b=e} (pd_{b,p,st} - pc_{b,p,st}) \Bigg) dur_{st} \quad \forall e, p \end{aligned} \tag{10}$$

The rest of the constraints, such as commitment decisions, contracts, ramps, hydro reserves limitations, and other techno-economic details, were excluded from this section for simplicity since they do not represent an original contribution. Nevertheless, they were taken into account in the simulations performed for the case study in Section 3.

An overview of the complete methodology is depicted in Fig. 1. In this figure, the resulting aggregation determined by the clustering process is used for both transforming the inputs to the model as well as de-transforming the outputs to obtain the final hourly results.

3. Case study and results

This section presents a comprehensive real-size case study intended to illustrate the methodology proposed in this paper and provide a sensitivity analysis of the results to the possible temporal aggregation configurations. Note that the data and assumptions in this section should not be construed as an accurate representation of the system. They are used as an example of the application of the methodology.

3.1. System description

A real-size detailed representation of the Iberian electricity market (MIBEL) is used for this case study. This system consists of Spain and Portugal, where all thermal and hydro units are individually considered, including every technical and economic characteristic. Non-dispatchable generation is also taken into account, aggregated by technology for every hour. Furthermore, interconnections between areas are modeled too, including the representation of the electricity trade with external regions.

Regarding energy storage systems, this electricity system accounts for several pumped hydro storage plants, which are incorporated in the model. To analyze the performance of the proposed methodology under the current energy transition where the presence of storage units is growing significantly, a fast-ramping Li-ion BESS is also considered. This BESS is characterized by a total energy capacity of 10 MWh and a charging/discharging capacity of 2.5 MW, corresponding to a 4-hour discharge from the maximum capacity. The efficiency of both charging and discharging processes is assumed to be the same, equal to 0.9. Initially, the battery is assumed to be empty, and it holds a daily cycle.

Finally, the model covers all policies that regulate every aspect of the operation of the market.

In this regard, it is important to note that uncertainty is not included in this case study. Although the current market situation demands the inclusion of present uncertainty in the model to have accurate results, this integration falls outside the scope of this study. The case study presented in this section describes a sensitivity analysis based on a deterministic equilibrium model. In this sense, multiple tests are carried out to illustrate the methodology's performance in all possible configurations of the temporal structure of the problem.

3.2. Temporal aggregation configurations

The selected horizon for this case study is the year 2019, with an hourly resolution. Withal, temporal aggregation according to the methodology presented is implemented. The entire horizon is divided first into periods corresponding to months due to the monthly nature of the hydro management signals in this system. Subsequently, the hours of every month are aggregated into a clustering configuration. This clustering configuration is the same for every month and varies with each test performed. The input control variable used for the clustering process is the net demand, as explained in Section 2.1. Because nuclear generation in MIBEL is dispatched as baseload, and its production level keeps invariant most of the time, it will be subtracted to calculate the net demand. Thus, net demand would be equal to the demand minus the renewable and nuclear generation, minus the run-off-the-river. As previously mentioned, this definition may vary depending on the system being analyzed.

As a point of reference, a complete hourly resolution execution is carried out. This execution is characterized by a temporal configuration in which every month is represented by as many types of days as there are days in the month (a maximum of 31), and every type of day is modeled with 24 system states, corresponding to every hour. The obtained results are the benchmark against which the rest of the executions will be measured.

Apart from the hourly model, a total of 712 more model configurations are run. If every month of each execution is modeled in the same way, the number of combinations of types of days and system states represents the total number of executions. The 31 configurations with one system state are discarded as it is necessary to have at least two system states to have a feasible representation of the BESS constraint. Therefore, the number of executions of the model covers the total number of feasible combinations between types of days within each month and system states within these types of days. This wealth of executions of a real-size fundamental model allows us to better analyze the effect that the temporal reduction proposed in this paper brings to the results.

The machine used to run all these tests was a computer with Intel (R) Xeon (R) Silver 4116 CPU @2.10 GHz with 40 logical processors and 128 GB of installed RAM running 64-bit Windows Server 2019. The model was executed in GAMS 32.2.0, using the commercial solver CPLEX 12.10.0.0.

To compare the different temporal aggregation configurations, it is necessary to establish reasonable indicators. In this case study, the mean absolute error (MAE) and the mean absolute percentage error (MAPE) were used as the main criteria to compare the executions. In order to have a clear understanding of the model's results for each clustering configuration, the following variables are selected: thermal generation, hydropower production, non-dispatchable generation, electricity prices, and energy storage levels, both in the modeled reservoirs and the incorporated BESS. Overall, both metrics calculated across these variables will determine to a great extent the accuracy of the results of every test performed against the hourly realization.

3.3. Results

3.3.1. Size and CPU time

Firstly, since the methodology is intended to render computationally tractable models, it is worth looking at the model dimensions and CPU time in every test performed. This information is represented in Fig. 3, where the number of variables and equations, maximum RAM required, and CPU times are portrayed for the total number of system states modeled by month. To have a reference, the hourly model consists of 3,385,516 equations and 9,110,800 variables. It took 5299 s to run and required 148.8 GB of RAM., This CPU time may be a practical duration depending on the computational setting. However, in the medium term, it is usual to make probabilistic forecasts, for example, through Monte Carlo simulations. This process will usually require executing the same model hundreds of times under different scenarios of uncertain variables. In this context, CPU times of this magnitude are not suitable for the daily activities of an agent in his decision-making process. Therefore, although this article is centered on a deterministic application of the model, the proposed methodology becomes even more relevant when it is applied in a setting where uncertainty is considered. Uncertainty modeling has been extensively researched in the literature. A thorough review of different modeling techniques for the inclusion of risk factors in power systems models can be found in [31,32] . In this context, the most common approaches to model uncertainty in power systems are stochastic programming and Monte Carlo simulation. For instance, Ref. [33] develops a methodology for scenario tree reduction in multi-stage stochastic programming. On the other hand, Ref. [34] implements an efficient procedure to execute Monte Carlo simulations based on spatial interpolation. In the context of this paper, for the medium term and considering the size of the model presented here, stochastic programming would be difficult to apply and could result in a longer execution time. On the contrary, Monte Carlo simulation allows performing scenario analysis under multiple uncertain factors with a high degree of complexity to compute statistical outputs. As aforementioned, this is very valuable information for the decision-making process of a market agent. Despite the potential benefits of this methodology in a probabilistic framework, note that the application of a Monte Carlo simulation is out of the scope of this paper.

According to the formulation in Section 2.2, the bigger the total number of system states, the bigger the model and its complexity. It can be observed that, in general, both the number of variables and equations increase linearly with the number of system states, while the memory shows a quadratic growth and the CPU time seems to increase exponentially.

3.3.2. Objective function

To keep the model computable and manageable, a small number of system states are needed. However, reducing the number of states has a detrimental effect on the accuracy of the results. This impact is represented in Fig. 4, where the absolute percentage error of the objective function is drawn.

As can be seen, the objective function error decreases as the number of system states increases. More specifically, this decrease retains an exponential trend. Hence, adding more states will dramatically reduce the error when a small number of system states are defined but slowly decrease it when more states are modeled. This allows getting a reasonable accuracy when just a few system states are modeled. Exponential decay is also detected when looking at the error curves for each number of types of days. Also, Fig. 4 shows that with the minimum number of types of days considered (1 type of day), the error is always above $1e-3\%$. However, if we slightly increase this number of types of days, the error decreases significantly. For instance, with five types of days, the error is lower than $4e-4\%$

3.3.3. Productions, reserves, and prices

Whereas the objective function value provides a good measure of the methodology's effectiveness, it is worth analyzing the errors found in the results for the main variables of the model to have a more meaningful comparison. These results are depicted in Fig. 5. While the same general trend is noticed across all variables, the sensitivity of energy storage levels to the temporal configuration for the BESS deserves special attention.

When paying close attention to the sensitivity curves in Fig. 5, the hourly evolution of the state of charge of the battery appears to be more sensitive to a better representation of every type of day, rather than to an increase in the number of types of days. Alternatively, the opposite occurs in the case of productions, where a steeper decay in error is identified with increments in the number of types of days. This

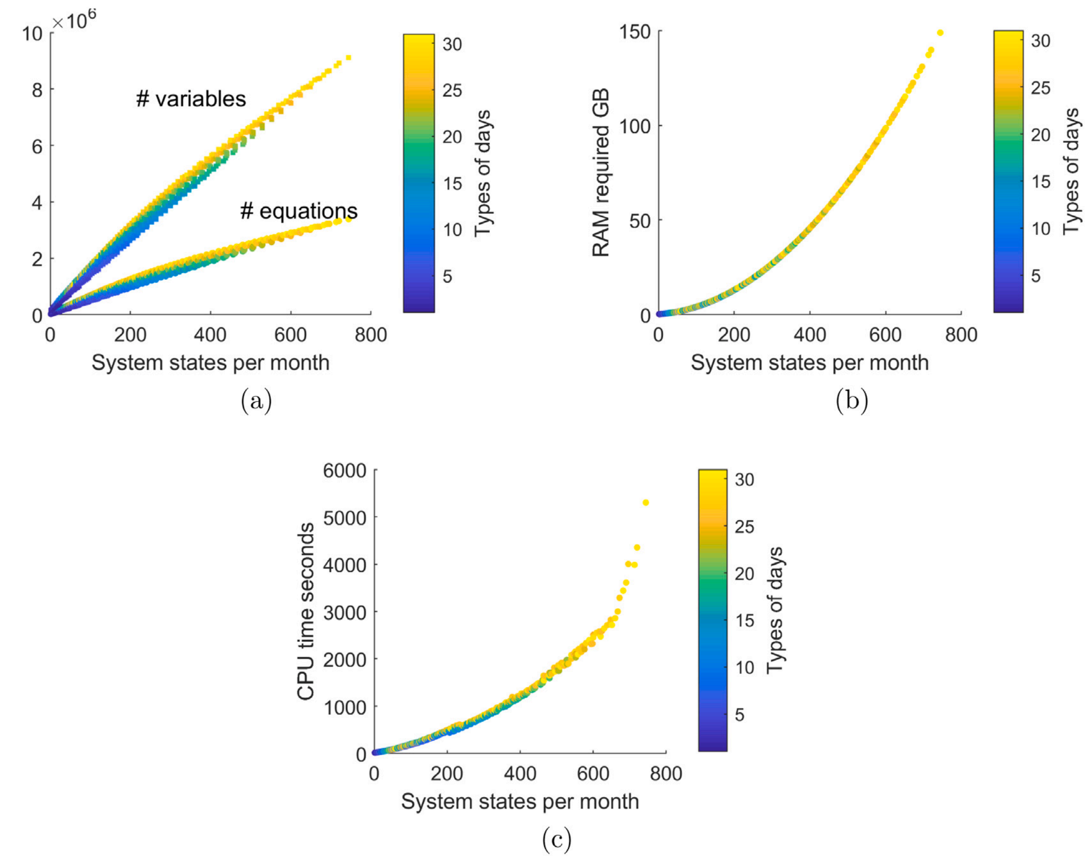

Fig. 3. Number of equations and variables 3(a), RAM required 3(b) and CPU time taken 3(c) for each execution with respect to the number of system states per month represented. Every point corresponds to a realization of the model for a specific clustering configuration.

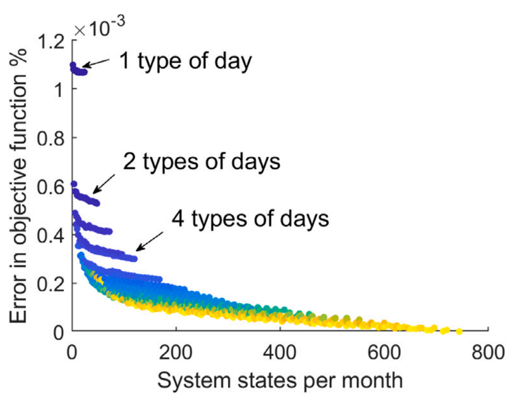

Fig. 4. Evolution of the objective function error with respect to the number of systems states modeled per month.

Table 1
MAE sensitivity for the possible configurations with 48 total system states per month.

td	$st \in td$	MAPE pt_t %	MAE SOC_b %
2	24	12.86	14.76
3	16	10.44	14.99
4	12	9.40	15.78
6	8	8.53	16.87
8	6	7.77	18.54
12	4	6.85	21.32
16	3	6.46	23.57
24	2	6.15	31.14

idea can be better appreciated in the sensitivity curves in Fig. 6(a) and Fig. 6(b).

Therefore, as far as the battery charge level is concerned, the model performs better when more daily system states are represented for a fixed number of total system states. For aggregated productions, the opposite holds true. Table 1 shows an error comparison for all the configurations with the same total number of system states, equal to 48. The results can be explained by the variability of the output being analyzed. In particular, the daily cycle of batteries is characterized by a significant level of fluctuation when integrated into power systems with high shares of renewables, as is the case. Therefore, the more system states are used, the more granularity is achieved. This temporal aggregation structure makes it possible to model this kind of variability correctly.

Alternatively, variables with a very flat profile are less sensitive to the defined system states since just a few states are necessary to follow a low-variability profile.

Regarding electricity prices, the methodology performs accurately, even when a high temporal aggregation is defined. Overall, by increasing the granularity in types of days, the model yields more accurate electricity prices than by increasing the time resolution within each day.

4. Conclusions

This paper proposes a novel methodology that reduces the temporal dimension of medium- and long-term power system models. The primary purpose is to achieve computationally tractable problems while yielding accurate results.

Focused on the correct characterization of the variability of the net demand, a clustering process has been implemented to transform the

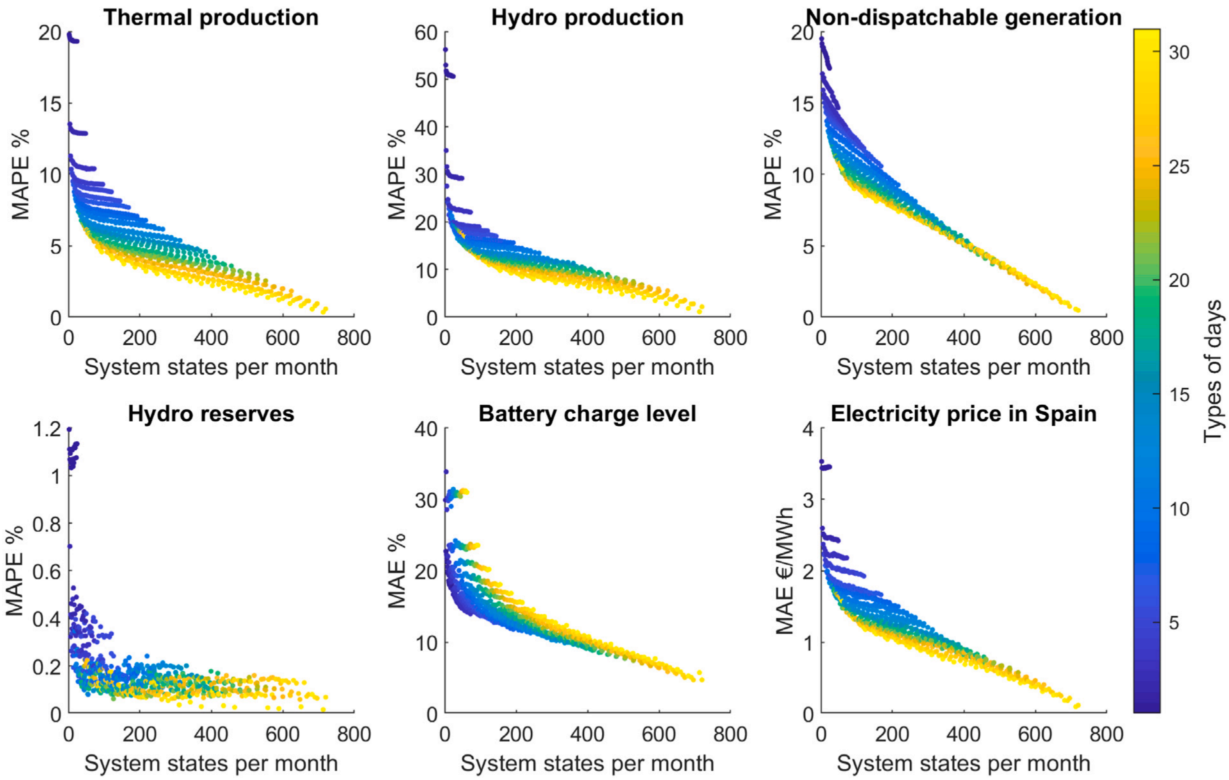

Fig. 5. MAE for every analyzed output against the system states modeled per month. Note that x-axis retain the same form as in Fig. 3 to have a clear understanding of the trade-off between computational effort (CPU time and RAM memory) and solution quality.

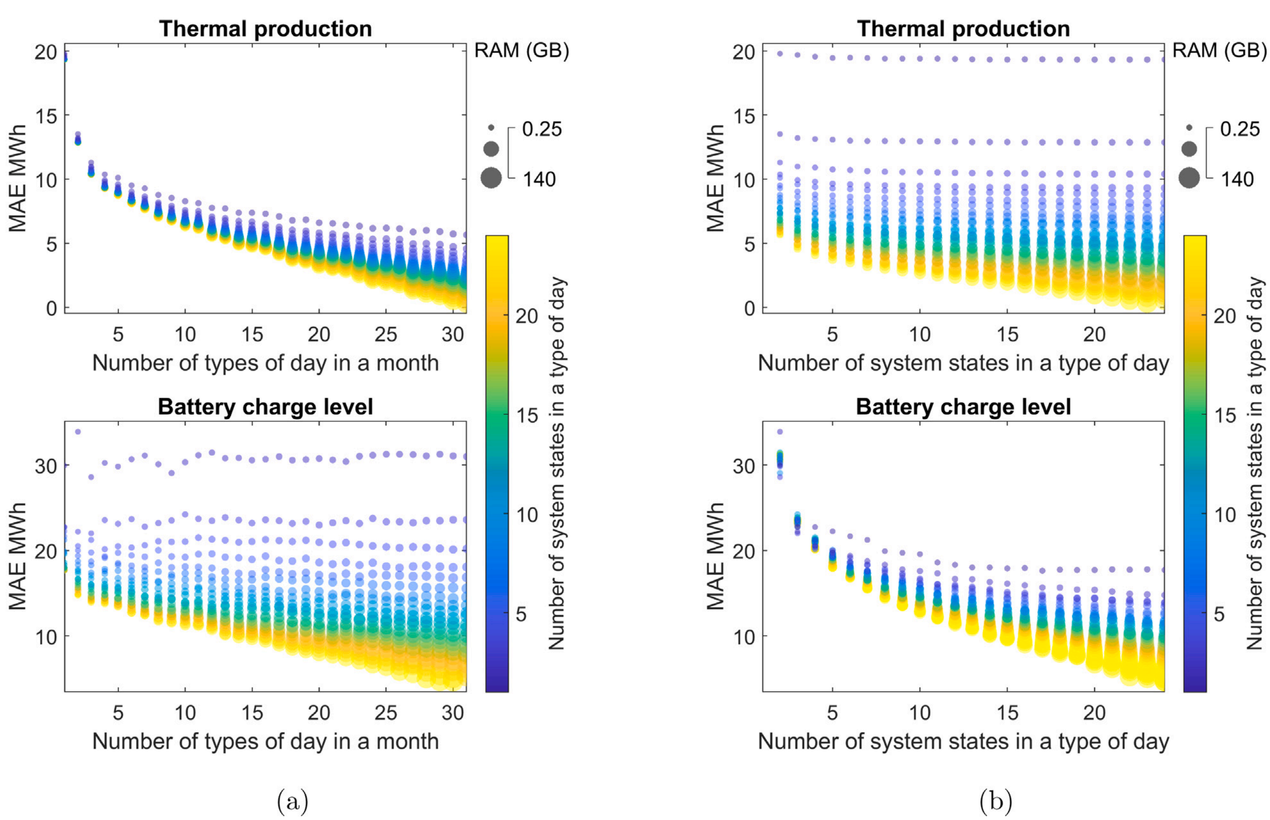

Fig. 6. MAE of thermal production and battery charge level obtained for every number of types of day represented in a month 6(a) and for every number of systems states modeled in a day 6(b).

temporal structure of the input parameters into periods, types of days, and system states. This arrangement is flexible and highly configurable. At the same time, incorporating an efficient formulation of the operation of energy storage systems following the methodology makes a real-size problem manageable and computationally not intensive. Following the proposed approach, it is possible to determine an accurate dispatch of the system's units, including the careful management of backup gas power plants and short- and long-term energy storage technologies. Ultimately, this makes it possible to capture power systems' intricacies and realistically represent the outcomes.

This methodology has been tested through its application to a real-size case study of the MIBEL electricity market. In general, this methodology has proven to effectively capture both short- and medium-term variability in real power systems. Following the performed sensitivity analysis, it can be seen how the level of accuracy is highly dependent on the selected temporal aggregation configuration. Hence, it is challenging to opt for a specific one. This selection will be limited by the available computational resources and the desired level of accuracy in the results. Still, results show that a very reasonable solution will be obtained even with a somewhat simplified temporal structure, such as one consisting of five types of days and four states within each day. This configuration represents a good trade-off between the accuracy of the results and the computational complexity.

Although the methodology proposed in this paper has been applied to a specific case study, it would be equally valid and applicable to other medium-term market models in this context. In this sense, the sensitivity results should be similar.

CRediT authorship contribution statement

Alberto Orgaz: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing, Writing – review & editing, Visualization. **Antonio Bello:** Conceptualization, Methodology, Validation, Resources, Writing – review & editing, Supervision. **Javier Reneses:** Conceptualization, Methodology, Resources, Writing – review & editing, Supervision.

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.

References

[1] UNFCCC C. Paris agreement. 2015, FCCCC/CP/2015/L. 9/Rev. 1. URL https://unfccc.int/resource/docs/2015/cop21/eng/l09r01.pdf.

[2] Hirth L, Ziegenhagen I. Balancing power and variable renewables: Three links. Renew Sustain Energy Rev 2015;50:1035–51. http://dx.doi.org/10.1016/j.rser.2015.04.180.

[3] Ulbig A, Andersson G. Analyzing operational flexibility of electric power systems. Int J Electr Power Energy Syst 2015;72:155–64. http://dx.doi.org/10.1016/j.ijepes.2015.02.028.

[4] Khojasteh M, Faria P, Vale Z. Energy-constrained model for scheduling of battery storage systems in joint energy and ancillary service markets based on the energy throughput concept. Int J Electr Power Energy Syst 2021;133:107213. http://dx.doi.org/10.1016/j.ijepes.2021.107213.

[5] Zhang Y, Xu Y, Yang H, Dong ZY, Zhang R. Optimal whole-life-cycle planning of battery energy storage for multi-functional services in power systems. IEEE Trans Sustain Energy 2020;11(4):2077–86. http://dx.doi.org/10.1109/TSTE.2019.2942066.

[6] Deng Y, Zhang Y, Luo F, Mu Y. Operational planning of centralized charging stations utilizing second-life battery energy storage systems. IEEE Trans Sustain Energy 2021;12(1):387–99. http://dx.doi.org/10.1109/TSTE.2020.3001015.

[7] Ringkjøb H-K, Haugan PM, Solbrekke IM. A review of modelling tools for energy and electricity systems with large shares of variable renewables. Renew Sustain Energy Rev 2018;96:440–59. http://dx.doi.org/10.1016/j.rser.2018.08.002.

[8] Tejada-Arango DA, Domeshek M, Wogrin S, Centeno E. Enhanced representative days and system states modeling for energy storage investment analysis. IEEE Trans Power Syst 2018;33(6):6534–44. http://dx.doi.org/10.1109/TPWRS.2018.2819578.

[9] Green R, Staffell I, Vasilakos N. Divide and conquer? *k*-means clustering of demand data allows rapid and accurate simulations of the british electricity system. IEEE Trans Eng Manage 2014;61(2):251–60. http://dx.doi.org/10.1109/TEM.2013.2284386.

[10] Nahmmacher P, Schmid E, Hirth L, Knopf B. Carpe diem: A novel approach to select representative days for long-term power system modeling. Energy 2016;112:430–42. http://dx.doi.org/10.1016/j.energy.2016.06.081.

[11] Flores-Quiroz A, Palma-Behnke R, Zakeri G, Moreno R. A column generation approach for solving generation expansion planning problems with high renewable energy penetration. Electr Power Syst Res 2016;136:232–41. http://dx.doi.org/10.1016/j.epsr.2016.02.011.

[12] Poncelet K, Delarue E, Six D, Duerinck J, D'haeseleer W. Impact of the level of temporal and operational detail in energy-system planning models. Appl Energy 2016;162:631–43. http://dx.doi.org/10.1016/j.apenergy.2015.10.100.

[13] Gabrielli P, Gazzani M, Martelli E, Mazzotti M. Optimal design of multi-energy systems with seasonal storage. Appl Energy 2018;219:408–24. http://dx.doi.org/10.1016/j.apenergy.2017.07.142.

[14] Pfenninger S. Dealing with multiple decades of hourly wind and PV time series in energy models: A comparison of methods to reduce time resolution and the planning implications of inter-annual variability. Appl Energy 2017;197:1–13. http://dx.doi.org/10.1016/j.apenergy.2017.03.051.

[15] Kotzur L, Markewitz P, Robinius M, Stolten D. Impact of different time series aggregation methods on optimal energy system design. Renew Energy 2018;117:474–87. http://dx.doi.org/10.1016/j.renene.2017.10.017.

[16] Gonzato S, Bruninx K, Delarue E. Long term storage in generation expansion planning models with a reduced temporal scope. Appl Energy 2021;298:117168. http://dx.doi.org/10.1016/j.apenergy.2021.117168.

[17] Bahl B, Kümpel A, Seele H, Lampe M, Bardow A. Time-series aggregation for synthesis problems by bounding error in the objective function. Energy 2017;135:900–12. http://dx.doi.org/10.1016/j.energy.2017.06.082.

[18] Tejada-Arango DA, Wogrin S, Centeno E. Representation of storage operations in network-constrained optimization models for medium- and long-term operation. IEEE Trans Power Syst 2018;33(1):386–96. http://dx.doi.org/10.1109/TPWRS.2017.2691359.

[19] Domínguez-Muñoz F, Cejudo-López JM, Carrillo-Andrés A, Gallardo-Salazar M. Selection of typical demand days for CHP optimization. Energy Build 2011;43(11):3036–43. http://dx.doi.org/10.1016/j.enbuild.2011.07.024.

[20] Teichgraeber H, Brandt AR. Clustering methods to find representative periods for the optimization of energy systems: An initial framework and comparison. Appl Energy 2019;239:1283–93. http://dx.doi.org/10.1016/j.apenergy.2019.02.012.

[21] Hoffmann M, Kotzur L, Stolten D, Robinius M. A review on time series aggregation methods for energy system models. Energies 2020;13(3):641. http://dx.doi.org/10.3390/en13030641.

[22] Ortiga J, Bruno JC, Coronas A. Selection of typical days for the characterisation of energy demand in cogeneration and trigeneration optimisation models for buildings. Energy Convers Manage 2011;52(4):1934–42. http://dx.doi.org/10.1016/j.enconman.2010.11.022.

[23] Baumgärtner N, Bahl B, Hennen M, Bardow A. RiSES3: Rigorous synthesis of energy supply and storage systems via time-series relaxation and aggregation. Comput Chem Eng 2019;127:127–39. http://dx.doi.org/10.1016/j.compchemeng.2019.02.006.

[24] Teichgraeber H, Lindenmeyer CP, Baumgärtner N, Kotzur L, Stolten D, Robinius M, et al. Extreme events in time series aggregation: A case study for optimal residential energy supply systems, URL https://arxiv.org/pdf/2002.03059.

[25] Barquin J, Centeno E, Reneses J. Medium-term generation programming in competitive environments: a new optimisation approach for market equilibrium computing. IEE Proc, Gener Transm Distrib 2004;151(1):119. http://dx.doi.org/10.1049/ip-gtd:20040055.

[26] Murphy FH, Smeers Y. Generation capacity expansion in imperfectly competitive restructured electricity markets. Oper Res 2005;53(4):646–61. http://dx.doi.org/10.1287/opre.1050.0211.

[27] Wogrin S, Duenas P, Delgadillo A, Reneses J. A new approach to model load levels in electric power systems with high renewable penetration. IEEE Trans Power Syst 2014;29(5):2210–8. http://dx.doi.org/10.1109/TPWRS.2014.2300697.

[28] Wogrin S, Galbally D, Reneses J. Optimizing storage operations in medium- and long-term power system models. IEEE Trans Power Syst 2016;31(4):3129–38. http://dx.doi.org/10.1109/TPWRS.2015.2471099.

[29] Pineda S, Morales JM. Chronological time-period clustering for optimal capacity expansion planning with storage. IEEE Trans Power Syst 2018;33(6):7162–70. http://dx.doi.org/10.1109/TPWRS.2018.2842093.

[30] Wogrin S, Tejada-Arango DA, Pineda S, Morales JM. What time-period aggregation method works best for power system operation models with renewables and storage? In: 2019 international conference on smart energy systems and technologies (SEST). IEEE; 2019, p. 1–6. http://dx.doi.org/10.1109/SEST.2019.8849027.

[31] Aien M, Hajebrahimi A, Fotuhi-Firuzabad M. A comprehensive review on uncertainty modeling techniques in power system studies. Renew Sustain Energy Rev 2016;57:1077–89. http://dx.doi.org/10.1016/j.rser.2015.12.070.

[32] Soroudi A, Amraee T. Decision making under uncertainty in energy systems: State of the art. Renew Sustain Energy Rev 2013;28:376–84. http://dx.doi.org/10.1016/j.rser.2013.08.039.

[33] Dupačová J, Gröwe-Kuska N, Römisch W. Scenario reduction in stochastic programming. Math Program 2003;95(3):493–511. http://dx.doi.org/10.1007/s10107-002-0331-0.

[34] Dueñas P, Reneses J, Barquín J. Dealing with multi-factor uncertainty in electricity markets by combining Monte Carlo simulation with spatial interpolation techniques. IET Gener Transm Distrib 2011;5(3):323–31. http://dx.doi.org/10.1049/iet-gtd.2010.0264, URL https://digital-library.theiet.org/content/journals/10.1049/iet-gtd.2010.0264.