


Chronological Time-Period Clustering for Optimal Capacity Expansion Planning With Storage

Salvador Pineda , *Member, IEEE*, and Juan M. Morales , *Senior Member, IEEE*

Abstract—To reduce the computational burden of capacity expansion models, power system operations are commonly accounted for in these models using representative time periods of the planning horizon such as hours, days, or weeks. However, the validity of these time-period aggregation approaches to determine the capacity expansion plan of future power systems is arguable, as they fail to capture properly the mid-terms dynamics of renewable power generation and to model accurately the operation of electricity storage. In this paper, we propose a new time-period clustering method that overcomes the aforementioned drawbacks by maintaining the chronology of the input time series throughout the whole planning horizon. Thus, the proposed method can correctly assess the economic value of combining renewable power generation with interday storage devices. Numerical results from a test case based on the European electricity network show that our method provides more efficient capacity expansion plans than existing methods while requiring similar computational needs.

Index Terms—Time-period aggregation, capacity expansion, clustering techniques, energy storage, renewable power generation.

NOMENCLATURE

A. Indexes and Sets

- g Generation technology index.
- \mathcal{G}^r Subset of renewable generation technologies.
- n Bus index.
- s Storage technology index.
- t Time period index.

B. Parameters

- η_s Energy capacity of storage technology s (h).
- κ Renewable share target (%).
- ξ_s Round-trip efficiency of storage technology s (p.u.).
- ρ_{gnt} Capacity factor of technology g at bus n and time t (p.u.).
- τ_t Duration of time period t (h).
- a_n Yearly availability of hydro energy at bus n (p.u.).

Manuscript received January 8, 2018; revised April 16, 2018; accepted May 25, 2018. Date of publication May 30, 2018; date of current version October 18, 2018. This work was supported in part by the Spanish Ministry of Economy, Industry and Competitiveness through Projects ENE2016-80638-R and ENE2017-83775-P and in part by the Research Funding Program for Young Talented Researchers of the University of Málaga through Project PPIT-UMA-B1-2017/18. Paper no. TPWRS-00039-2018. (Corresponding author: Salvador Pineda.)

S. Pineda is with the Department of Electrical Engineering, University of Malaga, Malaga 29071, Spain (e-mail: spinedamorente@gmail.com).

J. M. Morales is with the Department of Applied Mathematics, University of Malaga, Malaga 29071, Spain (e-mail: juanmi82mg@gmail.com).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TPWRS.2018.2842093

- \bar{b}_{sn}^0 Initial capacity of storage s at bus n (MW).
- c_g Linear cost parameter of technology g (€/MWh).
- d_{nt} Demand level at bus n and time period t (MW).
- \bar{f}_{nm}^0 Initial capacity of transmission line nm (MW).
- \hat{f}_{nm} Maximum capacity of transmission line nm (MW).
- \bar{h}_{nt}^0 Initial hydro capacity at bus n (MW).
- \hat{h}_n Maximum hydro power capacity at bus n (MW).
- i_g^F Investment cost of transmission lines (€/MW·Km).
- i_g^G Investment cost of technology g (€/MW).
- i_g^H Investment cost of hydro power capacity (€/MW).
- i_s^S Investment cost of storage s (€/MW).
- l_{nm} Length of transmission line nm (Km).
- \bar{p}_{gn}^0 Initial capacity of technology g at bus n (MW).
- \bar{p}_{gn} Maximum capacity of technology g at bus n (MW).
- r_g^- Relative ramp-down limit of technology g (p.u.).
- r_g^+ Relative ramp-up limit of technology g (p.u.).
- sc_n Load shedding cost at bus n (€/MWh).
- w_t Weight of time period t .
- y_g^F Expected lifetime of transmission lines (years).
- y_g^G Expected lifetime of technology g (years).
- y_g^H Expected lifetime of hydro power units (years).
- y_s^S Expected lifetime of storage s (years).

C. Variables

- b_{snt}^+ Charge of storage s at bus n and time t (MW).
- b_{snt}^- Discharge of storage s at bus n and time t (MW).
- b_{snt} Energy level of storage s at bus n and time t (MWh).
- \bar{b}_{sn} Additional capacity of storage s at bus n (MW).
- d_{nt} Satisfied demand at bus n and time t (MW).
- f_{nmt} Power flow through line nm at time t (MW).
- \bar{f}_{nm} Additional transmission capacity of line nm (MW).
- h_{nt} Hydro power generation at bus n , time t (MW).
- \bar{h}_n Additional hydro capacity at bus n (MW).
- p_{gnt} Generation of technology g at bus n and time t (MW).
- \bar{p}_{gn} Additional capacity of technology g at bus n (MW).

I. INTRODUCTION

THE expansion of the generation and transmission capacity of a power system is a long-term planning problem that spans several years [1] and that must account for the short-term operating requirements and constraints. Since an hourly time representation of the whole planning horizon would render the capacity expansion model computationally intractable, most energy planning models consider a stylized temporal representation [2], [3].

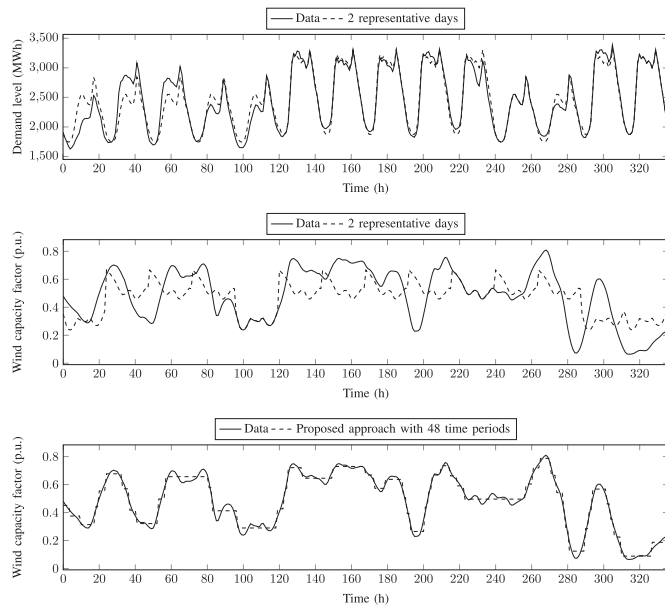


Fig. 1. Time aggregation for demand and wind capacity factor.

Probably the simplest approach consists in characterizing the time-dependent parameters, such as the electricity demand, using duration curves that are approximated by a limited amount of blocks [4], [5]. Despite its computational efficiency, this time representation completely disregards the chronology of the varying parameters and therefore, it does not allow for the inclusion of inter-temporal constraints (e.g., generation ramping limits) into the capacity expansion model.

A more sophisticated and broadly used time representation relies on the selection of a reduced set of representative days or weeks from the planning horizon [6]. Unlike the duration curves, the chronology of the varying parameters is preserved, at least, within each representative day or week and therefore, some inter-temporal constraints such as ramping limits can be considered. To illustrate this approach, the upper subplot of Fig. 1 represents the aggregated electricity demand in Denmark during the first two weeks of 2010 (solid line) and the approximation using only two representative days (dashed line). Once the two representative days are chosen, the data linked to each day is replaced by that corresponding to its closest representative day. Due to the strong daily pattern of electricity demand, this approach proves to be accurate in that case.

Representative days or weeks have been used in several generation and transmission capacity expansion models. Authors of [7] determine optimal capacity additions using a representative day for each month of the planning horizon. Reference [8] employs a generation investment planning model to evaluate the contribution of solar generation to resource adequacy using sample days. Authors of [9] propose a bilevel model to decide the optimal size and location of storage devices using representative days. The value of energy storage is estimated in [10] using representative weeks.

In most capacity expansion models, the representative days or weeks are chosen using classical clustering techniques such as K-means or hierarchical clustering [11], [12]. New methods to

select representative days in capacity expansion problems have been recently proposed in the technical literature. For instance, reference [13] provides a novel optimization-based approach to select representative periods. Similarly, the authors of [14] propose a modified hierarchical clustering procedure to choose a reduced set of representative days that retains important statistical features of the input data such as correlation.

However, today's power systems are undergoing major changes that may undermine the validity of current approaches to account for short-term operating conditions in long-term planning problems, namely, the large penetration of renewable power generation and the growing integration of storage technologies.

Indeed, unlike electricity demand, wind power production, for example, does not exhibit a strong daily pattern. Actually, it is not rare to observe 2–3 days of high wind followed by 2–3 days of low wind [10]. Obviously, this dynamic behavior cannot be properly captured by representative days, as illustrated in the middle subplot of Fig. 1. This figure shows the capacity factor of the aggregated wind power production in Denmark for the first two weeks of 2010 (solid line). Notice the poor approximation achieved if two representative days are considered (dashed line).

In parallel with the growth of fluctuating renewable power generation, the increasing role of energy storage in power systems also questions the use of representative days in capacity expansion models [15]–[17].

While there is a consensus on the benefits of exploiting *intraday* storage to alleviate the increased balancing needs created by renewable generation [18], [19], some authors have also started to argue in favor of the installation of *interday* storage in power systems with a high share of renewable power production [20]–[23]. This type of storage technologies become crucial to overcome the “dark calm” periods happening in central Europe, for example. These periods are characterized by prolonged high pressure weather conditions that reduce the wind power generation to a minimum, combined with clouds, fog or snow that hinder the solar power generation [24]. Keeping track of the energy level of interday storage devices in capacity expansion models is not possible using representative days [2], [25]. Consequently, most existing models only consider intraday storage technologies [9], [14], [16].

Furthermore, growing efforts have been dedicated to the development of computationally efficient methods for capturing the inter-temporal operation of storage. For example, the authors in [26] propose a heuristic system-state procedure to approximate the energy level of storage devices. However, this procedure can only be applied, up to date, for operational problems. Likewise, a smiling-based approach is developed in [17] to determine solely optimal investments in storage. An extension of the approach in [17] is, therefore, required for its application to greenfield systems. Besides, this approach only accounts for an operational horizon of 24 hours, which makes it unable to properly estimate the value of interday storage.

In this paper we propose a method to cluster the time periods of a capacity expansion model while keeping, as much as possible, the chronological information of the time-dependent parameters throughout the whole planning horizon. In doing so, the resulting capacity expansion model is capable of capturing

the longer dynamics of renewable power generation and properly modeling the energy conservation constraints of interday storage devices. In short, the proposed method performs a hierarchical clustering of consecutive hours according to conventional distance measures [11]. Just as an illustration, the bottom subplot of Fig. 1 depicts the time-period aggregation of the wind capacity factor in Denmark during the first two weeks of 2010 following the procedure proposed in this paper and considering 48 time periods (i.e., the same number of periods as in two representative days). It can be seen that our method approximates the variability of wind power production during the whole time horizon more accurately than the approach based on two representative days, which is displayed in the middle subplot.

The contributions of this paper are thus threefold:

- The proposal of a modified hierarchical clustering technique to aggregate consecutive time periods according to conventional distance measures.
- The formulation of a capacity expansion model that determines the optimal investment in conventional and renewable generation, intraday and interday storage technologies, and transmission facilities taking into account the long dynamics of time-dependent parameters.
- The comparison of the proposed time aggregation method with those most commonly used in the technical literature, which consist in selecting representative days or weeks using classical clustering techniques. We evaluate the performance of each approach in terms of computational time and accuracy of the investment decisions when compared with a benchmark model that does not aggregate time periods.

The rest of this paper is organized as follows. The modified hierarchical clustering technique whereby consecutive time periods are aggregated is described in Section II, while our capacity expansion model is formulated in Section III. A case study based on the European power system is used in Section IV to compare the different time-aggregation approaches. Finally, conclusions are duly drawn in Section V.

II. TIME-PERIOD CLUSTERING

Clustering consists in grouping a collection of objects into subsets or “clusters” such that the elements of a cluster are more similar to each other than to the elements belonging to a different cluster [11]. In the present subject-matter, clustering techniques are commonly used to select a reduced number of representative time periods (hours, days or weeks) in order to cut down the computational burden of long-term planning models [2], [6], [8], [13], [26]–[28]. Among the different techniques, hierarchical clustering is widely used because its outcome does not depend on the initialization of the algorithm. Unlike other clustering techniques such as K-means, hierarchical clustering also allows to readily include additional conditions on how clusters are merged. On the contrary, hierarchical clustering requires a measure of dissimilarity between two groups of observations and a linkage criterion to determine which clusters are merged or divided in agglomerative or divisive hierarchical clustering, respectively [11], [14], [29].

The time-period aggregation methods we consider in this paper, including the one we propose, are all based on Ward’s method for agglomerative hierarchical clustering [30]. In a few words, Ward’s method recursively merges the pair of clusters that minimally increases within-cluster variance. We explain below the clustering algorithms used by the time-period aggregation methods considered in this paper.

Let N be the total number of days/weeks of the planning horizon. Let \mathbf{x}_i be a vector containing the normalized values of all time-dependent parameters for day/week i . For example, if the demand and wind speed for two different locations are considered, the vector \mathbf{x}_i would have $2 \times 2 \times 24$ or $2 \times 2 \times 168$ elements for each day or week, respectively. The steps to select N' representative days/weeks using agglomerative hierarchical clustering are:

- 1) Set the initial number of clusters n to the total number of days/weeks N .
- 2) Determine the centroid $\bar{\mathbf{x}}_I$ of each cluster I as

$$\bar{\mathbf{x}}_I = \frac{1}{|I|} \sum_{i \in I} \mathbf{x}_i \quad (1)$$

- 3) Compute the dissimilarity between each pair of clusters I, J according to Ward’s method as follows [30]

$$D(I, J) = \frac{2|I||J|}{|I| + |J|} \|\bar{\mathbf{x}}_I - \bar{\mathbf{x}}_J\|^2 \quad (2)$$

- 4) Merge the two closest clusters (I', J') according to the dissimilarity matrix, i.e., $(I', J') \in \operatorname{argmin} D(I, J)$ s.t. $I \neq J$.
- 5) Update $n \leftarrow n - 1$.
- 6) If $n = N'$ go to step 7). Otherwise go to step 2).
- 7) Determine the set of representative days/weeks as the clusters’ medoids, i.e., the elements with the minimum dissimilarity to the rest of elements in each cluster.
- 8) The number of days/weeks belonging to each cluster corresponds to the value of the weight parameter w_t .

This agglomerative hierarchical clustering procedure is the one used to select two representative days out of the fourteen days considered in the upper and middle subplots of Fig. 1. It is also the method whereby we choose the representative days/weeks for the capacity expansion model described in Section III.

Next we present the proposed chronological time-period clustering method. Let N and N' be the initial and reduced number of time periods of the planning horizon. The steps to implement the proposed algorithm are the following:

- 1) Set the initial number of clusters n to the total number of hours N .
- 2) Determine the centroid $\bar{\mathbf{x}}_I$ of each cluster I using (1).
- 3) Compute the dissimilarity between each pair of adjacent clusters I, J according to Ward’s method using (2).
- 4) Merge the two closest *adjacent* clusters (I', J') according to the dissimilarity matrix, i.e., $(I', J') \in \operatorname{argmin} D(I, J)$ s.t. $J \in \mathcal{A}(I)$, where $\mathcal{A}(I)$ is the set of clusters adjacent to cluster I . Two clusters I and J are said to be adjacent if I contains an hour that is

consecutive to an hour in J , or vice versa, according to the original time series.

- 5) Update $n \leftarrow n - 1$.
- 6) If $n = N'$ go to step 7). Otherwise go to step 2).
- 7) Determine the set of representative periods as the clusters' centroids \bar{x}_I .
- 8) The number of hours belonging to each cluster corresponds to the value of the time-period duration τ_t .

This is the clustering algorithm used to aggregate the two-week wind capacity factor depicted in the bottom subplot of Fig. 1 using 48 time periods.

III. CAPACITY EXPANSION MODEL

In this section, we formulate a capacity expansion model that determines the optimal size, type and location of generation technologies and energy storage devices as well as the capacity of transmission lines in order to reach a given share of renewable power generation at the minimum cost. For the sake of simplicity, the proposed model provides capacity expansion decisions for a single target year [31]–[33]. The capacity expansion model is formulated as a deterministic optimization problem and therefore, uncertainty of input parameters is disregarded. Stochastic versions of this problem can be found in [34], [35]. In order to keep the computational tractability of the expansion planning model within reasonable limits, investment decisions are modeled as continuous variables. Assuming also a pipeline representation of the transmission network [14], the capacity expansion model is formulated as the following linear programming problem:

$$\begin{aligned} \min \quad & \sum_{gnt} \tau_t w_t c_g p_{gnt} + \sum_{nt} \tau_t w_t s c_n (\bar{d}_{nt} - d_{nt}) + \\ & \sum_{gn} \frac{i_g^G}{y_g^G} \bar{p}_{gn} + \sum_n \frac{i^H}{y^H} \bar{h}_n + \sum_{sn} \frac{i_s^S}{y_s^S} \bar{b}_{sn} + \sum_{nm} \frac{i^F l_{nm}}{y^F} \bar{f}_{nm} \end{aligned} \quad (3a)$$

s.t.

$$\sum_{g \in \mathcal{G}^r nt} \tau_t w_t p_{gnt} \geq \kappa \sum_{gnt} \tau_t w_t p_{gnt} \quad (3b)$$

$$0 \leq \bar{p}_{gn} \leq \hat{p}_{gn}, \quad \forall g \in \mathcal{G}^r, n \quad (3c)$$

$$0 \leq \bar{h}_n \leq \hat{h}_n, \quad \forall n \quad (3d)$$

$$0 \leq \bar{f}_{nm} \leq \hat{f}_{nm}, \quad \forall n, m \quad (3e)$$

$$\sum_g p_{gnt} + \sum_s (b_{snt}^- - b_{snt}^+) + h_{nt} = d_{nt} + \sum_m f_{nmt}, \quad \forall n, t \quad (3f)$$

$$0 \leq p_{gnt} \leq \rho_{gnt} (\bar{p}_{gn}^0 + \bar{p}_{gn}), \quad \forall g, n, t \quad (3g)$$

$$p_{gnt} - p_{gnt-1} \geq -r_g^- (\bar{p}_{gn}^0 + \bar{p}_{gn}), \quad \forall g, n, t \notin T_1 \quad (3h)$$

$$p_{gnt} - p_{gnt-1} \leq r_g^+ (\bar{p}_{gn}^0 + \bar{p}_{gn}), \quad \forall g, n, t \notin T_1 \quad (3i)$$

$$0 \leq h_{nt} \leq \bar{h}_n^0 + \bar{h}_n, \quad \forall n, t \quad (3j)$$

$$\sum_t \tau_t w_t h_{nt} \leq a_n \sum_t \tau_t w_t (\bar{h}_n^0 + \bar{h}_n), \quad \forall n \quad (3k)$$

$$0 \leq b_{snt}^+ \leq \bar{b}_{sn}^0 + \bar{b}_{sn}, \quad \forall s, n, t \quad (3l)$$

$$0 \leq b_{snt}^- \leq \bar{b}_{sn}^0 + \bar{b}_{sn}, \quad \forall s, n, t \quad (3m)$$

$$b_{snt} = b_{snt-1} - \tau_t b_{snt}^- + \tau_t \xi_s b_{snt}^+, \quad \forall s, n, t \notin T_1 \quad (3n)$$

$$0 \leq b_{snt} \leq \eta_s (\bar{b}_{sn}^0 + \bar{b}_{sn}), \quad \forall s, n, t \quad (3o)$$

$$b_{snt} = b_{snt'}, \quad \forall s, n, (t, t') \in T_2 \quad (3p)$$

$$0 \leq d_{nt} \leq \bar{d}_{nt}, \quad \forall n, t \quad (3q)$$

$$-\bar{f}_{nm}^0 - \bar{f}_{nm} \leq f_{nmt} \leq \bar{f}_{nm}^0 + \bar{f}_{nm}, \quad \forall n, m, t \quad (3r)$$

Objective function (3a) minimizes the total system cost over the planning horizon, which includes fuel cost (first term), load shedding cost (second term), and the costs of investing in conventional and renewable power capacity (third term), hydro power (fourth term), storage (fifth term) and transmission (sixth term). Constraint (3b) ensures that the renewable generation target κ is met. Equations (3c), (3d) and (3e) set the maximum installable capacity of renewable generation units, hydro power units, and transmission lines, respectively. Equation (3f) imposes the power balance for each bus of the network and time period of the planning horizon. Constraints (3g), (3h) and (3i) enforce the maximum capacity, the down-ramping limit, and the up-ramping limit of the generation technologies, respectively. The maximum generation from hydro power units is restricted by (3j), while constraint (3k) sets the maximum hydro energy available over the planning horizon. The charge and discharge of storage devices are constrained by (3l) and (3m), in that order. Equation (3n) computes the storage energy level for each time period using the round-trip efficiency [14], [26], which is the available data for most storage technologies [15]. The storage energy level for each time period is bounded by (3o). Constraint (3p) forces the storage level to be the same for the pairs of time periods contained in T_2 . Finally, the served demand and the power flows are limited through (3q) and (3r), respectively.

Generic formulation (3) is used to evaluate the performance of different time-period aggregation algorithms. To do so, we define the following four models:

- Full model, denoted by “F”. All hours of the planning horizon are considered and therefore, $\tau_t = w_t = 1, \forall t$. Constraints involving consecutive time periods such as (3h), (3i) and (3n) can be imposed for all time periods except the first one, i.e., $T_1 = \{1\}$. Likewise, the energy level of the storage devices is forced to be the same in the first and last hour of the year, which means that $T_2 = \{(1, 8760)\}$.
- Model based on representative days, which we denote by “D- N_D ”, where N_D is the number of representative days. Each of these days consists of 24 hourly values of the input data and thus $\tau_t = 1, \forall t$. However, the weight

of each representative day is determined according to the clustering algorithm described in Section II. For instance, if a cluster includes 20 days, the value of w_t corresponding to each hour of the representative day associated with that cluster is equal to 20. Since representative days are not necessarily consecutive, constraints (3h), (3i) and (3n) cannot be applied to the first hour of each of these days, i.e., $T_1 = \{1, 25, 49, \dots\}$. Besides, the storage level in the first and last hour of each representative day must be the same, that is, $T_2 = \{(1, 24), (25, 48), (49, 72), \dots\}$.

- Model based on representative weeks, denoted by “W- N_W ”, where N_W is the number of representative weeks. Following the same reasoning as for the model with representative days, we have that $\tau_t = 1, \forall t$; w_t is determined by the clustering procedure of Section II; $T_1 = \{1, 169, 337, \dots\}$, and $T_2 = \{(1, 168), (169, 336), (337, 504), \dots\}$.
- Chronological time-period aggregation model, which we refer to as “C- N_T ”, where N_T is the number of reduced time periods. This model is based on the aggregation of consecutive time periods. Therefore, $w_t = 1, \forall t$ and τ_t is determined by the clustering procedure presented in Section II. For example, if a cluster includes 20 hours, the value of τ_t corresponding to that period is equal to 20. Since chronology is maintained in this model, constraints involving consecutive time periods can be imposed throughout the whole planning horizon, i.e., $T_1 = \{1\}$. For the same reason, the energy level of the storage devices must be the same at the beginning and end of the planning horizon, i.e., $T_2 = \{(1, N_T)\}$.

Note that for all models above, $\sum_t \tau_t w_t = 8760$, which are the total number of hours of the planning horizon, that is, of the target year.

IV. CASE STUDY

Next we analyze the performance of the four models described in Section III using a test case in which we plan the European power system for 2030. For this purpose, we solve planning model (3) starting from a “greenfield” system, as in [10], [13], [14], [25], i.e., with no existing capacity for generation, transmission or storage. We do so to highlight the differences in the expansion plans resulting from the application of each time-period aggregation technique. Those discrepancies would be diluted if existing capacities were considered.

We consider two generic dispatchable generation technologies, namely, a base and a peak technology, whose economic and technical characteristics are listed in Table I. The base technology has a lower fuel cost but a tighter ramping limit. We do not take into account scheduled maintenance periods or unexpected failures of dispatchable power units. Consequently, we set the capacity factors of dispatchable generation technologies, i.e., ρ_{gnt} , to 1 in all simulations. Table I also includes the costs of investing in wind and solar power capacity, both with a fuel cost equal to 0. We use the hourly wind and solar capacity factors in each country in year 2010 available in [36].

TABLE I
GENERATION TECHNOLOGY DATA

Technology	i_g^G (€/MW)	y_g^G (years)	c_g (€/MWh)	r_g^+/r_g^- (p.u.)
Base	$4 \cdot 10^6$	60	10	0.1
Peak	$1.5 \cdot 10^6$	40	40	1.0
Wind	$1.5 \cdot 10^6$	25	-	-
Solar	$1 \cdot 10^6$	25	-	-

TABLE II
STORAGE TECHNOLOGY DATA

Storage	η_s (h)	ξ_s (p.u.)	i_s^S (€/MW)	y_s (years)
intraday	6	0.8	$1.5 \cdot 10^6$	80
interday	48	0.7	$2 \cdot 10^6$	60

TABLE III
MAXIMUM CAPACITY OF WIND, SOLAR AND HYDRO POWER GENERATION AND YEARLY AVAILABILITY FACTOR OF HYDRO POWER PRODUCTION

n	at	be	bg	cz	de	dk	ee
$\hat{p}_{wind,n}$ (GW)	45.8	17.7	68.5	56.1	222.6	32.6	15.8
$\hat{p}_{solar,n}$ (GW)	29.2	21.4	39.6	38.3	200.4	22.5	7.3
\hat{h}_n (GW)	11.1	0.1	2.2	1.2	3.9	0	0
a_n (p.u.)	0.44	0.37	0.2	0.26	0.6	0.3	0.49
n	es	fi	fr	gb	gr	hr	hu
$\hat{p}_{wind,n}$ (GW)	366.9	71.8	381.7	212.5	105.6	19.8	68.2
$\hat{p}_{solar,n}$ (GW)	221.6	20.3	251.8	179.3	62.8	13.4	44.3
\hat{h}_n (GW)	14.7	3.4	22.6	2.2	2.5	1.8	0.1
a_n (p.u.)	0.25	0.5	0.37	0.34	0.23	0.39	0.46
n	ie	it	lt	lu	lv	nl	pl
$\hat{p}_{wind,n}$ (GW)	56.3	190.2	37.6	1.7	28.4	23.6	193.9
$\hat{p}_{solar,n}$ (GW)	32.9	159.9	20.7	1.4	13.6	31.8	134.4
\hat{h}_n (GW)	0.2	15.3	0.1	0	1.5	0.1	1.2
a_n (p.u.)	0.37	0.35	0.43	0.39	0.23	0.32	0.29
n	pt	ro	se	si	sk	ch	no
$\hat{p}_{wind,n}$ (GW)	51	183	93.4	8.3	27.2	20.8	32.2
$\hat{p}_{solar,n}$ (GW)	35.1	111.2	30	5.4	18.3	18.7	12
\hat{h}_n (GW)	5.5	7.4	17.2	1.4	2	12.5	30.2
a_n (p.u.)	0.3	0.31	0.46	0.44	0.32	0.34	0.52

We also account for the possibility of installing intraday and interday storage technologies, whose main features are provided in Table II. The intraday storage technology has a lower investment cost and a higher round-trip efficiency, whereas the energy capacity of the interday storage is higher [15], [29], [37].

Each country is represented as a single bus and therefore, the European power system is modeled as a 28-bus network [37]. For each country, Table III provides the maximum installable capacity for wind, solar and hydro power generation, as well as the yearly availability of hydro power generation [37]. The maximum transmission capacities that could be reached in year 2030, which are shown in Table IV, are assumed twice the existing transmission capacities in year 2010. Table IV also includes the length of the transmission lines calculated as the distance between countries. The investment cost of transmission lines is set to €1000/MW·km with a lifetime of 50 years [37].

TABLE IV
MAXIMUM CAPACITY AND LENGTH OF LINES

nm	at-cz	at-de	at-hu	at-it	at-si	at-ch	be-fr	be-nl	bg-gr
\hat{f}_{nm} (GW)	1.4	3.2	0.9	0.3	1.8	1.5	4.2	4.5	0.9
l_{nm} (Km)	268	604	553	649	155	708	537	202	553
nm	bg-ro	cz-de	cz-pl	cz-sk	de-dk	de-fr	de-lu	de-nl	de-pl
\hat{f}_{nm} (GW)	0.8	2.9	2.7	3.2	3.7	5.8	1.9	7.9	2.0
l_{nm} (Km)	328	578	466	486	569	1067	504	584	971
nm	de-se	de-ch	dk-se	sk-no	ee-fi	ee-lv	es-fr	es-pt	fi-se
\hat{f}_{nm} (GW)	1.2	6.5	3.7	1.9	0.7	1.0	1.7	2.9	3.7
l_{nm} (Km)	1327	538	1100	477	345	179	923	512	812
nm	fr-gb	fr-it	fr-ch	gb-ie	gb-it	hr-hu	hr-si	hu-ro	hu-sk
\hat{f}_{nm} (GW)	4.0	3.3	4.1	0.5	1.0	1.5	1.5	1.1	1.7
l_{nm} (Km)	1197	1249	673	578	1072	527	115	621	159
nm	it-si	it-ch	lt-lv	nl-no	pl-se	pl-sk	se-no		
\hat{f}_{nm} (GW)	0.5	4.9	2.4	1.4	0.6	1.1	7.2		
l_{nm} (Km)	536	726	191	987	866	347	1133		

TABLE V
RESULTS FOR A RENEWABLE TARGET OF 50%

	F	C-672	D-28	W-4
Base (GW)	208	206	235	207
Peak (GW)	20	16	41	0
Wind (GW)	772	747	692	790
Solar (GW)	217	255	276	155
Hydro (GW)	160	160	160	160
Intraday (GW)	48	31	135	100
Interday (GW)	144	151	0	69
Network (GW)	23	18	34	19
Cost (10^9 €)	95.13	99.31	102.17	133.72
Share (%)	50	49.5	46.1	47.1
Shed (%)	0	0.1	0.2	1.1
Spil (%)	2	3.3	3.4	3.3

To work with sensible values for the electricity demand in 2030, we consider a yearly demand growth of 1%, starting from 2010. This means that the European electricity demand in 2030 is assumed to be 20% higher than that in 2010. The realized values of the European electricity demand in 2010 are publicly available in [38]. The cost of load shedding is assumed to be €1000/MWh.

For a renewable target of 50%, Table V shows the expansion plans provided by models F, C-672, D-28 and W-4 described in Section III. Note that $672 = 28 \times 24 = 4 \times 168$ and hence, the number of variables and constraints are the same for the latter three models. Results from model F are used as a benchmark. Capacity investments in generation, storage and transmission are provided in aggregated values for the whole system.

Each expansion plan is also evaluated in terms of total cost (investment plus operating costs), realized share of renewable energy in the electricity supply, percentage of load shedding and percentage of renewable power generation that is spilled. Each evaluation involves running model (3) considering the full set of hourly periods of year 2030 (i.e., without time-period aggregation) with the expansion decisions fixed to the values given by the time-period clustering technique under analysis. The target constraint (3b) is removed whenever investment decisions are fixed and all time periods are considered. The results of these evaluations are also shown in the last four rows of Table V. Notice that while other works compare the performance of different time aggregation methods in terms of how statistical properties

(such as correlation) of the original time series are preserved [13], [14], we compare them in terms of the optimality of the investment decisions obtained.

The clustering method that consists in selecting the four most representative weeks of the year (W-4) shows the worst performance by large, with a total cost and a percentage of load curtailment significantly higher than those prompted by the other two clustering algorithms. This happens because W-4 fails to capture those hours where the net load peaks, leading to expensive load curtailment if its investment decisions are evaluated considering all time periods. Actually, the peak power capacity under W-4 is significantly lower than for the other methods. The percentage of involuntary load curtailment is, in contrast, considerably smaller for the clustering technique that selects the most 28 representative days of the year (D-28). This technique, however, puts all the emphasis on the short-term dynamics of the power system, preventing the expansion planner from anticipating any potential benefit that could be reaped by interday energy shifting. As a result, expansion planning model D-28 leads to no investment at all in interday storage. Besides, D-28 underinvests in wind power generation capacity at the expense of overinvesting in solar power, which usually exhibits a more recurrent daily power profile than wind power generation. Consequently, the expansion plan delivered by D-28 fails to achieve the 50% renewables target by almost 4%. On the contrary, the proposed time-period clustering technique is, without any doubt, the one featuring a performance closest to the benchmark in terms of total cost, share of renewable energy in the electricity supply, percentage of load shedding and percentage of renewable power spillage. This is so because method C-672 manages to keep a balance in the representation of the intra and interday dynamics of the involved time-series data (wind, solar and demand) and also model more accurately the operation of energy storage devices. In fact, the installed storage capacities for C-672 are quite similar to those provided by the benchmark model.

In order to reduce load shedding costs of W-4 one could argue that those hours of the year with the highest electricity demand should always be accounted for in the expansion planning model. Notice, however, that load shedding can also occur when renewable generation is particularly low or the network is very congested at a local level. Since the capacity of renewable generating units and transmission lines are decision variables, identifying the hours that may involve load shedding actions a priori is not a trivial task.

Figure 2 represents the total cost associated with the expansion plans delivered by each of the considered clustering techniques as a function of the achieved share of renewable energy in the electricity supply. It should be noticed that the proposed clustering technique (C-672) outperforms the other two (D-28 and W-4) for almost all levels of share, with differences in performance that become more and more significant as the contribution of renewable power generation increases. The reason for this is that as the electricity supply relies more and more on wind and solar power generation, the accurate modeling of the storage operation becomes more relevant. By definition, the approach based on representative days is not able to model interday storage and thus, the cost yielded by this method is comparatively

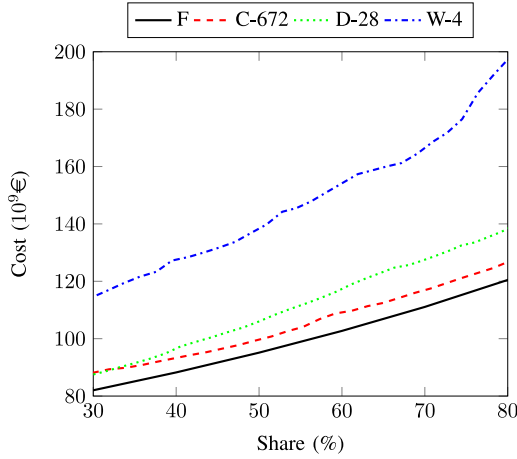


Fig. 2. Cost vs. renewable share.

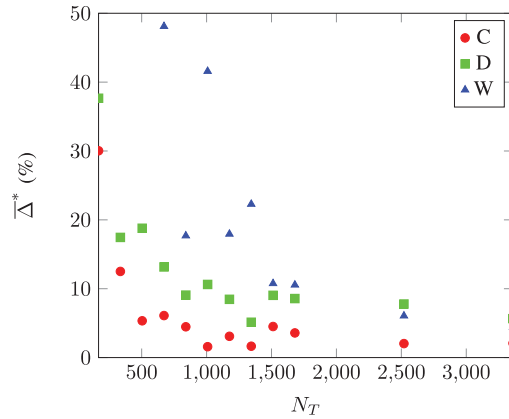


Fig. 3. Average cost error as a function of the number of periods.

higher. The cost increase of the clustering technique based on representative weeks is due to a different reason. Even if this approach is able to partially capture the benefits of interday storage, its underestimation of the net-load peaks remarkably worsens its performance.

In order to quantify and summarize the information provided by Fig. 2, we define the *average cost error* as follows

$$\bar{\Delta}^* = \frac{\overline{\text{Cost}}^* - \overline{\text{Cost}}^F}{\overline{\text{Cost}}^F} \times 100$$

where $\overline{\text{Cost}}^*$, with $*$ = {D, W, C} is the total cost of the expansion plan associated with clustering method “*” averaged over all levels of renewable share in between 30% and 80%. For instance, the average cost errors corresponding to C-672, D-28 and W-4 are 6.1%, 13.1% and 48.1%, respectively.

Next we investigate the performance of the methods compared in this paper for an increasing number of time periods, representative days and representative weeks. To do so, Fig. 3 plots the average cost error $\bar{\Delta}^*$ as a function of the number of periods N_T into which the whole year 2030 is aggregated. For the methods based on representative days and weeks, N_T is computed as $24 \times N_D$ and $168 \times N_W$, respectively.

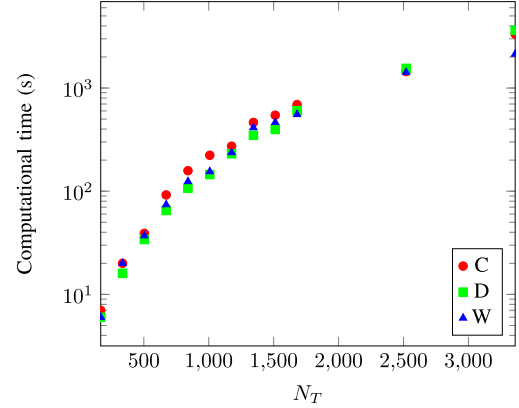


Fig. 4. Computational time.

TABLE VI
AVERAGE COST ERRORS UNDER DIFFERENT SCENARIOS

Scenario	C-672	D-28	W-4
No_solar	8.6%	18.3%	31.5%
No_wind	9.0%	7.2%	32.8%
No_hydro	2.3%	13.2%	60.3%
No_storage	11.1%	7.2%	6.3%

Observe that the average error incurred by the proposed chronological time-period clustering technique is remarkably lower than that of the other two clustering methods for a wide range of number of periods. For example, the number of periods required to keep the average cost error below 6% are, approximately, 3300, 2600 and 700 for methods D, W and C, respectively. Notice also that the clustering technique based on the selection of representative weeks is the one that benefits the most in relative terms from augmenting the number of periods, surpassing the performance of the clustering method based on representative days when the number of time periods is sufficiently increased.

Despite the substantial discrepancies in performance featured by the three clustering methods, all of them involve computational burdens of the same order of magnitude, as highlighted in Fig. 4. Note that the computational time shown in this figure accounts for both the time taken by the clustering method and the time required to solve the expansion planning model (3), although the former is very negligible compared to the latter. Furthermore, the average time needed to solve the expansion planning model (3) considering the full set of hourly periods is about ten hours, which highlights the ability of all the three clustering algorithms to speed up the solution process remarkably. The simulations results of this case study have been obtained using CPLEX 12.6.3 under Pyomo 5.2 on a Linux-based server with one CPU clocking at 2.6 GHz and 20 GB of RAM.

To conclude this section we analyze the performance of each time-period clustering technique under four scenarios. In each of these scenarios we exclude the possibility of installing a certain type of technology. This way, we denote those scenarios as *No_wind*, *No_solar*, *No_hydro*, and *No_storage*, with the short name indicating which technology has been ruled out.

Table VI includes the average cost error incurred by the three clustering algorithms under the four scenarios considered. It can be seen that the proposed method significantly outperforms the others in two out of those four scenarios, namely, *No_solar*, and *No_hydro*, while it displays a lower performance in scenarios *No_wind* and *No_storage*. To understand these results, we need to bear in mind that the proposed chronological time-period clustering method retains the mid-term dynamics of the input time series and allows for a more accurate modeling of the operation of interday storage. In fact, the advantage of the proposed approach stems precisely from the close interaction of these two aspects, that is, from its ability to capture the value of using interday storage for integrating renewable generation with relevant mid-term dynamics. On the other hand, the proposed approach loses information on the hour-to-hour variations of time-dependent parameters as illustrated in the bottom subplot of Fig 1.

The *No_wind* scenario is characterized by two time-dependent inputs (solar and demand) with strong daily patterns. In such a case, the value of interday storage is low, the accurate modeling of hour-to-hour variations becomes relatively more relevant and as a result, the proposed approach performs worse than the method based on the selection of representative days. In the *No_storage* scenario, installing storage is simply not a possibility and accordingly, the proposed approach results in the highest average cost error. Conversely, the value of interday storage is comparatively very significant in scenarios *No_solar* and *No_hydro*, where investments in solar and hydro power generation are excluded and, therefore, the proposed approach performs much better than the other methods in these two scenarios.

Finally, it is worth mentioning that the results provided in Table VI have been determined using the most commonly used approaches to aggregate time periods in order to compare them with the one proposed in this paper. Although there is a wide variety of methods to choose representative days and weeks, any of these methods will be unable to properly capture mid-term system dynamics and therefore, storage investments will be essentially inefficient and will involve, in turn, average cost errors similar to those included in Table VI.

V. CONCLUSION

Most existing time-period aggregation methods aim at reducing the computational burden of capacity expansion problems through the selection of a set of representative hours, days or weeks of the planning horizon. These methods do not properly capture the mid-term dynamics of renewable energy sources such as wind and fail to model the operation of electricity storage technologies accurately. To overcome these drawbacks, we propose in this paper a new time-period clustering technique that retains the chronology of the time-dependent parameters throughout the whole planning horizon. In this way, the proposed method determines capacity expansion plans that take into account the economic value of using interday storage to handle prolonged periods of high or low renewable power generation.

Using a case study based on the European electricity network, we compare the proposed method with existing ones in terms of their average cost error with respect to a benchmark capacity expansion model that works with the full set of time periods. Numerical results show the superior performance of our method, which determines more efficient capacity expansion plans than those prompted by existing methods without increasing the computational burden. For the base case scenario, the average cost error considering 672 time periods amounts to 13.1%, 48.1% and 6.1% for the methods based on representative days and representative weeks, and the proposed clustering technique, respectively. Using the European power network under different scenarios we also illustrate that the performance of the chronological time-period aggregation worsens for power systems that do not expect to include renewable power generation with mid-term dynamics (like wind) or energy storage devices. However, considering the global trend towards renewables-dominant power systems, with various forms of storage as enablers, the proposed clustering method proves to be an efficient and effective way to facilitate the solution of large-scale expansion planning models.

As shown in the bottom subplot of Fig. 1, the chronological time-period aggregation smooths out the input time series and therefore, filters out their shorter-term dynamics (e.g., peak values of a short duration). How to enhance or complement the proposed clustering technique to fully account for the short-term dynamics of the input time series in capacity expansion models is left as future research. Another topic that requires further investigation is how to adapt the time aggregation procedure here proposed to take into account the uncertainty of the time dependent parameters.

REFERENCES

- [1] R. Hemmati, R.-A. Hooshmand, and A. Khodabakhshian, "Comprehensive review of generation and transmission expansion planning," *IET Gener., Transmiss. Distrib.*, vol. 7, no. 9, pp. 955–964, Sep. 2013.
- [2] B. A. Frew and M. Z. Jacobson, "Temporal and spatial tradeoffs in power system modeling with assumptions about storage: An application of the POWER model," *Energy*, vol. 117, pp. 198–213, 2016.
- [3] K. Poncelet, E. Delarue, D. Six, J. Duerinck, and W. D'haeseleer, "Impact of the level of temporal and operational detail in energy-system planning models," *Appl. Energy*, vol. 162, pp. 631–643, 2016.
- [4] F. H. Murphy and Y. Smeers, "Generation capacity expansion in imperfectly competitive restructured electricity markets," *Oper. Res.*, vol. 53, no. 4, pp. 646–661, 2005.
- [5] J. H. Roh, M. Shahidehpour, and L. Wu, "Market-based generation and transmission planning with uncertainties," *IEEE Trans. Power Syst.*, vol. 24, no. 3, pp. 1587–1598, Aug. 2009.
- [6] J. H. Merrick, "On representation of temporal variability in electricity capacity planning models," *Energy Econ.*, vol. 59, pp. 261–274, 2016.
- [7] N. E. Koltsaklis and M. C. Georgiadis, "A multi-period, multi-regional generation expansion planning model incorporating unit commitment constraints," *Appl. Energy*, vol. 158, pp. 310–331, 2015.
- [8] F. D. Munoz and A. D. Mills, "Endogenous assessment of the capacity value of solar PV in generation investment planning studies," *IEEE Trans. Sustain. Energy*, vol. 6, no. 4, pp. 1574–1585, Oct. 2015.
- [9] Y. Dvorkin, R. Fernandez-Blanco, D. S. Kirschen, H. Pandžić, J. P. Watson, and C. A. Silva-Monroy, "Ensuring profitability of energy storage," *IEEE Trans. Power Syst.*, vol. 32, no. 1, pp. 611–623, Jan. 2017.
- [10] F. J. de Sisternes, J. D. Jenkins, and A. Botterud, "The value of energy storage in decarbonizing the electricity sector," *Appl. Energy*, vol. 175, pp. 368–379, 2016.

- [11] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer Series in Statistics), 2nd ed. New York, NY, USA: Springer, 2009.
- [12] M. S. ElNozahy, M. M. A. Salama, and R. Seethapathy, "A probabilistic load modelling approach using clustering algorithms," in *Proc. IEEE Power Energy Soc. General Meeting*, 2013, pp. 1–5.
- [13] K. Poncet, H. Hoshle, E. Delarue, A. Virag, and W. D'haeseleer, "Selecting representative days for capturing the implications of integrating intermittent renewables in generation expansion planning problems," *IEEE Trans. Power Syst.*, vol. 32, no. 3, pp. 1936–1948, May 2017.
- [14] Y. Liu, R. Sioshansi, and A. J. Conejo, "Hierarchical clustering to find representative operating periods for capacity-expansion modeling," *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 3029–3039, May 2018.
- [15] M. Beaudin, H. Zareipour, A. Schellenberg, and W. Rosehart, "Energy storage for mitigating the variability of renewable electricity sources: An updated review," *Energy Sustain. Develop.*, vol. 14, no. 4, pp. 302–314, 2010.
- [16] S. Wogrin and D. F. Gayme, "Optimizing storage siting, sizing, and technology portfolios in transmission-constrained networks," *IEEE Trans. Power Syst.*, vol. 30, no. 6, pp. 3304–3313, Nov. 2015.
- [17] H. Pandzic, Y. Wang, T. Qiu, Y. Dvorkin, and D. S. Kirschen, "Near-Optimal method for siting and sizing of distributed storage in a transmission network," *IEEE Trans. Power Syst.*, vol. 30, no. 5, pp. 2288–2300, Sep. 2015.
- [18] J. P. Barton and D. G. Infield, "Energy storage and its use with intermittent renewable energy," *IEEE Trans. Energy Convers.*, vol. 19, no. 2, pp. 441–448, Jun. 2004.
- [19] C. O'Dwyer and D. Flynn, "Using energy storage to manage high net load variability at sub-hourly time-scales," *IEEE Trans. Power Syst.*, vol. 30, no. 4, pp. 2139–2148, Jul. 2015.
- [20] P. Denholm and M. Hand, "Grid flexibility and storage required to achieve very high penetration of variable renewable electricity," *Energy Policy*, vol. 39, no. 3, pp. 1817–1830, 2011.
- [21] S. Spiecker and C. Weber, "The future of the European electricity system and the impact of fluctuating renewable energy—A scenario analysis," *Energy Policy*, vol. 65, pp. 185–197, 2014.
- [22] S. Weitemeyer, D. Kleinhans, T. Vogt, and C. Agert, "Integration of renewable energy sources in future power systems: The role of storage," *Renewable Energy*, vol. 75, pp. 14–20, 2015.
- [23] A. van Stiphout, S. Vaack, and G. Deconinck, "The role of long-term energy storage in investment planning of renewable power systems," in *Proc. IEEE Int. Energy Conf.*, Apr. 2016, pp. 1–6.
- [24] G. Fuchs, B. Lunz, M. Leuthold, and D. U. Sauer, "Technology Overview on Electricity Storage," Inst. Power Electron. Elect. Drives (ISEA), Aachen Univ., Germany, Tech. Rep. Jun., 2012. doi: [10.13140/RG.2.1.5191.5925](https://doi.org/10.13140/RG.2.1.5191.5925).
- [25] T. Brijis, A. van Stiphout, S. Siddiqui, and R. Belmans, "Evaluating the role of electricity storage by considering short-term operation in long-term planning," *Sustain. Energy, Grids Netw.*, vol. 10, pp. 104–117, 2017.
- [26] D. A. Tejada, S. Wogrin, and E. Centeno, "Representation of storage operations in network-constrained optimization models for medium- and long-term operation," *IEEE Trans. Power Syst.*, vol. 33, no. 1, pp. 386–396, Jan. 2018.
- [27] M. Nicolosi, A. Mills, and R. Wiser, "The importance of high temporal resolution in modeling renewable energy penetration scenarios," in *Proc. 9th Conf. Appl. Infrastructure Res.*, Tech. Rep. LBNL-4197E, 2010.
- [28] Q. Ploussard, L. Olmos, and A. Ramos, "An operational state aggregation technique for transmission expansion planning based on line benefits," *IEEE Trans. Power Syst.*, vol. 32, no. 4, pp. 2744–2755, Jul. 2017.
- [29] P. Nahmmacher, E. Schmid, L. Hirth, and B. Knopf, "Carpe diem: A novel approach to select representative days for long-term power system modeling," *Energy*, vol. 112, pp. 430–442, 2016.
- [30] J. H. Ward, "Hierarchical grouping to optimize an objective function," *J. Amer. Statist. Assoc.*, vol. 58, pp. 236–244, Mar. 1963.
- [31] G. Latorre, S. Member, J. J. M. Areiza, R. Cruz, and A. Villegas, "Classification of publications and models on transmission expansion planning," *IEEE Trans. Power Syst.*, vol. 18, no. 2, pp. 938–946, May 2003.
- [32] J. Wang, M. Shahidehpour, Z. Li, and A. Botterud, "Strategic generation capacity expansion planning with incomplete information," *IEEE Trans. Power Syst.*, vol. 24, no. 2, pp. 1002–1010, May 2009.
- [33] L. Baringo and A. J. Conejo, "Transmission and wind power investment," *IEEE Trans. Power Syst.*, vol. 27, no. 2, pp. 885–893, May 2012.
- [34] P. D. Brown, J. A. Peças Lopes, and M. A. Matos, "Optimization of pumped storage capacity in an isolated power system with large renewable penetration," *IEEE Trans. Power Syst.*, vol. 23, no. 2, pp. 523–531, May 2008.
- [35] P. Harsha and M. Dahleh, "Optimal management and sizing of energy storage under dynamic pricing for the efficient integration of renewable energy," *IEEE Trans. Power Syst.*, vol. 30, no. 3, pp. 1164–1181, May 2015.
- [36] I. Staffell and S. Pfenninger, "Using bias-corrected reanalysis to simulate current and future wind power output," *Energy*, vol. 114, pp. 1224–1239, 2016.
- [37] P. Nahmmacher, E. Schmid, and B. Knopf, "Documentation of Limes-EU—A long-term electricity system model for Europe," Potsdam Inst. Climate Impact Res. (PIK), 2014. [Online]. Available: <https://www.pik-potsdam.de/members/paulnah/limes-eu-documentation-2014.pdf>
- [38] ENTSO-e, 2016. [Online]. Available: <http://www.transparency.entsoe.eu>



Salvador Pineda (S'07–M'11) received the Ingeniero Industrial degree from the University of Málaga, Málaga, Spain, in 2006, and the Ph.D. degree in electrical engineering from the University of Castilla-La Mancha, Ciudad Real, Spain, in 2011. He is currently an Associate Professor with the University of Málaga. His research interests include power system operation and planning, electricity markets, renewable integration, energy policy, game theory, and optimization.



Juan M. Morales (S'07–M'11–SM'16) received the Ingeniero Industrial degree from the University of Málaga, Málaga, Spain, in 2006, and the Ph.D. degree in electrical engineering from the University of Castilla-La Mancha, Ciudad Real, Spain, in 2010.

He is currently an Associate Professor with the Department of Applied Mathematics, University of Málaga. His research interests include fields of power systems economics, operations and planning, energy analytics and optimization, smart grids, decision-making under uncertainty, and electricity markets.