Applied Energy 101 (2013) 475–482

Contents lists available at SciVerse ScienceDirect

Applied Energy

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

# Correlated wind-power production and electric load scenarios for investment decisions

L. Baringo, A.J. Conejo *

*Department of Electrical Engineering, Univ. Castilla-La Mancha, Campus Universitario s/n, 13071 Ciudad Real, Spain*

## HIGHLIGHTS

- ▶ Investment models require an accurate representation of the involved uncertainty.
- ▶ Demand and wind power production are correlated and uncertain parameters.
- ▶ Two methodologies are provided to represent uncertainty and correlation.
- ▶ An accurate uncertainty representation is crucial to get optimal results.

ARTICLE INFO

*Article history:*
Received 30 March 2012
Received in revised form 26 May 2012
Accepted 2 June 2012
Available online 19 July 2012

*Keywords:*
Clustering
Correlated scenarios
Electric load
Investment
Stochastic programming
Wind-power

ABSTRACT

Stochastic programming constitutes a useful tool to address investment problems. This technique represents uncertain input data using a set of scenarios, which should accurately describe the involved uncertainty. In this paper, we propose two alternative methodologies to efficiently generate electric load and wind-power production scenarios, which are used as input data for investment problems. The two proposed methodologies are based on the load- and wind-duration curves and on the K-means clustering technique, and allow representing the uncertainty of and the correlation between electric load and wind-power production. A case study pertaining to wind-power investment is used to show the interest of the proposed methodologies and to illustrate how the selection of scenarios has a significant impact on investment decisions.



## 1. Introduction

One key issue when dealing with investment problems is the modeling of the uncertain parameters that influence the investment decisions. These parameters include electric load, investment costs, fuel prices, market participants behavior, etc.

Two main techniques are available in the technical literature to deal with optimization problems involving uncertain data, namely stochastic programming [1,2] and robust optimization [3–5]. Stochastic programming, which is used in this paper, represents the uncertainty in the input data via scenarios [1] and thus, an adequate modeling of these scenarios is essential to achieve the best investment decisions. The selection of these scenarios and their influence on investment decisions are analyzed in this paper.

As an example, we consider a wind-power investment problem, which seeks to determine the wind-power capacity to be built in an existing electric energy system. Among the references addressing this problem [6–8], in this paper we consider the model proposed in [8] consisting in a stochastic mathematical program with equilibrium constraints (MPEC). In this particular model, there are two parameters subject to uncertainty that significantly influence the investment decisions: the electric load and the wind-power production.

Wind-power, as other renewable sources, is subject to uncertainty and thus requires models rather different to those developed for conventional generating units [9,10]. Moreover, wind-power production for the same facility is different in different locations depending on the wind-power conditions. These wind-power conditions can be represented via scenarios, which are generated using historical data in the location under study.

On the other hand, we face the uncertainty of the electric load of the system that has an important impact on market prices, which in turn modify the investment decisions. The electric load uncer-

* Corresponding author. Tel.: +34 926 295433; fax: +34 926 295361.
*E-mail addresses:* Luis.Baringo@uclm.es (L. Baringo), Antonio.Conejo@uclm.es (A.J. Conejo).

0306-2619/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.apenergy.2012.06.002

tainty can also be represented through scenarios based on historical data.

Finally, we should note that in most systems electric load and wind-power production are not statistically independent magnitudes. Low values of electric load usually occur during the night, when wind-power production is comparatively higher. Thus, considering electric load and wind-power production as independent phenomena may render suboptimal and inefficient investment decisions. It is thus necessary to properly represent the statistical correlation between electric load and wind-power production.

We propose two methods to address the uncertainty of and the correlation between the electric load and the wind-power production in different locations of an electric energy system: the load- and wind-duration curves technique [8] and the K-means clustering technique [11–13].

These methodologies use as input historical data of electric load and wind-power production in different locations of an electric energy system, and provide as output a reduced data set of electric load and wind-power production in different locations that keeps the information and correlation of the historical data. This reduced data set consists of a set of scenarios, each one comprising a value for the electric load and wind-power production in each location of the system. Note that each set of values of electric load and wind-power production in different locations represents a system operating condition, i.e., a scenario. For example, the K-means technique is used in [12] to generate load and wind input data for the probabilistic evaluation of the total transfer capability in power systems.

Within the context above, the contributions of this paper are threefold:

1. To propose, analyze and compare in detail two methodologies to precisely characterize the electric load and wind-power production in several locations of an electric energy system, considering the uncertainty of and the correlation between these two uncertain parameters.
2. To use these two methodologies to derive electric load and wind-power production scenarios used as input data for an investment decision problem (i.e., a long-term planning problem).
3. To analyze through a case study how different scenario representations of the electric load and the wind-power production influence wind-power investment decisions.

The remaining of this paper is organized as follows. In Section 2, we describe two techniques to characterize electric load and wind-power production uncertainty, as well as the correlation between the two parameters. In Section 3, these two techniques are used to characterize the electric load and the wind-power production in a given electric energy system. These data are then used to solve a wind-power investment problem in Section 4. Finally, Section 5 concludes the paper with some relevant remarks.

## 2. Techniques

In this section, we describe the two techniques used to represent the uncertain nature of the electric load and the wind-power production as well as their correlation: the load- and wind-duration curves and the K-means technique.

### 2.1. Load- and wind-duration curves

The first method to model the uncertainty and correlation of the electric load and the wind-power production at each bus of an existing electric energy system is based on the load- and wind-duration curves. Both electric load and wind-power production are jointly modeled as described below.

We consider available historical data of electric load consumption throughout one or several years in the electric energy system under study. These historical data are adequately scaled to account for demand growth. Using these scaled data, we build a load–duration curve as represented by the continuous curve in the lower plot of Fig. 1. This load–duration curve is approximated by a set of demand blocks, e.g., four blocks. Note that the first block is narrower (in terms of hour span) than the others in order to represent the peak load. Since the peak load can have a great impact on system-wide decisions, it should be adequately represented. The electric load uncertainty within each demand block is represented by considering different demand levels. In order to do this, we build the cumulative distribution function (cdf) of the electric load within each block, as depicted in Fig. 2. This cdf plot is divided into a selected number of segments (three in the example of Fig. 2), each one with an associated probability. The average values of the values within each segment give the electric demand levels represented in the lower plot of Fig. 1.

Once the electric load of the system is represented, we model the wind-power production. As explained in the introduction of this paper, wind-power production is correlated with the electric load of the system and thus, electric load and wind-power conditions have to be jointly represented.

We use historical data of wind-power capacity factors (defined as the wind-power production divided by the installed wind-power capacity) throughout the same period as the electric load. For each demand block used to adjust the load–duration curve, we consider the corresponding wind-power capacity factors (i.e., the wind-power capacity factor realizations corresponding to each

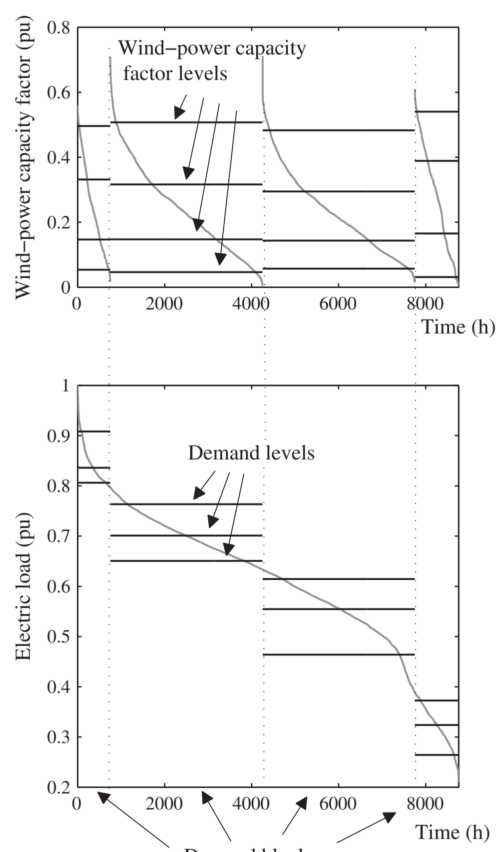


**Fig. 1.** Load- and wind-duration curves.

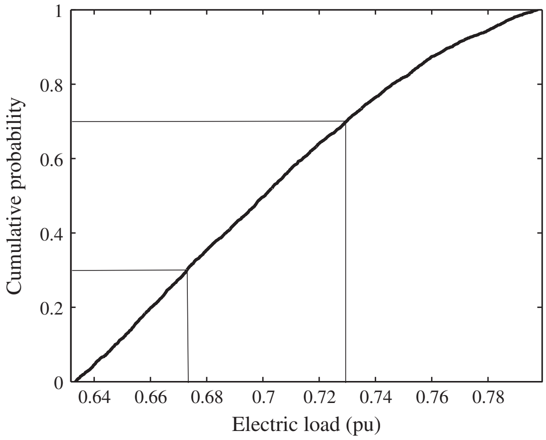


**Fig. 2.** Electric load cumulative distribution function.

load consumption within each demand block). With these data, we build, for each demand block, a wind-duration curve (these are the continuous curves represented in the upper plot of Fig. 1), which is approximated by a set of wind-power capacity factor levels (horizontal lines in the upper plot of Fig. 1). The procedure that we use to select the wind-power capacity factor levels is identical to that explained for the electric demand levels, i.e., we build the corresponding wind-power cdf as depicted in Fig. 3, we divide this cdf plot into a selected number of segments (four in the example shown in Fig. 3) and we compute the average value of each segment, which represent the wind-power capacity factor levels used in the study and represented in the upper plot of Fig. 1.

Within each demand block, we consider all the possible combinations of demand levels and wind-power capacity factor levels, which constitute the electric load and wind-power conditions for each block. Each condition is assigned a probability equal to the probability of the electric demand level times the probability of the wind-power capacity factor level.

For the example in Fig. 1, we have four demand blocks, three demand levels and four wind-power capacity factor levels, i.e., a total of 48 operating conditions. The weight of each of these operating conditions is computed as the number of hours in the corresponding demand block times the probability of each operating condition within the demand block.

The number of demand blocks used to adjust the load–duration curve, as well as the number of demand levels and wind-power capacity factor levels used to represent the electric load and wind-power production uncertainty should be selected taking into account the nature of the study to be carried out. A large number of blocks and levels may result in intractability while a small number may lead to a poor representation of the electric load and wind-power production of the system.

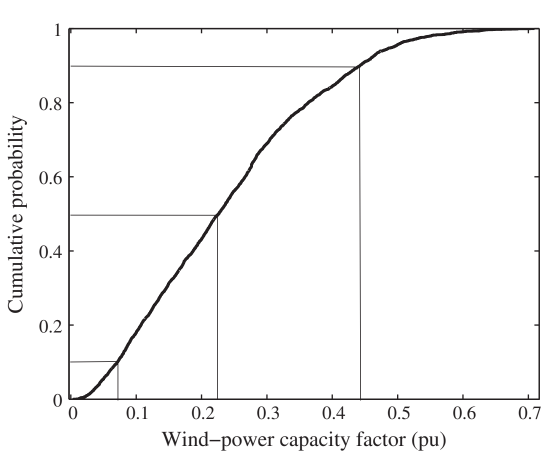


**Fig. 3.** Wind-power capacity factor cumulative distribution function.

The above technique is simple and easy to implement; however, it has the disadvantages that it can only be used if the same correlation among electric load and wind-power production is considered in all the locations of the system. The K-means technique described below solves this drawback.

### *2.2. Clustering technique: K-means method*

The aim of a generic clustering algorithm is to arrange data into groups according to similarities. The data are observations of physical processes (e.g., the electric load and wind-power production in a given location of an existing electric energy system).

Among different clustering algorithms, in this paper we use the K-means algorithm [11] due to its simplicity and good performance. Its working is described below.

As it is customary, we define a cluster as a group of observations that are similar among them and different from observations in other clusters. The observations are hourly historical data of electric load and wind-power production in different locations of the electric energy system under study. The objective of the clustering technique is to reduce these historical observations into a small enough set of clusters, each one defined by the value of the electric load and the wind-power production in the different locations, and the number of the original observations that are allocated to it.

As it is customary, we define the centroid of each cluster as the mean value of electric load and wind-power production in each location of all the historical observations allocated to the cluster.

Given these definitions, the K-means algorithm is based on the iterative algorithm below:

- *Step 1:* Select the number of required clusters according to the needs of the problem. Note that a low number of clusters can originate a poor representation of the electric load and wind-power production while a high number can lead to intractability.
- *Step 2:* Define the initial centroid of each cluster, e.g., randomly assigning a historical observation to each cluster.
- *Step 3:* Compute the distances between each original observation and each cluster centroid [11]. In this paper, we consider quadratic distances.
- *Step 4:* Allocate each historical observation to the closest cluster according to the distances calculated in Step 3.
- *Step 5:* Recalculate the cluster centroids using the historical observations allocated to each cluster.

Steps 3–5 are repeated iteratively until there are no changes in the cluster compositions between two consecutive iterations. The output of this algorithm are the cluster centroids as well as the number of observations allocated to each cluster. Note that each cluster centroid is defined by the values of the electric load and the wind-power production in different locations, which represent a system operating condition. On the other hand, the number of observations in each cluster provides the weight of each scenario.

The K-means technique requires specifying the number of clusters. As in the case of selecting the number of demand blocks to adjust the load–duration curve and the number of electric demand levels and wind-power capacity factor levels, the number of clusters should be large enough to adequately represent the electric load and wind-power production in the system.

It is important to note that the K-means technique allows representing different correlations between electric load and wind-power production across the considered electric energy system.

Other clustering techniques such as hierarchical clustering methods [14] are relevant to the problem addressed in this paper and may be used to derive the electric load and wind-power production scenarios. The main difference between these techniques and the K-means clustering technique is that the former ones do not require specifying in advance the number of clusters.

**Table 1**
Bus distribution in the IEEE RTS.

| Wind\demand | $t+2$ zone | $t+1$ zone | $t$ zone |
|---|---|---|---|
| North zone | 15, 16, 17, 18, 24 | 11, 12, 14, 19, 20, 21, 22 | 13, 23 |
| South zone | 1, 3, 4 | 2, 5, 9, 10 | 6, 7, 8 |

## 3. Input data analysis

In this section, we use the techniques described in Section 2 to reduce the historical observations of electric load and wind-power production into a set of scenarios.

For this purpose, we consider the IEEE Reliability Test System (RTS) [15] depicted in Fig. 4. This system comprises 24 buses, 38 lines, 32 generating units and 17 demands. Data defining this system is based on those provided in [15]. We assume that the IEEE RTS is divided in three demand zones: $t$, $t+1$ and $t+2$. The electric load in zones $t+1$ and $t+2$ is the same as in zone $t$ but with 1 and 2 h delay, respectively. On the other hand, we differentiate between two wind zones, north and south, which are considered to have the same wind levels but with 2 h delay, i.e., the wind-power capacity factors at buses in the south zone in hour $t$ are the same as the wind-power capacity factors at buses in the north zone in hour $t+2$. Additionally, we consider that wind-power conditions are better in the north than in the south assuming that wind-power capacity factors in the north are 10% higher than in the south. Table 1 identifies the IEEE RTS bus locations within the different demand and wind zones. Note that these assumptions are made for the sake of simplicity and due to the lack of historical disaggregated data. If historical data at different zones and buses were available, the proposed procedure is equally valid and results would be more realistic.

Fig. 5 depicts the aggregated historical data of normalized electric load and wind-power capacity factors throughout year 2007 in the Iberian Peninsula [17,18]. Note that electric load and wind-power capacity factors are correlated and an adequate modeling of this correlation is essential. These pairs of normalized electric load and wind-power capacity factor are considered to correspond with demand zone $t$ and the south wind zone. The electric load and wind-power capacity factors in the remaining zones are correlated as described in the previous paragraph.

The clustering mechanisms explained in Section 2 allow taking into account the electric load and wind-power production uncertainty and correlation. The output of these techniques is a reduced

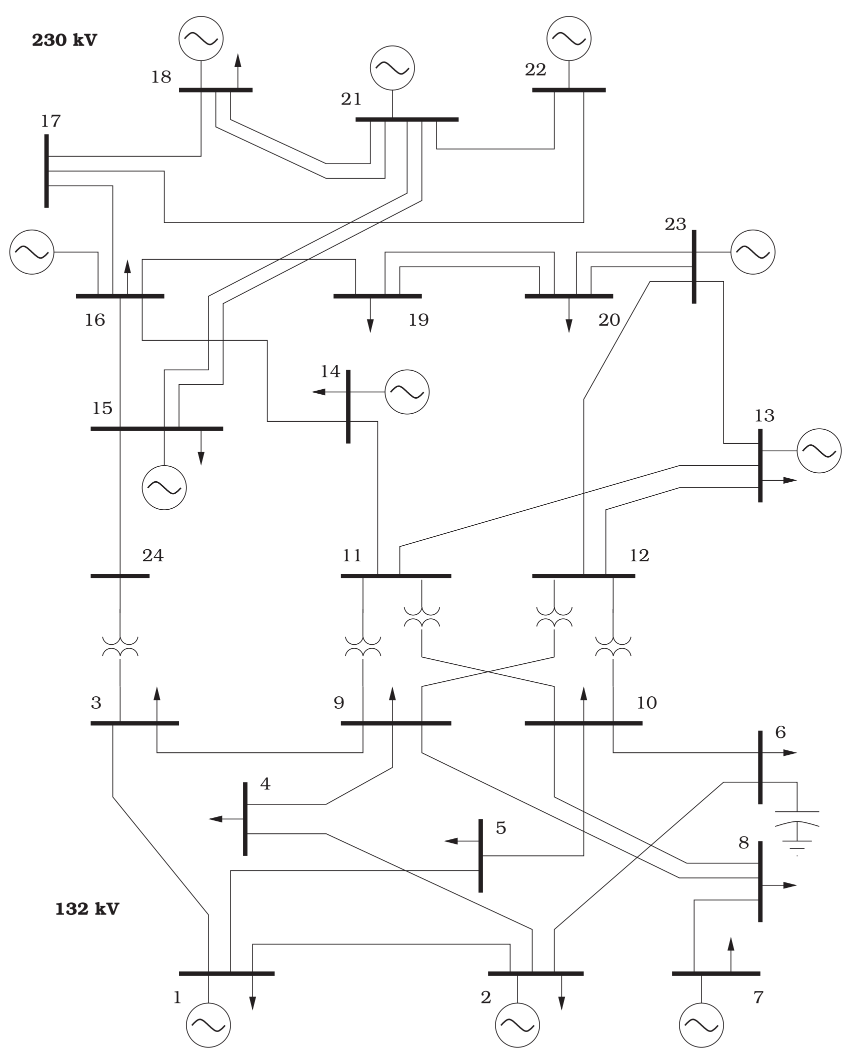


**Fig. 4.** IEEE 24-bus Reliability Test System.

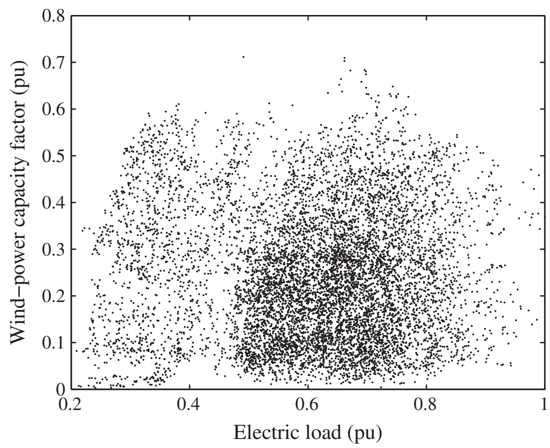


**Fig. 5.** Historical data of electric load and wind-power capacity factors.

**Table 2**
Load- and wind-duration curve data.

| Block # | Number of hours | Demand levels | Weight | Wind-power capacity factor levels | Weight |
|---|---|---|---|---|---|
| 1 | 750 | 0.9083 | 0.3 | 0.4961 | 0.1 |
| | | 0.8359 | 0.4 | 0.3312 | 0.4 |
| | | 0.8062 | 0.3 | 0.1465 | 0.4 |
| | | | | 0.0538 | 0.1 |
| 2 | 3500 | 0.7633 | 0.3 | 0.5071 | 0.1 |
| | | 0.7008 | 0.4 | 0.3161 | 0.4 |
| | | 0.6507 | 0.3 | 0.1470 | 0.4 |
| | | | | 0.0464 | 0.1 |
| 3 | 3500 | 0.6144 | 0.3 | 0.4828 | 0.1 |
| | | 0.5544 | 0.4 | 0.2948 | 0.4 |
| | | 0.4638 | 0.3 | 0.1433 | 0.4 |
| | | | | 0.0570 | 0.1 |
| 4 | 1010 | 0.3725 | 0.3 | 0.5402 | 0.1 |
| | | 0.3238 | 0.4 | 0.3887 | 0.4 |
| | | 0.2641 | 0.3 | 0.1653 | 0.4 |
| | | | | 0.0313 | 0.1 |

data set that maintains the information of the historical data set and also the correlation between electric loads and wind-power capacity factors at a given bus and among different buses.

In the next subsections, we use the two techniques described in Section 2 (load- and wind-duration curves and K-means) to analyze the data described above.

### 3.1. Load- and wind-duration curves

Using the hourly historical data provided in the previous subsection, we build the load-duration curve depicted in the lower plot of Fig. 1 which is adjusted by four demand blocks. Within each demand block, we consider three electric demand levels and four wind-power capacity factor levels, by building the electric demand and wind-power cdfs for each demand block and following the procedure described in Section 2.1. For example, Figs. 2 and 3 depict the cdfs of electric load and wind-power, respectively, within the second demand block. The electric load cdf is divided into three segments: the first segment comprising the load values with cumulative probabilities lower than 0.3; the second, load values with cumulative probabilities between 0.3 and 0.7; and the last one, load values with cumulative probabilities higher than 0.7. We compute the average values of the load values within each segment and use these values as the electric demand levels with probabilities of 0.3, 0.4 and 0.3 for segments 1, 2 and 3, respectively. We proceed similarly with the wind-power cdf depicted in Fig. 3. In this case we consider four segments with probabilities of 0.1, 0.4, 0.4 and 0.1.

Jointly considering demand blocks, electric demand levels and wind-power capacity factor levels, we obtain 48 different electric load and wind-power scenarios, each one representing a system operating condition. The weight of each condition is computed as the number of hours within each demand block times the probability of the particular electric load and wind-power condition within this block.

**Table 3**
Clustering technique results: electric load data.

| $\omega$ | $k^{D}_{t+2}$ | $k^{D}_{t+1}$ | $k^{D}_{t}$ | $\omega$ | $k^{D}_{t+2}$ | $k^{D}_{t+1}$ | $k^{D}_{t}$ |
|---|---|---|---|---|---|---|---|
| 1 | 0.7010 | 0.7194 | 0.7131 | 25 | 0.4727 | 0.4716 | 0.4749 |
| 2 | 0.3358 | 0.3367 | 0.3381 | 26 | 0.4018 | 0.3330 | 0.3698 |
| 3 | 0.6465 | 0.6496 | 0.6609 | 27 | 0.6409 | 0.6722 | 0.6562 |
| 4 | 0.4181 | 0.4193 | 0.4187 | 28 | 0.8444 | 0.8412 | 0.8485 |
| 5 | 0.6882 | 0.7115 | 0.7010 | 29 | 0.2991 | 0.2988 | 0.2974 |
| 6 | 0.6930 | 0.6594 | 0.6780 | 30 | 0.5932 | 0.5916 | 0.5875 |
| 7 | 0.6519 | 0.6478 | 0.6521 | 31 | 0.7386 | 0.7254 | 0.7335 |
| 8 | 0.3902 | 0.3953 | 0.3952 | 32 | 0.5237 | 0.5290 | 0.5217 |
| 9 | 0.7606 | 0.7619 | 0.7667 | 33 | 0.6915 | 0.6953 | 0.6903 |
| 10 | 0.2760 | 0.2756 | 0.2727 | 34 | 0.5890 | 0.5725 | 0.5762 |
| 11 | 0.5236 | 0.5200 | 0.5168 | 35 | 0.7822 | 0.7703 | 0.7809 |
| 12 | 0.7346 | 0.7474 | 0.7436 | 36 | 0.6230 | 0.5706 | 0.5932 |
| 13 | 0.5943 | 0.6590 | 0.6261 | 37 | 0.4642 | 0.4680 | 0.4649 |
| 14 | 0.3579 | 0.3563 | 0.3558 | 38 | 0.6932 | 0.6440 | 0.6709 |
| 15 | 0.6068 | 0.5956 | 0.5989 | 39 | 0.6491 | 0.6920 | 0.6933 |
| 16 | 0.5633 | 0.6296 | 0.5927 | 40 | 0.7377 | 0.7461 | 0.7450 |
| 17 | 0.5566 | 0.5548 | 0.5523 | 41 | 0.8428 | 0.8474 | 0.8495 |
| 18 | 0.2666 | 0.2665 | 0.2631 | 42 | 0.7136 | 0.6993 | 0.7058 |
| 19 | 0.3658 | 0.3583 | 0.3611 | 43 | 0.7606 | 0.7548 | 0.7621 |
| 20 | 0.6527 | 0.6141 | 0.6309 | 44 | 0.3609 | 0.3702 | 0.3673 |
| 21 | 0.2899 | 0.2925 | 0.2888 | 45 | 0.6674 | 0.6841 | 0.6786 |
| 22 | 0.3760 | 0.3738 | 0.3777 | 46 | 0.8487 | 0.8488 | 0.8541 |
| 23 | 0.5860 | 0.5908 | 0.5832 | 47 | 0.3206 | 0.3220 | 0.3194 |
| 24 | 0.5315 | 0.5292 | 0.5262 | 48 | 0.4488 | 0.4458 | 0.4490 |

**Table 4**
Clustering technique results: wind-power capacity factor data and number of hours within each final cluster.

| $\omega$ | $k^{W}_{N}$ | $k^{W}_{S}$ | $\vartheta$ | $\omega$ | $k^{W}_{N}$ | $k^{W}_{S}$ | $\vartheta$ |
|---|---|---|---|---|---|---|---|
| 1 | 0.2027 | 0.1803 | 297 | 25 | 0.5673 | 0.5154 | 84 |
| 2 | 0.1429 | 0.1283 | 108 | 26 | 0.3034 | 0.0878 | 4 |
| 3 | 0.5865 | 0.5348 | 62 | 27 | 0.5036 | 0.4593 | 116 |
| 4 | 0.2014 | 0.1927 | 101 | 28 | 0.4967 | 0.4503 | 176 |
| 5 | 0.2998 | 0.2630 | 232 | 29 | 0.4869 | 0.4465 | 89 |
| 6 | 0.1443 | 0.1280 | 308 | 30 | 0.5610 | 0.5182 | 79 |
| 7 | 0.2394 | 0.2100 | 378 | 31 | 0.5232 | 0.4707 | 143 |
| 8 | 0.3623 | 0.3198 | 56 | 32 | 0.0961 | 0.0945 | 529 |
| 9 | 0.4164 | 0.3726 | 217 | 33 | 0.7125 | 0.6469 | 21 |
| 10 | 0.0846 | 0.0765 | 153 | 34 | 0.1926 | 0.1773 | 353 |
| 11 | 0.2225 | 0.2122 | 400 | 35 | 0.1953 | 0.1718 | 256 |
| 12 | 0.6255 | 0.5662 | 51 | 36 | 0.3011 | 0.2826 | 255 |
| 13 | 0.1416 | 0.1332 | 234 | 37 | 0.3320 | 0.2934 | 89 |
| 14 | 0.5747 | 0.5203 | 156 | 38 | 0.3251 | 0.2973 | 254 |
| 15 | 0.0901 | 0.0833 | 368 | 39 | 0.3747 | 0.3409 | 161 |
| 16 | 0.2724 | 0.2523 | 230 | 40 | 0.0908 | 0.0798 | 353 |
| 17 | 0.4718 | 0.4257 | 87 | 41 | 0.1110 | 0.0925 | 245 |
| 18 | 0.3622 | 0.3367 | 62 | 42 | 0.4475 | 0.4121 | 129 |
| 19 | 0.4604 | 0.4173 | 106 | 43 | 0.3040 | 0.2686 | 251 |
| 20 | 0.4337 | 0.4003 | 137 | 44 | 0.2912 | 0.2673 | 92 |
| 21 | 0.2529 | 0.2288 | 91 | 45 | 0.0650 | 0.0579 | 322 |
| 22 | 0.0673 | 0.0623 | 122 | 46 | 0.2915 | 0.2574 | 216 |
| 23 | 0.3902 | 0.3574 | 181 | 47 | 0.3647 | 0.3379 | 94 |
| 24 | 0.3365 | 0.3140 | 229 | 48 | 0.4507 | 0.4107 | 93 |

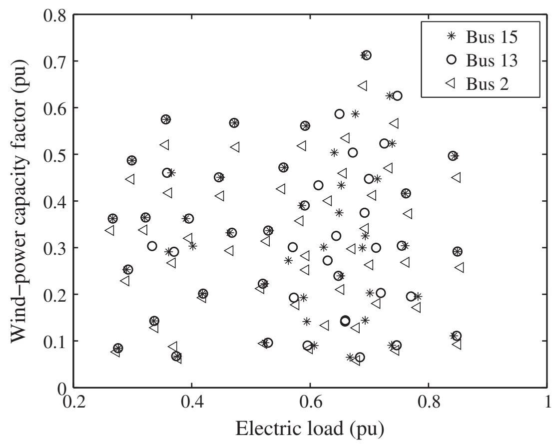


**Fig. 6.** Reduced data set of electric load and wind-power capacity factors.

Data defining the electric load and wind-power production scenarios are depicted in Fig. 1 and provided in Table 2.

### 3.2. Clustering mechanism: K-means method

The historical data set is reduced to a tractable data set using the clustering mechanism explained in Section 2.2. The K-means algorithm solution is obtained using SPSS [16]. We reduce the historical 8760 operating conditions (one for each hour of the year) into a reduced set of 48 scenarios, each one characterized by the electric load level in each demand zone, the wind-power capacity factor in each wind zone and the number of hours of the original data set that are included in each element of the reduced data set. Data defining the electric load conditions of this reduced data set are provided in Table 3, while data related to the wind-power capacity factors and the number of hours within each scenario are provided in Table 4. Note that $k_t^{\mathrm{D}}$, $k_{\mathrm{N/S}}^{\mathrm{W}}$ and $\vartheta$ are defined as the normalized electric load in demand zone $t$, the wind-power capacity factor in north/south wind zone and the number of hours within scenario $\omega$, respectively.

Fig. 6 depicts the reduced data set at three different buses located in three different zones of the system. As expected, operating conditions are similar but not equal at different buses.

## 4. Wind-power investment problem

In this section we solve the wind-power investment problem proposed in [8] and analyze how different modeling techniques to represent the electric load and wind-power production scenarios can result in different wind-power investment decisions.

### 4.1. Problem description

Ref. [8] considers a wind-power investor participating in a pool-based electricity market and interested in building new wind-power facilities to maximize its profit. The wind-power producer sells its wind-power production in the pool and is paid at the locational marginal price (LMP) of the bus at which it is located.

In order to maximize its profit, the wind-power investor has to decide both the location (among different buses) and the size of the new wind-power units to be built in an existing electric energy system. For this purpose, the wind-power investor faces the uncertainty of both the electric load and the wind-power production. For simplicity, the decision variables defining the wind-power capacity to be built at each bus of the system are considered continuous variables, which constitutes a difference with the assumption made in [8], where it is assumed that wind-power is available in discrete blocks.

Mathematically, the model proposed in [8] is an MPEC which can be easily recast as a mixed-integer linear programming (MILP) problem, which can be solved using commercially available branch-and-cut solvers. However, if a large system and a large number of electric load and wind-power scenarios are considered, the problem proposed in [8] becomes intractable. Nevertheless, this problem has a decomposable structure and becomes increasingly convex as the number of considered scenarios increases [19]. Thus, the Benders decomposition approach proposed in Ref. [20] can be applied. This is the method we use in this paper to obtain the optimal investment decisions.

All the results of this case study are obtained using CPLEX 11.2.1 [21] under GAMS [22] on a Linux-based server with four processors clocking at 2.9 GHz and 250 GB of RAM.

### 4.2. Data

We consider the IEEE RTS depicted in Fig. 4. Wind-power capacity can be built at buses 1, 2, 7, 13, 14 and 15 until a maximum of 800 MW per bus. The annualized wind-power investment cost is considered equal to $160,000 per MW and the investment budget is not limited.

We use the electric load and wind-power data provided in Section 3. Note that we consider three demand zones and two wind zones. While most of the load of the system is located in the south, the highest wind-power potential is in the north, where wind-power capacity factors are 10% higher than in the south. This constitutes a typical situation in realistic energy systems, in which wind-power facilities are generally located far away from load centers.

### 4.3. Results

In order to analyze how different modeling scenario-generation approaches influence wind-power investment decisions, we solve the investment problem stated in [8] considering the following cases:

1. The electric load and wind-power production scenarios are modeled using the load- and wind-duration curves described in Section 2.1. We use the data provided in Table 2.
2. The electric load and wind-power production scenarios are modeled using the clustering mechanism described in Section 2.2. In this case we consider that the electric load and wind-power capacity factors are identical at all the buses in the system, and equal to the values in demand zone $t+2$ and the north wind zone. We consider that wind-power capacity factors in the south zone are 10% lower than in the north as indicated in Section 3. Data defining wind-power and electric load scenarios are provided in Tables 3 and 4 (demand zone $t+2$ and north wind zone).
3. The electric load and wind-power production scenarios are modeled using the clustering mechanism described in Section 2.2. In this case, we consider different electric load and wind-power capacity factors in different demand and wind zones. Data for this case are provided in Tables 3 and 4.

Investment results for the three considered cases are provided in Table 5. Columns 2–7 provide the wind-power capacities to be built at buses 1, 2, 7, 13, 14 and 15, respectively. The eighth column provides the total wind-power capacity newly built in the system. Finally, the ninth column gives the expected profit.

**Table 5**
Investment results for the IEEE RTS.

| Case | New capacity (MW) built at bus | | | | | | Total new capacity (MW) | Profit (M$) |
|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 7 | 13 | 14 | 15 | | |
| 1 | 0 | 0 | 0 | 411.9 | 800 | 800 | 2011.9 | 16.340 |
| 2 | 0 | 0 | 0 | 333 | 800 | 800 | 1933 | 16.535 |
| 3 | 0 | 0 | 0 | 273.8 | 800 | 800 | 1873.8 | 16.430 |

**Table 6**
Investment results for the IEEE RTS considering limited transmission capacity.

| Case | New capacity (MW) built at bus | | | | | | Total new capacity (MW) | Profit (M$) |
|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 7 | 13 | 14 | 15 | | |
| 1 | 0 | 0 | 506.4 | 0 | 800 | 800 | 2106.4 | 18.846 |
| 2 | 45.9 | 0 | 346.9 | 0 | 800 | 592.7 | 1785.5 | 18.362 |
| 3 | 143.7 | 0 | 291.9 | 0 | 800 | 545.1 | 1780.7 | 18.547 |

As expected, results are different if different electric load and wind-power scenario characterizations are considered. Note however that this difference is not very significant and there are only differences in the wind-power capacity to be built at bus 13. The main differences are between cases 1 and 3, which use the load- and wind-duration curve and the clustering technique, respectively. Additionally, cases 1 and 3 consider the same (case 1) and different (case 3) electric load and wind-power capacity factors in different zones.

Note that despite being most of the load located in the south zone, the wind-power capacity is installed in the north zone, which has better wind-power conditions. This is so because there are no transmission congestion problems, as the lines connecting the south and north zones have capacity enough to allow the wind-power to flow south.

Next, we consider that the transmission capacity of lines connecting south and north zones is limited to 100 MW. We solve again cases 1–3 and obtain the results provided in Table 6.

In this case and due to transmission bottlenecks, it becomes optimal in all cases to install wind-power capacity at buses in the south zone despite having the south zone worse wind-power conditions than the north zone.

In the case of transmission bottleneck, which is common in real-world systems, the differences between investment decisions considering the same or different electric load and wind-power capacity factors in different zones (cases 1 and 2 vs. case 3) become important. Note that the total wind-power capacity to be built in cases 2 and 3 is very similar but not its distribution throughout the network. Due to the transmission capacity limits, prices are different at different buses, and thus it does matter where to build the wind-power facilities.

Finally, observe that the expected profit in the case of considering limited the transmission capacity between north and south zones is higher than in the original case. This can be explained by the fact that most of the loads are located in the south while most of the conventional generation facilities are in the north and thus, in case of transmission congestion between the south and the north, prices in the south become comparatively high and the wind-power investor takes advantages of this situation.

The wind-power investment problem analyzed in [8] assumes that transmission lines are fixed. However, as shown in this case study, if wind-power production is far away from the load centers and transmission congestion occurs, the whole wind-power potential is not used. A transmission and wind-power investment analysis could be carried out following the procedure described in [23]. The qualitative results of this analysis are expected to be similar to those obtained in this paper, i.e., modeling different electric load and wind-power conditions in different zones is important mainly if the system experiences transmission congestion.

## 5. Conclusions

This paper proposes and analyzes two methodologies to generate correlated electric load and wind-power production scenarios: the load- and wind-duration curves technique and the K-means clustering technique. The scenarios are used as input data in a decision making problem which aims to determine the wind-power facilities to be built in an existing electric energy system.

The conclusions below are in order:

1. Since the electric load and the wind-power production in a system are uncertain and correlated, an accurate modeling of this uncertainty and correlation is required in decision making (e.g., investment) problems.
2. The two proposed methodologies, the load- and wind-duration curve and the K-means clustering technique, allow efficiently representing the uncertainty and correlation in the electric load and the wind-power production. Additionally, the K-means technique allows representing the correlation between electric load and wind-power production in different locations, which entails comparatively higher accuracy.
3. Considering either identical or different electric load and wind-power capacity factors in different locations results in different investment decisions, which highlights the importance of an adequate scenario modeling.
4. The differences in item 3 become more relevant if transmission congestion occurs.

## Acknowledgments

L. Baringo and A.J. Conejo are partly supported by the Ministry of Economy and Competitiveness of Spain through CICYT Project DPI2009-09573.

## References

[1] Birge JR, Louveaux F. Introduction to stochastic programming. 2nd ed. Springer; 2011.
[2] Conejo AJ, Carrión M, Morales JM. Decision making under uncertainty in electricity markets. Springer; 2010.

[3] Ben-Tal A, Nemirovski A. Robust solutions of linear programming problems contaminated with uncertain data. Math Prog 2000;88:411–24.
[4] Bertsimas D, Sim M. The price of robustness. Oper Res 2004;52:35–53.
[5] Bertsimas D, Brown DB, Caramanis C. Theory and applications of robust optimization. SIAM Rev 2011;53:464–501.
[6] Burke DJ, O'Malley M. Maximizing firm wind connection to security constrained transmission networks. IEEE Trans Power Syst 2010;25:749–59.
[7] Zhou Y, Wang L, McCalley JD. Designing effective and efficient incentive policies for renewable energy in generation expansion planning. Appl Energy 2011;88:2201–9.
[8] Baringo L, Conejo AJ. Wind power investment within a market environment. Appl Energy 2011;88:3239–47.
[9] Murphy FH, Smeers Y. Generation capacity expansion in imperfectly competitive restructured electricity markets. Oper Res 2005;53:646–61.
[10] Kazempour J, Conejo AJ, Ruiz C. Strategic generation investment using a complementarity approach. IEEE Trans Power Syst 2011;26:940–8.
[11] Hoppner F, Klawonn F, Kruse R, Rumkler T. Fuzzy cluster analysis. Chichester: Wiley; 1999.
[12] Ramezani M, Singh C, Haghifam MR. Role of clustering in the probabilistic evaluation of TTC in power systems including wind power generation. IEEE Trans Power Syst 2009;24:849–58.
[13] Vallée F, Brunieau G, Pirlot M, Deblecker O, Lobry J. Optimal wind clustering methodology for adequacy evaluation in system generation studies using nonsequential Monte Carlo simulation. IEEE Trans Power Syst 2011;26:2173–83.
[14] Ward JH. Hierarchical grouping to optimize an objective function. J Amer Stat Assoc 1963;58:236–44.
[15] Reliability Test System Task Force. The IEEE reliability test system-1996. IEEE Trans Power Syst, vol. 14; 1999. p. 1010–20.
[16] SPSS. Statistical package for the social sciences. SPSS Inc; 2012. <http://www.spss.com>.
[17] Iberian Electricity Pool. OMEL; 2012. <http://www.omel.es/>.
[18] Red Eléctrica de España; 2011. <http://www.ree.es/>.
[19] Bertsekas DP, Sandell NR. Estimates of the duality gap for large-scale separable nonconvex optimization problems. In: IEEE conf on decision and control, Miami Beach, Florida;1982. p. 782–5.
[20] Baringo L, Conejo AJ. Wind power investment: a Benders decomposition approach. IEEE Trans Power Syst 2012;27:433–41.
[21] The ILOG CPLEX; 2012. <http://www.ilog.com/products/cplex/>.
[22] Rosenthal RE. GAMS, A user's guide. Washington, DC (USA): GAMS Development Corporation; 2011.
[23] Baringo L, Conejo AJ. Transmission and wind power investment. IEEE Trans Power Syst 2012;27:885–93.