

Article

# Wind Power Long-Term Scenario Generation Considering Spatial-Temporal Dependencies in Coupled Electricity Markets

Geovanny Marulanda <sup>1,\*</sup>, Antonio Bello <sup>1</sup>, Jenny Cifuentes <sup>2</sup> and Javier Reneses <sup>1</sup>

<sup>1</sup> Institute for Research in Technology (IIT), ICAI School of Engineering, Comillas Pontifical University, 28015 Madrid, Spain; antonio.bello@iit.comillas.edu (A.B.); javier.reneses@iit.comillas.edu (J.R.)

<sup>2</sup> Santander Big Data Institute, Universidad Carlos III de Madrid, 28903 Getafe, Spain; jcifuent@inst.uc3m.es

\* Correspondence: geovanny.marulanda@iit.comillas.edu

Received: 5 June 2020; Accepted: 30 June 2020; Published: 3 July 2020



**Abstract:** Wind power has been increasing its participation in electricity markets in many countries around the world. Due to its economical and environmental benefits, wind power generation is one of the most powerful technologies to deal with global warming and climate change. However, as wind power grows, uncertainty in power supply increases due to wind intermittence. In this context, accurate wind power scenarios are needed to guide decision-making in power systems. In this paper, a novel methodology to generate realistic wind power scenarios for the long term is proposed. Unlike most of the literature that tackles this problem, this paper is focused on the generation of realistic wind power production scenarios in the long term. Moreover, spatial-temporal dependencies in multi-area markets have been considered. The results show that capturing the dependencies at the monthly level could improve the quality of scenarios at different time scales. In addition, an evaluation at different time scales is needed to select the best approach in terms of the distribution functions of the generated scenarios. To evaluate the proposed methodology, several tests have been made using real data of wind power generation for Spain, Portugal and France.

**Keywords:** ARIMA; long-term forecasting; multi-area electricity markets; SARIMA; wind power forecasting

## 1. Introduction

Wind energy is one of the fastest-growing renewable energy technologies around the world [1]. Power from wind produces no air pollution or carbon dioxide emissions, no water consumption, and it is an inexhaustible renewable energy source [2]. In addition, operational costs of wind generators are close to zero once turbines have been built [3]. Global efforts to mitigate climate change and environmental restrictions have motivated an increase of wind power participation in electricity markets in many countries and it will continue growing. In fact, annual capacity additions for onshore wind power are expected to increase to more than 200 GW around the world until 2050, that is, more than four times the wind power added to the operation in 2018 [4].

In this context, generating realistic scenarios for wind power production in the medium and long term is fundamental to guide decision-making in power systems. Accurate scenarios are needed to design support schemes for wind energy promotion and to define wind power integration targets. Moreover, these scenarios are needed to attract new investments, to achieve a proper level of deployment according to wind resource availability, and to assess possible impacts on the energy mix [5]. Likewise, these scenarios are used by firms to mitigate the risk exposure, to design energy portfolios, to estimate electricity prices and to build realistic expansion planning scenarios in power systems [6].

In general, the high variability of renewable energies makes it difficult to predict their energy output. This supposes big challenges to firms and system operators. In fact, the intermittence of renewable resources could produce power curtailments and reduce system reliability of power systems. Moreover, an increase of the operational cost could be expected if more reserves are required to cover sudden energy decreases [2]. Consequently, additional support mechanisms are needed to deal with renewable energies uncertainty [7]. In this way, many countries have been making efforts to increase cross-border interconnections and to improve the market integration with other areas [8]. These actions not only boost the security of electricity supply, but also increase market competition, reduce the need to build new power plants and allow to integrate more renewable technologies into electricity markets.

The current trend towards market integration is changing the paradigm in which power systems are operated. As the interconnection among different areas increases, wind power generation will not only be dependent on the resource availability in local markets, but will also be affected by wind power behavior in other interconnected areas both in the short and the long term. However, in most instances, scenario generation methodologies have been proposed at local scales and just a few works include spatial-temporal dependencies among areas [9]. Particularly, the works presented in [10,11] propose methodologies to generate multiple scenarios considering spatial-temporal correlations among time series of wind turbines at different locations. Nevertheless, these approaches analyze time series from wind turbines and a whole market analysis is not performed.

Two kinds of strategies have been widely reported in the literature to generate wind power scenarios: direct and indirect [12]. In direct approaches, wind power generation is predicted from historical power data of turbines. In contrast, indirect approaches predict the wind speed in a first step, and then wind power is forecasted through the curves that relates wind speed and wind power for each turbine [12]. Classic forecasting techniques include physical and statistical approaches. While physical approaches are based on fundamental physical principles like conservation of mass and energy, statistical approaches generally use previous historical data to build statistical models [1]. Most of these approaches are proposed for a period of less than one year [13,14], and just a few are focused on scenario generation with longer time horizons [3]. Particular examples of these kinds of work have been reported in [15] and [16].

Recently, forecasting techniques based on Artificial Intelligence (AI) and hybrid models have been tested [3,17–21]. Usually, AI-based and hybrid models have shown less prediction errors than classic techniques [22,23]. However, classic techniques are easier to interpret and are less intensive from a computational point of view [18,24]. These characteristics are valuable, especially when the objective is to generate many scenarios that simulate the hourly behavior of wind generation in the long term. In fact, an hourly resolution in long-term scenarios could help to better deal with the problems originated by wind energy intermittence.

Most of the literature proposes methodologies to generate scenarios for wind power in a turbine or a group of turbines located in one area, and a whole system analysis is not performed. Actually, there is a lack of scenario generation methodologies capable of capturing the spatial-temporal dependencies of wind power in multi-area electricity markets. This paper fills this gap in the literature. The proposed approach is focused on generating realistic scenarios for wind power as local methodologies fail to reproduce properly the dependency relationships of wind power behavior among areas and markets. In fact, aggregated wind power time series from several systems not only include implicit information of wind behavior in all interconnected markets, but also network status and firm strategies. In addition, generating wind power scenarios from a market-aggregated perspective decreases the error when compared with desegregated approaches [25]. In particular, this approach with an hourly resolution allows system operators to simulate the operation of power systems and to assess wind energy production in a more effective way [16]. Finally, an hourly resolution allows to evaluate the quality of scenarios generated at different temporal scales, which has received little attention in the literature.

Based on the previous considerations, a new methodology to generate realistic scenarios for wind power in the long term in interconnected electricity markets is proposed in this paper. This approach considers spatial-temporal dependencies among different electricity markets. Furthermore, the effectiveness of this proposal has been assessed by backtesting approaches at distinct temporal scales. Finally, several tests have been made using real time series of wind power generation for Spain, Portugal and France. The contrast between the size of each market and its wind installed capacity allows to capture a wide range of scenarios from a local and global perspective.

This proposal is divided into five main steps. In the first two steps, a data pre-processing and a time series decomposition is performed for each market. Four data normalization transformations are assessed, and trends and seasonality are captured for each area. In a third step, the temporal dependencies for each area are identified for wind power through Seasonal Auto Regressive Integrated Moving Average models (SARIMA). In this step, the dynamic of wind power for each area is captured and dependencies among areas are treated in detail. Moreover, in order to identify the best model, two approaches for SARIMA have been compared. In the fourth step, the estimation of probability distribution functions considering dependencies, found at the third stage, is conducted through correlated and multivariate Monte Carlo simulations. With this, the idiosyncrasy of wind generation dynamics for each time series and the dependencies among areas are replicated in each market. Finally, backtesting and benchmarking strategies are proposed to assess the effectiveness of the two proposed approaches at different time scales.

Summarizing, the main contributions of this paper are:

1. A new methodology to obtain realistic scenarios for wind power generation considering spatial-temporal dependencies among electricity markets is proposed. There is a lack of methodologies to generate realistic scenarios of wind power from a whole market perspective. The proposed methodology contributes to filling this gap.
2. Two approaches to generate realistic wind power scenarios are proposed for the long term with a statistical approach. Most forecasting statistical approaches tackle the wind power forecasting problem from a short-term perspective. In this paper, a long-term approach with an hourly resolution is proposed.
3. Finally, different strategies to evaluate the accuracy of generated scenarios are proposed and assessed at different time scales. Statistical properties of distribution functions of generated scenarios have been compared with historical data on an hourly, daily, weekly and monthly scale.

This paper is organized as follows. In Section 2, the methodology is detailed. In Section 3, the results obtained with the proposed methodology are analyzed and compared. Finally, in Section 4, the main conclusions are presented and discussed.

## 2. Methodology

As mentioned, five steps have been proposed in this paper to generate scenarios of wind power production. The first step includes data pre-processing, i.e., the treatment of data mistakes, removing the dependencies between wind power and installed capacity for each area, and the transformation to a Gaussian-normalized time series. Second, a time series decomposition is performed. In this step, trends, seasonality and residuals for each area have been identified. Third, SARIMA models have been fitted to the historical data to detect temporal dependencies. Fourth, in order to generate wind power scenarios, several paths of multivariate and correlated Monte Carlo simulation of SARIMA models has been simulated. At last, backtesting and benchmarking stages are carried out from a bottom-up approach in order to compare Monte Carlo simulations with the original data. A description of each one of these steps is included in the next sections.

### 2.1. Data Pre-Processing

Missing data from time series for each area were avoided using linear interpolation methods and all outliers were considered for the study cases. Moreover, possible mistakes in historical data have been identified and amended. Contrary to short-term approaches, installed capacity of wind power needs to be removed from the time series of each area. Thus, the trend of the time series will only depend on the dynamics of wind power generation in each market. In addition, a better representation of spatial-temporal correlations among areas is obtained. In this paper, dependencies of generated wind power due to the installed capacity are removed using per unit values, i.e., time series are built using the utilization of wind power instead of real power. In this way, the generated scenarios could be easily scaled by multiplying by new capacity expansions.

As mentioned in [2], wind power shows more variability than wind speed. In fact, the time series of wind power reflects the joint effects of the high variability of wind speed, the controllability in wind turbines and the network status. Moreover, wind power is proportional to the cube of wind speed in the operational region of turbines. Still, some events could decrease the power output of wind turbines to zero suddenly, like maintenance activities and low wind speed conditions. In addition, aggregated time series reflects the dynamics of multiple geographical areas with different weather characteristics.

In this context, wind power data shows a non-normal distribution. This asymmetrical and nonlinear behavior may distort time series decomposition, add spurious interactions in data, and complicate the model analysis. Several transformations have been proposed to normalize time series [26]. In this paper, four of the most popular transformations have been compared: Box-Cox, Yeo-Johnson, Logit, and Ordered Quantile [27]. To the best of our knowledge, these transformations have not been compared for data normalization of wind power generation in electricity markets. However, applications of these transformations can be found in [27–30]. Table 1 shows each one of the transformations considered in this paper:  $y$  refers to the original data, both  $y^{(\lambda)}$  and  $z$  are transformed data,  $\lambda$  is a parameter adjusted in both cases via maximum likelihood,  $\Phi$  represents the standard normal of the cumulative distribution function, and rank and length are the observation’s rank and the number of samples of data series, respectively.

In order to select the best transformation, the Pearson P statistic (divided by its degrees of freedom) is calculated for each approach. Pearson P statistic allows to compare among transformations as an absolute distance of the departure of normality and it is a relatively interpretable goodness-of-fit measure. In this case, the closer the data is to a normal distribution, the closer the Pearson P value will be 1.

**Table 1.** Data normalization techniques.

| Name             | Transformation   |
|------------------|--|
| Box-Cox          | $y^{(\lambda)} = \begin{cases} \frac{y^\lambda - 1}{\lambda}, & \text{if } \lambda \neq 0 \\ \log(y), & \text{if } \lambda = 0 \end{cases}$  |
| Yeo-Johnson      | $y^{(\lambda)} = \begin{cases} \frac{(y+1)^\lambda - 1}{\lambda}, & \text{if } y \geq 0, \lambda \neq 0 \\ \log(y + 1), & \text{if } y \geq 0, \lambda = 0 \\ -\frac{((1-y)^{2-\lambda} - 1)}{2-\lambda}, & \text{if } y < 0, \lambda \neq 2 \\ -\log(1 - y), & \text{if } y < 0, \lambda = 2 \end{cases}$ |
| Logit            | $z = \ln\left(\frac{y}{1-y}\right)$  |
| Ordered quantile | $z = \Phi^{(-1)}\left(\frac{\text{rank}(y) - \frac{1}{2}}{\text{length}(y)}\right)$  |

### 2.2. Time Series Decomposition

Time series decomposition refers to deconstructing a time series into different components; typically, trend, seasonality and residuals. In particular, additive decomposition is appropriated when the magnitude of seasonal variations or the fluctuation around the trend-cycle do not vary with

the level of the series [31]. In this step, the deterministic components of time series are identified. Additive decomposition models are represented by (1):

$$y_t = T_t + S_t + R_t \quad (1)$$

where  $y_t$ ,  $T_t$ ,  $S_t$  and  $R_t$  denotes the normalized time series, the trend, seasonality, and residuals of time series at period  $t$ , respectively. A linear trend, if any, is fitted by using a least-square approach. Annual and/or daily seasonality could be included in time series decomposition. Thus, the seasonality is given by (2):

$$S_t = S_{a_t} + S_{d_t} \quad (2)$$

where  $S_{a_t}$  and  $S_{d_t}$  are the annual and daily seasonality at time  $t$ , respectively. The annual seasonality may be represented through Fourier series [32–34], since this fits better than, for example, quadratic models [35]. It can be calculated as (3):

$$S_{a_t} = \sum_{j=1}^{\lfloor \frac{s}{2} \rfloor} \left\{ a_j \cos \left( \frac{2\pi jt}{s} \right) + b_j \sin \left( \frac{2\pi jt}{s} \right) \right\} \quad (3)$$

where  $s$  is the number of hours of the year, and  $a_j$  and  $b_j$  are constants optimized from historical data for each area using a least-squares approach. In addition, daily seasonality is calculated for each hour  $i$  of the day as shown in (4):

$$S_{d_i} = \mu_i - \bar{\mu} \quad (4)$$

where  $\mu_i$  is the mean for each of the twenty four hours of the day in the series and  $\bar{\mu}$  is the mean of  $\mu_i$ , that is (5)–(6):

$$\mu_i = \frac{1}{n_i} \sum_{\substack{j=1 \\ j \in \Omega_i}}^N y_j, \forall i = 1, 2, \dots, 24 \quad (5)$$

$$\bar{\mu} = \frac{1}{24} \sum_{i=1}^{24} \mu_i \quad (6)$$

being  $n_i$  the number of data for hour  $i$ ,  $\Omega_i$  the set of the existing data for hour  $i$ , and  $N$  the number of data for hour  $i$  in all the data series of the same area.

### 2.3. Detection of Spatial-Temporal Dependencies

Wind is produced by differences in the atmospheric pressure. In fact, air moves from higher to lower pressure areas developing different wind speeds. These differences in atmospheric pressures are mainly due to unequal solar heating around the world. Because of this, the dynamics of wind power depends on the season of the year and the hour of the day. In order to capture time dependencies of wind power, a SARIMA has been fitted for each area. In general, a SARIMA( $p, d, q$ ) ( $P, D, Q$ ) $_S$  process is represented as [36]:

$$\theta_p(B) \Theta_P(B^S) (1 - B)^d (1 - B^S)^D y_t = w_q(B) W_Q(B^S) a_t \quad (7)$$

where  $S$  is the periodicity, and  $p$  and  $q$  are integers that indicates the non-seasonal autoregressive and moving average polynomial orders, respectively.  $B$  is the backward shift operator, satisfying  $By_t = y_{t-1}$  and  $B^k y_t = y_{t-k}$ , and  $a_t$  is white noise with mean equal to zero and variance  $\sigma^2$ . In addition,  $P$  and  $Q$  are integers representing the seasonal autoregressive and moving average polynomial orders, and  $d$  and  $D$  are non-seasonal and seasonal differencing orders, respectively.  $w_q(B)$  and  $\theta_p(B)$  are polynomials in  $B$  of degrees  $p$  and  $q$ , respectively, and are given by:

$$\theta_p(B) = 1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3 - \dots - \theta_p B^p \quad (8)$$

$$w_q(B) = 1 - w_1B - w_2B^2 - w_3B^3 - \dots - w_qB^q \quad (9)$$

Finally,  $\Theta_p(B)$  and  $W_q(B)$  are polynomials in  $B$  of degrees  $P$  and  $Q$ , respectively, as is shown in (10) and (11):

$$\Theta_p(B^S) = 1 - \Theta_1B^S - \Theta_2B^{2S} - \Theta_3B^{3S} - \dots - \Theta_pB^{pS} \quad (10)$$

$$W_q(B^S) = 1 - W_1B^S - W_2B^{2S} - W_3B^{3S} - \dots - W_qB^{qS} \quad (11)$$

To fit the best SARIMA model, the Box-Jenkins method has been followed [37]. This method consists of an iterative process that can be summarized in three main steps [36]: model identification, parameter estimation, and diagnostic checking. In the model identification, suitable values for SARIMA orders are established. With this aim, an analysis of the Autocorrelation Function (ACF) and Partial Autocorrelation Functions (PACF) is performed. In the parameter estimation, the polynomial coefficients of the SARIMA are calculated. These parameters are estimated via maximum likelihood for all the suitable values found in the model identification. At last, the fitted models are checked for inadequacies. In addition to the visual inspection of residual autocorrelations for each series, the Bayesian Information Criterion (BIC) have been used to select the best final structure of the SARIMA models that allows us to capture the conditional mean. In this way, BIC is given by [38]:

$$\text{BIC} = -2(L(x_1, x_2, \dots, x_n; \hat{\theta})) + m \ln(n) \quad (12)$$

where  $L(x_1, x_2, \dots, x_n; \hat{\theta})$  is the maximized value of the likelihood function for the estimated model,  $\hat{\theta}$  are the values that maximize the likelihood function,  $m$  is the number of parameters to be estimated, and  $n$  is the number of observed data. In this context, for a set of possible SARIMA models, the preferred is the one with the minimum BIC value.

Capturing the wind power behavior throughout the year and the spatial-temporal dependencies at a multidimensional level, the question arises whether it may be interesting to use a dynamic model that is specifically calibrated for different periods of time throughout the year. Because of this, two kinds of SARIMA models have been fitted to be compared. First, one static SARIMA is fitted for each area. With this approach, the dependencies of all historical data have been considered in each SARIMA. That means that the scenarios will be generated using this only model for the whole time horizon. However, as data dependencies could change according to the time of the year, a dynamic methodology to fit one SARIMA per month is accomplished in the second approach. In this case, one time series for each month and area is built with historical data of the same month and area. In other words, twelve SARIMA have been fitted for each area using monthly time series. In contrast to the first approach, the SARIMA fitted for each area considers only information of the historical data for the same month. This way, the scenarios will be generated month to month through the twelve fitted SARIMA.

#### 2.4. Scenario Generation

In traditional wind power forecasting approaches, each scenario is generated from uncorrelated and univariate normal distributions. However, time series in multi-area electricity markets could be strongly correlated, not only due to weather condition in near areas, but to the status of the grid, the load profiles, and firms' strategies, among others. Moreover, these dependency relationships may change throughout the year when a modeling at market or country level is performed. To capture these correlations, hourly multivariate and correlated residuals with normal distribution functions have been generated. That is, the generated residuals include spatial correlations through covariance matrices obtained from monthly historical data series:

$$f_x(x_1, \dots, x_k) = \frac{\exp\left(-\frac{1}{2}(\mathbf{x} - \mu)^T \Sigma^{-1}(\mathbf{x} - \mu)\right)}{\sqrt{(2\pi)^k \Sigma^{1/2}}} \quad (13)$$

where  $k$  is the number of residuals generated,  $\mathbf{x}$  is a real  $k$  dimensional vector,  $\Sigma$  is the symmetric covariance matrix,  $|\Sigma|$  is the determinant of  $\Sigma$  and  $\mu$  is the mean of the residuals.

Once temporal dependencies have been modeled with SARIMA, and spatial relationships among areas have been included in residuals, Monte Carlo simulations are performed to generate the wind power scenarios. Contrary to deterministic approaches, Monte Carlo simulations allow to generate complete distribution functions for each hour. Moreover, each generated scenario includes temporal dependencies identified with the SARIMA model. In the case of the dynamic approach, presample responses and innovations for initial values of Monte Carlo simulation have been included from each path belonging to the previous month. Thus, scenarios are coherent among months and more realistic paths are achieved. Furthermore, the source of uncertainty that is related to the random error term is incorporated in the process of generating scenarios. Therefore, the variation in the errors has been considered. In the static case, the last values of historical data have been considered. Finally, in both approaches, trend and seasonality are added to simulated residuals, data series are inversely transformed, and a performance evaluation is carried out.

### 2.5. Performance Evaluation

As mentioned before, there is a lack of methodologies to assess the quality of wind power scenarios generated at different time scales, such as daily, weekly or monthly. Usually, a global benchmarking between distribution functions of generated scenarios and historical data is performed for an hourly scale. However, this approach does not give information about the performance of wind power scenario generator to model the behavior of historical data at different temporal scales. This information is essential when realistic wind paths are needed on a larger scale, for example, in generation planning, maintenance scheduling, and security analysis.

Hourly resolution for long-term scenarios as proposed in this paper allows to assess bottom-up analysis. In order to evaluate the generated scenarios, a comparison between the main statistical properties of the historical time series and the proposed scenarios is carried out at different time scales with a bottom-up approach. Moreover, a comparison between different SARIMA approaches was accomplished. In order to assess the quality of the scenarios, two stages have been used. In the first stage, distribution functions and the two first moments of the distribution have been contrasted to assess the global performance for each approach. In a second stage, the Pinball Loss Function (PLF) and the Winkler Score (WS) have been used to evaluate each approach at different intervals.

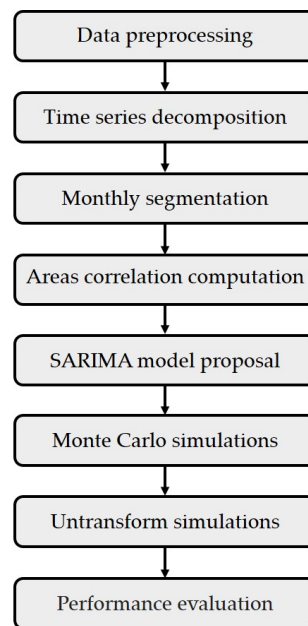
PLF and WS have been widely used to assess a forecast through a probabilistic approach [39]. In particular, PLF is an error measure for a quantile forecast [6,39]. Unlike other measures, a score is assigned according to the differences between the real values of a time series and the quantile values obtained in a forecast. The lower the score, the better the forecast. Let  $\hat{y}_a$  be the percentile of scenarios generated with  $a = 1, 2, \dots, 99$ . Then, PLF is given by [40,41]:

$$PLF(\hat{y}_a, y_t) = \begin{cases} (1 - \frac{a}{100})(\hat{y}_a - y_t) & \text{if } y_t < \hat{y}_a \\ (\frac{a}{100})(y_t - \hat{y}_a) & \text{if } y_t \geq \hat{y}_a \end{cases} \quad (14)$$

Unlike PLF, WS enables to compare between prediction intervals that takes into account both “sharpness” and “calibration” [42]. The WS penalizes when the forecasted value lies outside a predefined interval, and rewards a narrow prediction. As with PLF, the lower the achieved scores are, the better is the forecast. If the  $(100 - \alpha)\%$  prediction interval for time  $t$  is given by  $[L_t, U_t]$ , then the WS is calculated as [6,39]:

$$W_t = \begin{cases} (U_t - L_t) & \text{if } L_t \leq y_t \leq U_t \\ (U_t - L_t) + \frac{2}{\alpha}(L_t - y_t) & \text{if } y_t < L_t \\ (U_t - L_t) + \frac{2}{\alpha}(y_t - U_t) & \text{if } y_t > U_t \end{cases} \quad (15)$$

It should be noted that these parameters depend on the number of Monte Carlo simulations. However, as the number of simulations increase, the difference tends to be lower. To contrast the obtained results, a global benchmarking is first performed. Secondly, a monthly evaluation is conducted, and, finally, a weekly and an hourly comparison for each month are assessed. Furthermore, the monthly covariances of the generated scenarios have been assessed to ensure that spatial dependencies are accurately captured. Finally, Figure 1 shows a flow chart that summarizes the methodology proposed in this paper.



**Figure 1.** Flow chart for the methodology.

### 3. Results

The proposed methodology has been tested in a real study case using Matlab<sup>®</sup>, in particular, the 2018b version. For this purpose, real hourly wind power data for Spain from 2008 to 2019, and from 2012 to 2019 for Portugal and France have been analyzed. In this way, 105,192 samples for Spain and 70,128 for Portugal and France have been considered. This study case reflects the different dynamics in terms of wind resources in different regions. Data are available with the authors.

There are remarkable particularities that make this study case interesting to generate wind power scenarios with a multi-area approach. First, while electricity markets of Spain and Portugal are joined in the Iberian Electricity Market (MIBEL), energy exchange between MIBEL and France is limited by the market clearance of each area. In this way, energy exchanges not only depend on wind resource availability in the three areas, but also on the balance between offer and supply in the MIBEL and French markets. Second, the energy flow among areas is limited by the interconnection capacity. To date, these areas are joined through a cross-border interconnection, up to about 3000 MW between Spain and France and close to 4000 MW between Spain and Portugal [43]. Both market clearance and cross-boarder capacity could be inducing spatial-temporal correlations in wind power generation, as opposed to others produced by climate. These correlations are expected to be captured with the proposed approaches. Finally, the electricity mix is different in each country. Currently, France, Spain and Portugal have an installed generation capacity close to 130, 105, and 20 GW, respectively. From these values, 9%, 22% and 25% are from wind power [44]. The contrast between the size of each market and its wind installed capacity allows to capture a wide range of scenarios from a local and global perspective. This section shows the main findings for each step proposed in the Methodology section.



### 3.1. Data Pre-Processing

Figure 2 shows the real historical data of wind power per unit for each area. As it was expected, temporal series show an annual seasonal behavior for the three countries. Moreover, the differences among wind power utilization are evident, with Portugal being the country with the highest utilization.

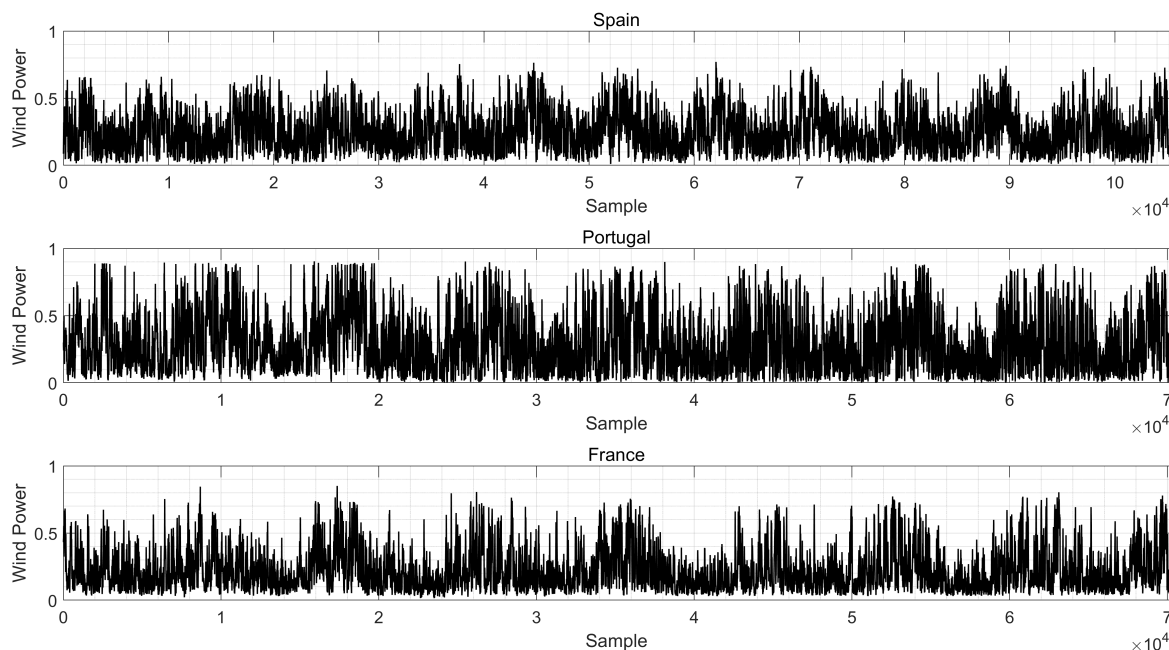


Figure 2. Historical data series.

To normalize data series Box-Cox, Yeo-Johnson, Logit and Ordered quantile transformation have been evaluated. As an example, Figure 3, shows at the top quantile-quantile (QQ) plots for original data, and below, the normalized time series obtained with each transformation for France. A visual inspection on these figures indicates that while tails remain in Box-Cox, Yeo-Johnson and Logit transformation, asymmetries have been removed with Ordered quantile transformation, showing the best normalized data. Similar plots were obtained for the other areas. To complement the visual inspection of these figures, the Pearson P statistic has been calculated for each transformation. The obtained results are shown in Table 2 for each area. It is clear that the Ordered quantile transformation shows the lowest value for the Pearson P statistic for all the areas. Based on these results, the Ordered quantile transform has been selected for data normalization in all areas.

Figure 4 shows the real data histograms and kernel density approaches for each area, as well as the same histograms and kernel densities for ordered quantile transformed series. Histograms of real data show asymmetries. In particular, long tails to the right for each area and heavy tails to the left are identified. In contrast, transformed data are much closer to a normal distribution and these asymmetries have been removed in all areas by means of ordered quantile transformation.

Table 2. Pearson P statistics for each area and transformation.

| Area     | No Transformation | Yeo-Johnson | Logit | Box-Cox | Ordered Quantile |
|----------|-------------------|-------------|-------|---------|------------------|
| Spain    | 24.77             | 8.75        | 9.67  | 3.87    | 1.14             |
| Portugal | 35.28             | 15.19       | 14.92 | 5.54    | 1.11             |
| France   | 52.91             | 13.87       | 4.97  | 5.13    | 1.08             |

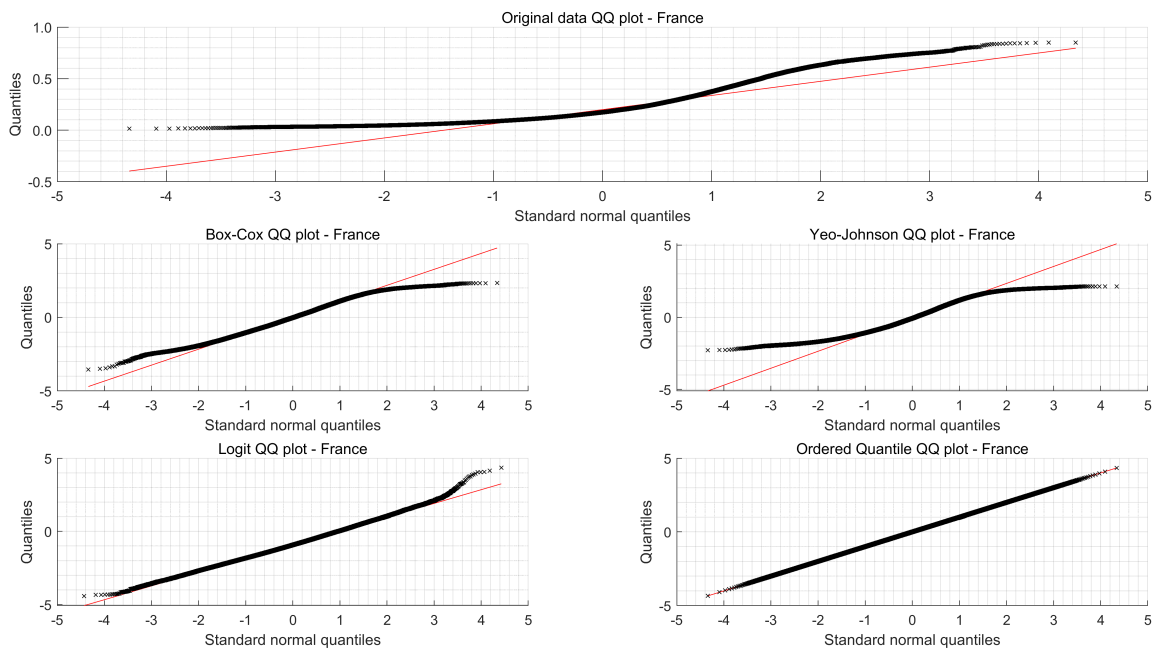


Figure 3. Quantile quantile plots for France.

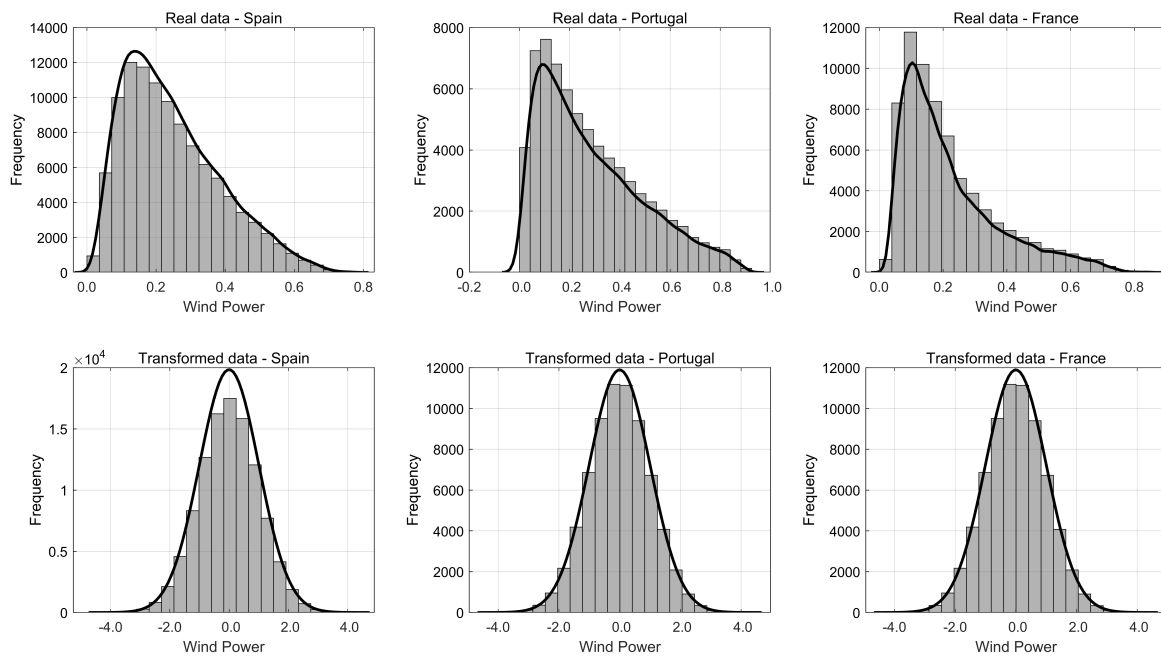


Figure 4. Historical and normalized histograms.

### 3.2. Time Series Decomposition

After the transformation, seasonality and trends from data series of each area have been removed. As an example, Figure 5 shows the data decomposition for Spain. Similar results have been obtained for Portugal and France. The horizontal and vertical axis of each graph represent the hours and the wind power, respectively. The first graph at the top shows time series transformed with ordered quantile. The second figure shows annual seasonality obtained from Equation (3). Usually, a few terms of Fourier transformation are needed to estimate seasonality. In this paper, they have been obtained from a harmonic function with order two. In this case, the daily component has not been included. The third graph identifies a trend with a low positive slope for Spain. This could be explained by

the installation of wind turbines with a higher performance in recent years. Finally, at the bottom, the white noise obtained after seasonality and trend subtraction is shown according to Equation (1). In this figure, a highly random behavior is identified after the decomposition when compared with the data series shown at the top of the figure. Figure 6 shows the obtained residuals for each area after the data decomposition.

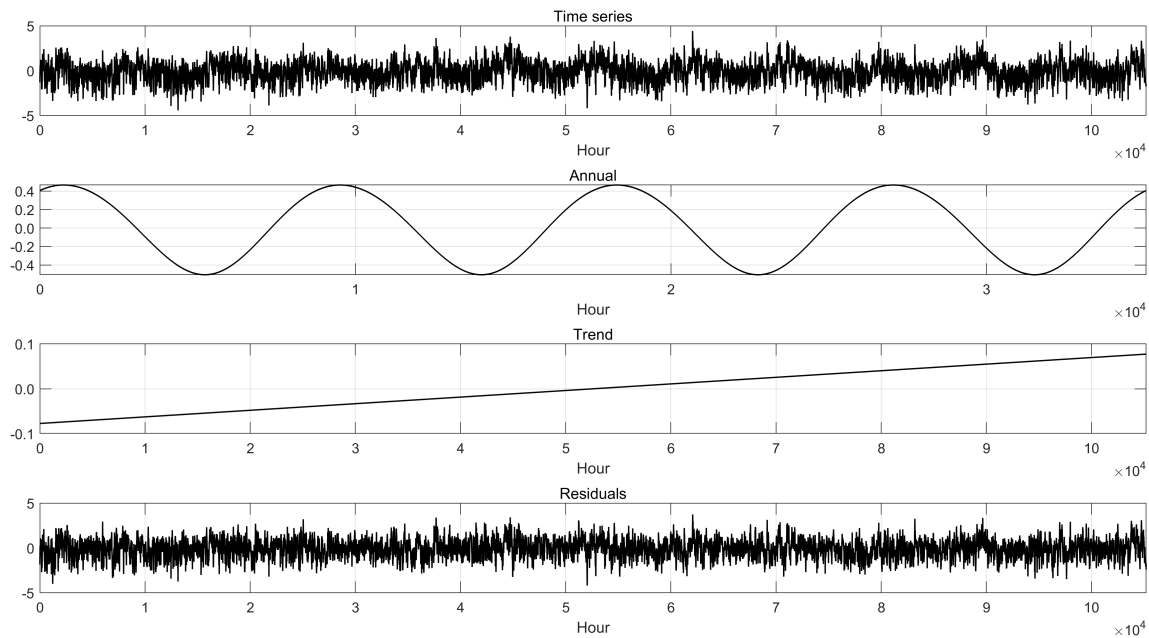


Figure 5. Data series decomposition of Spain.

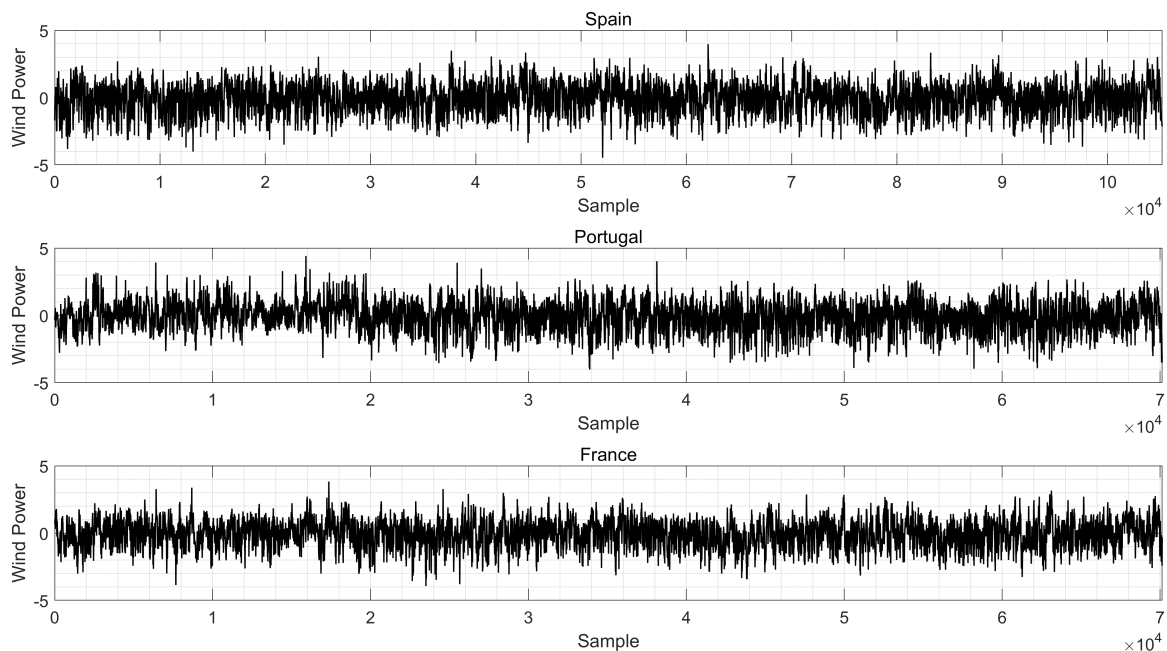


Figure 6. Remaining component obtained after data decomposition for each area.

### 3.3. Detection of Spatial-Temporal Dependencies

Weather and geographical aspects are internalized in aggregated form through the SARIMA models and the correlations among areas. As previously mentioned, two approaches have been proposed to evaluate spatial-temporal dependencies of wind power for each area at different time

scales. While in the first approach, dynamics of wind power are captured along the whole time series, in the second approach a dynamic segmentation is proposed to consider only dependencies among historical data within the same month. To find the order of the best SARIMA model, the Box-Jenkins methodology has been used [37]. In a first stage, a visual inspection of ACF and PACF has been carried out to find possible SARIMA orders. In a second stage, an exhaustive combination of these values and different autoregressive and moving average orders for seasonal and non-seasonal components of the SARIMA model was carried out. The differentiation has not been taken into account. Subsequently, BIC values are evaluated to select the SARIMA orders that produce the best fit for the data. Moreover, the residuals have been evaluated to ensure a white noise behavior. Finally, to calculate the coefficients of the polynomials  $\theta$ ,  $\Theta$ ,  $w$ , and  $W$  for selected SARIMA, a maximum likelihood estimation has been implemented.

### 3.4. Scenario Generation

Random and correlated paths have been simulated for static and dynamic approaches. With this aim, multivariate residuals with mean zero and considering the covariances among areas have been generated from the same normal distribution through Monte Carlo simulations. As a result, 1000 paths for each month  $m$  have been generated with a number of innovations equal to the number of hours of the historical data. The last values of each path in  $m - 1$  have been used as starting points for pre-sample responses and innovations of each path of month  $m$ , respectively. Once these scenarios are generated, the trend and seasonality are added to the simulated residuals and, finally, the data series are inversely transformed.

### 3.5. Performance Evaluation

The performance of the static and dynamic approach have been evaluated with a bottom-up approach. In the first stage, the results have been compared on an hourly basis. In order to quantify the differences between both approaches, PLF and WS have been evaluated. The results obtained for each area are shown in Tables 3 and 4, respectively. The best results are highlighted in bold. Both tables suggest a superior performance of the dynamic approach in all the percentiles and intervals.

**Table 3.** Pinball loss function comparison on an hourly scale.

| Area     | Approach | P5     | P30    | P50    | P70    | P95    | Mean          |
|----------|----------|--------|--------|--------|--------|--------|---------------|
| Spain    | Static   | 0.0100 | 0.0444 | 0.0570 | 0.0546 | 0.0198 | 0.0372        |
|          | Dynamic  | 0.0098 | 0.0419 | 0.0519 | 0.0475 | 0.0143 | <b>0.0331</b> |
| Portugal | Static   | 0.0132 | 0.0615 | 0.0793 | 0.0747 | 0.0231 | 0.0504        |
|          | Dynamic  | 0.0131 | 0.0608 | 0.0776 | 0.0724 | 0.0217 | <b>0.0491</b> |
| France   | Static   | 0.0087 | 0.0408 | 0.0550 | 0.0549 | 0.0188 | 0.0356        |
|          | Dynamic  | 0.0087 | 0.0402 | 0.0538 | 0.0527 | 0.0172 | <b>0.0345</b> |

**Table 4.** Winkler score comparison on an hourly scale.

| Area     | Approach | 50%  | 80%  | 90%  | 98%  | Mean        |
|----------|----------|------|------|------|------|-------------|
| Spain    | Static   | 0.36 | 0.50 | 0.60 | 0.90 | 0.59        |
|          | Dynamic  | 0.32 | 0.43 | 0.48 | 0.58 | <b>0.45</b> |
| Portugal | Static   | 0.49 | 0.64 | 0.73 | 0.89 | 0.69        |
|          | Dynamic  | 0.48 | 0.62 | 0.69 | 0.80 | <b>0.65</b> |
| France   | Static   | 0.35 | 0.47 | 0.55 | 0.73 | 0.52        |
|          | Dynamic  | 0.34 | 0.45 | 0.52 | 0.64 | <b>0.49</b> |

In the second stage, an evaluation of hourly data for each month has been performed. Figure 7 shows the mean of the hourly historical data for each month, as well as the values obtained for the dynamic approach in all the areas. The vertical axis shows the value for each month indicated on the horizontal axis. It can be seen from this figure that the dynamic approach is close to historical data the whole horizon. The obtained results suggest that the temporal relations are well captured with the dynamic approach from a monthly point of view. In other words, the dynamic approach not only captures appropriately the distribution function of all historical data (as shown in Tables 3 and 4), but also captures better the mean of each month. As an example, Figure 8 shows the monthly density distribution functions obtained with both approaches for Spain. A comparison with historical distribution functions indicates that the dynamic approach reproduces more accurately the behavior of the time series than the static approach.

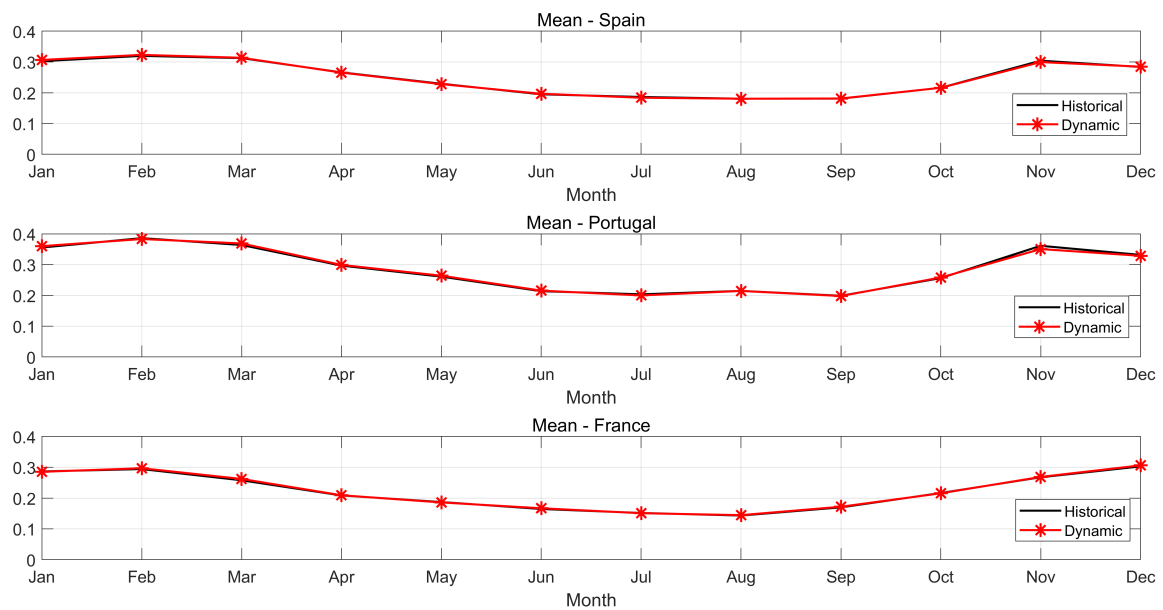


Figure 7. Monthly comparison for mean.

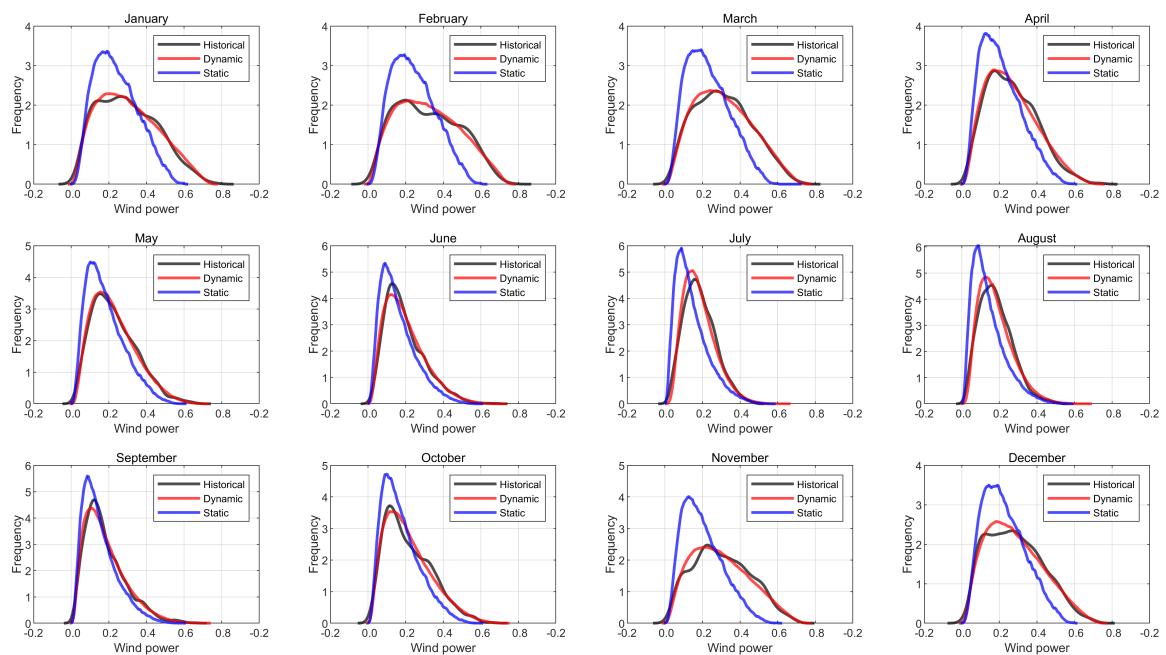


Figure 8. Monthly comparison of density distribution functions.

Likewise the results obtained for PLF at percentiles P5, P30, P50, P70 and P95, and the mean are provided in Table 5. The best results are highlighted in bold. In all cases, a higher score is obtained with the static approach. That means that the dynamic approach represents more accurately the distribution function of wind power than the static approach at different quantiles, when monthly distribution functions are considered. In addition, Table 6 shows the WS with intervals of 50%, 80%, 90% and 98%. Again, as it can be seen in bold, the dynamic approach shows a superior performance to reproduce the monthly historical behavior of the time series for each area at different intervals.

**Table 5.** Pinball loss function comparison on a monthly scale.

| Area     | Approach | P5     | P30    | P50    | P70    | P95    | Mean          |
|----------|----------|--------|--------|--------|--------|--------|---------------|
| Spain    | Static   | 0.0052 | 0.0230 | 0.0323 | 0.0370 | 0.0276 | 0.0250        |
|          | Dynamic  | 0.0044 | 0.0145 | 0.0160 | 0.0139 | 0.0042 | <b>0.0106</b> |
| Portugal | Static   | 0.0056 | 0.0198 | 0.0250 | 0.0244 | 0.0119 | 0.0173        |
|          | Dynamic  | 0.0053 | 0.0189 | 0.0227 | 0.0196 | 0.0056 | <b>0.0144</b> |
| France   | Static   | 0.0043 | 0.0158 | 0.0204 | 0.0209 | 0.0118 | 0.0146        |
|          | Dynamic  | 0.0040 | 0.0140 | 0.0169 | 0.0157 | 0.0055 | <b>0.0112</b> |

**Table 6.** Winkler score comparison on a monthly scale.

| Area     | Approach | 50%  | 80%  | 90%  | 98%  | Mean        |
|----------|----------|------|------|------|------|-------------|
| Spain    | Static   | 0.23 | 0.42 | 0.66 | 1.92 | 0.81        |
|          | Dynamic  | 0.10 | 0.15 | 0.17 | 0.26 | <b>0.17</b> |
| Portugal | Static   | 0.16 | 0.26 | 0.35 | 0.72 | 0.37        |
|          | Dynamic  | 0.14 | 0.19 | 0.22 | 0.25 | <b>0.20</b> |
| France   | Static   | 0.14 | 0.23 | 0.32 | 0.79 | 0.37        |
|          | Dynamic  | 0.11 | 0.15 | 0.19 | 0.26 | <b>0.18</b> |

In the third stage, the scenarios generated with both approaches have been evaluated on a weekly basis. With this aim, two-year scenarios have been generated and contrasted with the two last years of historical data. The mean for each week has been calculated both for the historical data and for each path of the dynamic and static approaches. Tables 7 and 8 show the results obtained for PLF and WS, respectively. The values of PLF and WS are always lower for the dynamic approach, as it is shown in bold values.

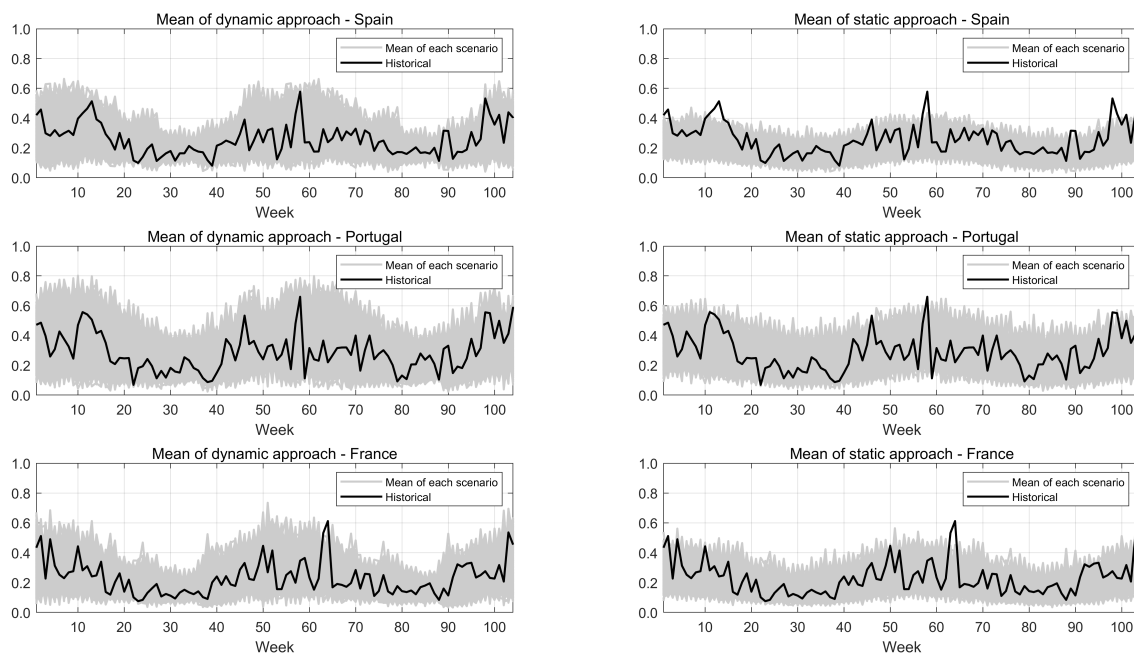
Figure 9 shows the mean obtained for each scenario and area compared with historical data, both for the dynamic and static approach. It can be seen that the dynamic model better reproduces the changes in the weekly behavior of the time series. This can be clearly evidenced between weeks 20 to 40, and 70 and 90.

**Table 7.** Pinball loss function for weekly approach.

| Area     | Approach | P5     | P30    | P50    | P70    | P95    | Mean          |
|----------|----------|--------|--------|--------|--------|--------|---------------|
| Spain    | Static   | 0.0071 | 0.0300 | 0.0396 | 0.0408 | 0.0231 | 0.0281        |
|          | Dynamic  | 0.0066 | 0.0254 | 0.0300 | 0.0269 | 0.0086 | <b>0.0195</b> |
| Portugal | Static   | 0.0092 | 0.0345 | 0.0423 | 0.0403 | 0.0151 | 0.0283        |
|          | Dynamic  | 0.0089 | 0.0329 | 0.0400 | 0.0372 | 0.0108 | <b>0.0260</b> |
| France   | Static   | 0.0060 | 0.0253 | 0.0320 | 0.0315 | 0.0138 | 0.0217        |
|          | Dynamic  | 0.0057 | 0.0241 | 0.0304 | 0.0284 | 0.0093 | <b>0.0196</b> |

**Table 8.** Winkler score for weekly approach.

| Area     | Approach | 50%  | 80%  | 90%  | 98%  | Mean        |
|----------|----------|------|------|------|------|-------------|
| Spain    | Static   | 0.26 | 0.43 | 0.60 | 1.42 | 0.68        |
|          | Dynamic  | 0.19 | 0.26 | 0.30 | 0.39 | <b>0.28</b> |
| Portugal | Static   | 0.28 | 0.40 | 0.49 | 0.77 | 0.48        |
|          | Dynamic  | 0.26 | 0.35 | 0.39 | 0.47 | <b>0.37</b> |
| France   | Static   | 0.21 | 0.31 | 0.40 | 0.66 | 0.39        |
|          | Dynamic  | 0.19 | 0.26 | 0.30 | 0.39 | <b>0.29</b> |



**Figure 9.** Comparison for mean on a weekly scale.

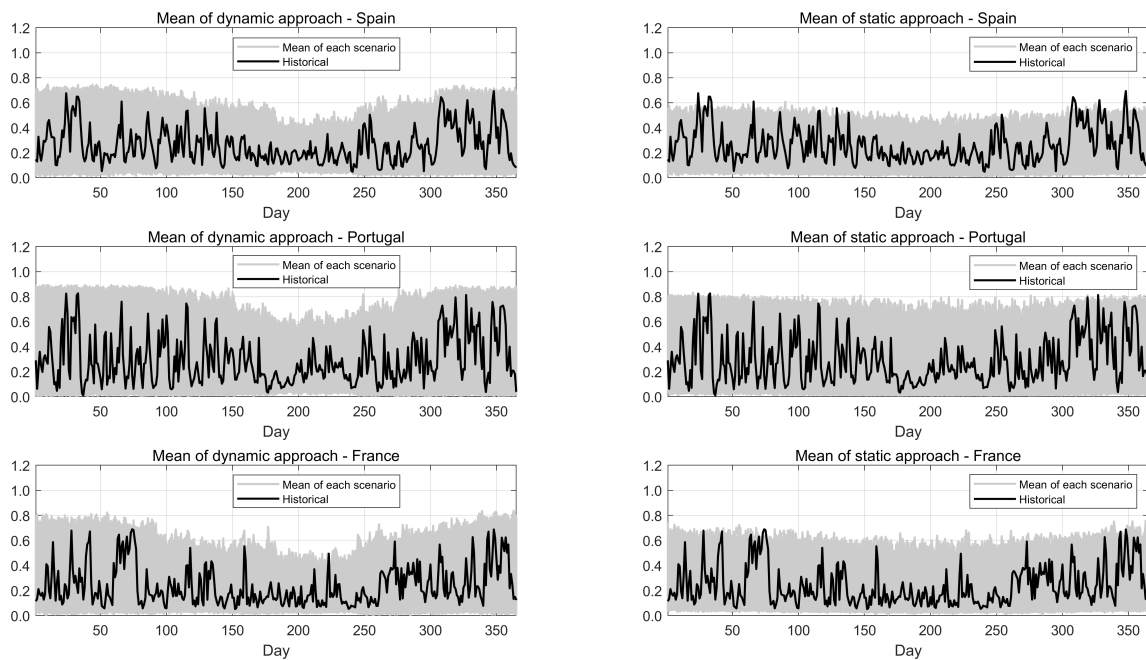
In the fourth stage, a daily evaluation has been performed. In this case, the scenarios have been generated for one year, and the mean of the wind generation for each day has been calculated for all the paths. The obtained results are shown in Tables 9 and 10. The best results are highlighted in bold. In both PLF and WS, the dynamic approach captures better the behavior of the time series at different percentiles and intervals on a daily scale. This can also be seen in Figure 10, where the daily means for each approach are shown and compared with the historical behavior. The figure shows that, for all the areas, the dynamic approach better follows the historical time series behavior.

**Table 9.** Pinball loss function for daily approach.

| Area     | Approach | P5     | P30    | P50    | P70    | P95    | Mean          |
|----------|----------|--------|--------|--------|--------|--------|---------------|
| Spain    | Static   | 0.0095 | 0.0413 | 0.0530 | 0.0515 | 0.0205 | 0.0352        |
|          | Dynamic  | 0.0089 | 0.0379 | 0.0471 | 0.0433 | 0.0128 | <b>0.0300</b> |
| Portugal | Static   | 0.0123 | 0.0535 | 0.0681 | 0.0642 | 0.0208 | 0.0438        |
|          | Dynamic  | 0.0118 | 0.0524 | 0.0664 | 0.0617 | 0.0186 | <b>0.0422</b> |
| France   | Static   | 0.0082 | 0.0372 | 0.0493 | 0.0488 | 0.0172 | 0.0321        |
|          | Dynamic  | 0.0080 | 0.0364 | 0.0480 | 0.0469 | 0.0153 | <b>0.0309</b> |

**Table 10.** Winkler score for daily approach.

| Area     | Approach | 50%  | 80%  | 90%  | 98%  | Mean        |
|----------|----------|------|------|------|------|-------------|
| Spain    | Static   | 0.34 | 0.49 | 0.60 | 0.96 | 0.60        |
|          | Dynamic  | 0.29 | 0.38 | 0.43 | 0.52 | <b>0.41</b> |
| Portugal | Static   | 0.43 | 0.58 | 0.66 | 0.82 | 0.62        |
|          | Dynamic  | 0.41 | 0.54 | 0.61 | 0.71 | <b>0.57</b> |
| France   | Static   | 0.31 | 0.43 | 0.51 | 0.65 | 0.48        |
|          | Dynamic  | 0.30 | 0.41 | 0.47 | 0.58 | <b>0.44</b> |

**Figure 10.** Comparison for mean at a daily scale.

Finally, a comparison of the correlations among areas have been performed. In particular, the time series of Spain and Portugal show the highest correlations, as expected due to the similarities among wind climate and topographical conditions [45]. However, these correlations could not be only due to the similarities in wind behavior but also because both are coupled in the MIBEL market. In contrast, energy exchanges with France are less coupled and they are limited by cross-border interconnections. Unlike the static approach, the dynamic model is able to capture the changes in dependency relationships at different scales throughout the year. Table 11 shows the correlations for the historical data and the dynamic approach, both on an hourly, daily, weekly, and monthly basis. It can be seen that the dynamic approach reproduces better the monthly correlations between Spain and Portugal than in the other scales. That is because the scenarios have been generated considering those monthly correlations. However, hourly, daily and weekly correlations are close to those observed in the historical data. Moreover, the lower correlations between Spain and Portugal with France are well captured with the dynamic approach as well.

To summarize, in this section, it has been shown that the dynamic approach proposed in this paper reproduces better than the static approach, the dynamics and the spatial-temporal dependencies observed in the historical data at different time scales. While the low value obtained for the PLF in the dynamic approach indicates that the distribution of the scenarios generated reproduces more accurately the historical behavior at different quantiles, and the low value obtained for the WS suggests that the forecasted scenarios achieve less error when different prediction intervals are considered.



In this way, the proposed model evidenced the robustness to simulate the idiosyncrasy of the historical time series of each area at different intervals, even when time series with dynamics so different as that have been used for the study case are considered. Finally, it is shown that the dynamic approach simulates closely the sudden changes observed in the mean of the historical data at different time scales with the highest reliability and that the scenarios generated preserve the correlation among areas observed in the historical data.

**Table 11.** Correlation comparison between historical data and dynamic approach.

| Areas   | Hourly |      | Daily |      | Weekly |      | Monthly |      |
|---------|--------|------|-------|------|--------|------|---------|------|
|         | Hist.  | Dyn. | Hist. | Dyn. | Hist.  | Dyn. | Hist.   | Dyn. |
| Spa-Por | 0.73   | 0.68 | 0.77  | 0.71 | 0.86   | 0.81 | 0.92    | 0.91 |
| Spa-Fra | 0.35   | 0.37 | 0.40  | 0.39 | 0.55   | 0.53 | 0.71    | 0.74 |
| Por-Fra | 0.24   | 0.25 | 0.29  | 0.28 | 0.48   | 0.45 | 0.70    | 0.70 |

#### 4. Conclusions

A new methodology to generate realistic scenarios for wind power generation in the long term has been proposed in this paper. Two approaches to include the spatial-temporal dependencies among different markets have been evaluated with a bottom-up approach. The results show firstly, that the ordered quantile transformation gives the best time series normalization for the study case when it is compared with the logit, Yeo-Johnson and Box-Cox transformations. Second, it is shown that a monthly segmentation helps to capture better the dynamics and the spatial-temporal dependencies observed in the historical data. Third, results obtained with both approaches have been compared using the Pinball Loss Function and the Winkler score metrics, showing that the dynamic approach reproduces more accurately the historical behavior of each area at different quantiles of the distribution function and prediction intervals. Fourth, it is shown that the dynamic approach simulates closely the sudden changes observed in the mean of the historical data at different time scales. Finally, the case study helps to understand that modeling the spatial-temporal dependencies is a key issue to simulate the aggregated wind power behavior in coupled electricity markets.

**Author Contributions:** Conceptualization, G.M., A.B., J.C., and J.R.; methodology, G.M., A.B., J.C., and J.R.; formal analysis, G.M.; investigation, G.M.; resources, G.M.; data curation, G.M.; writing—original draft preparation, G.M.; writing—review and editing, G.M., A.B., J.C., and J.R.; supervision, A.B. and J.R. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The authors declare no conflict of interest.

#### Abbreviations

The following abbreviations are used in this manuscript:

|        |  |
|--------|--|
| ACF    | Autocorrelation function                                 |
| AI     | Artificial Intelligence                                  |
| ARIMA  | Auto regressive integrated moving average model          |
| BIC    | Bayesian information criterion                           |
| MIBEL  | Iberian electricity market                               |
| PACF   | Partial autocorrelation function                         |
| PLF    | Pinball loss function                                    |
| QQ     | Quantile-quantile  |
| SARIMA | Seasonal auto regressive integrated moving average model |
| WS     | Winkler score  |

## References

1. Jung, J.; Broadwater, R.P. Current status and future advances for wind speed and power forecasting. *Renew. Sustain. Energy Rev.* **2014**, *31*, 762–777. [[CrossRef](#)]
2. Ren, G.; Liu, J.; Wan, J.; Guo, Y.; Yu, D. Overview of wind power intermittency: Impacts, measurements, and mitigation solutions. *Appl. Energy* **2017**, *204*, 47–65. [[CrossRef](#)]
3. Demolli, H.; Dokuz, A.S.; Ecemis, A.; Gokcek, M. Wind power forecasting based on daily wind speed data using machine learning algorithms. *Energy Convers. Manag.* **2019**, *198*, 111823. [[CrossRef](#)]
4. Agency, I.R.E. *Future of Wind: Deployment, Investment, Technology, Grid Integration and Socio-Economic Aspects*; Technical Report; IRENA: Abu Dhabi, UAE, 2019.
5. González, J.S.; Lacal-Arántegui, R. A review of regulatory framework for wind energy in European Union countries: Current state and expected developments. *Renew. Sustain. Energy Rev.* **2016**, *56*, 588–602. [[CrossRef](#)]
6. Bello, A.; Bunn, D.W.; Reneses, J.; Muñoz, A. Medium-term probabilistic forecasting of electricity prices: A hybrid approach. *IEEE Trans. Power Syst.* **2016**, *32*, 334–343. [[CrossRef](#)]
7. Eising, M.; Hobbie, H.; Möst, D. Future wind and solar power market values in Germany—Evidence of spatial and technological dependencies? *Energy Econ.* **2020**, *86*, 104638. [[CrossRef](#)]
8. Orgaz, A.; Bello, A.; Reneses, J. A New Model to Simulate Local Market Power in a Multi-Area Electricity Market: Application to the European Case. *Energies* **2019**, *12*, 2068. [[CrossRef](#)]
9. Zhao, Y.; Ye, L.; Pinson, P.; Tang, Y.; Lu, P. Correlation-constrained and sparsity-controlled vector autoregressive model for spatio-temporal wind power forecasting. *IEEE Trans. Power Syst.* **2018**, *33*, 5029–5040. [[CrossRef](#)]
10. Morales, J.M.; Minguez, R.; Conejo, A.J. A methodology to generate statistically dependent wind speed scenarios. *Appl. Energy* **2010**, *87*, 843–855. [[CrossRef](#)]
11. Villanueva, D.; Feijóo, A.; Pazos, J.L. Simulation of correlated wind speed data for economic dispatch evaluation. *IEEE Trans. Sustain. Energy* **2011**, *3*, 142–149. [[CrossRef](#)]
12. Shi, J.; Qu, X.; Zeng, S. Short-term wind power generation forecasting: Direct versus indirect ARIMA-based approaches. *Int. J. Green Energy* **2011**, *8*, 100–112. [[CrossRef](#)]
13. Kim, Y.; Hur, J. An Ensemble Forecasting Model of Wind Power Outputs based on Improved Statistical Approaches. *Energies* **2020**, *13*, 1071. [[CrossRef](#)]
14. Singh, S.; Mohapatra, A. Repeated wavelet transform based ARIMA model for very short-term wind speed forecasting. *Renew. Energy* **2019**, *136*, 758–768.
15. Kennedy, S.; Rogers, P. A probabilistic model for simulating long-term wind-power output. *Wind. Eng.* **2003**, *27*, 167–181. [[CrossRef](#)]
16. Kritharas, P.P.; Watson, S.J. A comparison of long-term wind speed forecasting models. *J. Sol. Energy Eng.* **2010**, *132*, 1–8. [[CrossRef](#)]
17. Khan, G.M.; Ali, J.; Mahmud, S.A. Wind power forecasting—An application of machine learning in renewable energy. In Proceedings of the 2014 International Joint Conference on Neural Networks (IJCNN), Beijing, China, 6–11 July 2014; IEEE: Piscataway, NJ, USA, 2014; pp. 1130–1137.
18. Barbosa de Alencar, D.; de Mattos Affonso, C.; Limão de Oliveira, R.C.; Moya Rodriguez, J.L.; Leite, J.C.; Reston Filho, J.C. Different models for forecasting wind power generation: Case study. *Energies* **2017**, *10*, 1976. [[CrossRef](#)]
19. Wan, C.; Xu, Z.; Pinson, P.; Dong, Z.Y.; Wong, K.P. Probabilistic forecasting of wind power generation using extreme learning machine. *IEEE Trans. Power Syst.* **2013**, *29*, 1033–1044. [[CrossRef](#)]
20. Chen, C.; Liu, H. Medium-term wind power forecasting based on multi-resolution multi-learner ensemble and adaptive model selection. *Energy Convers. Manag.* **2020**, *206*, 112492. [[CrossRef](#)]
21. Wang, G.; Jia, R.; Liu, J.; Zhang, H. A hybrid wind power forecasting approach based on Bayesian model averaging and ensemble learning. *Renew. Energy* **2020**, *145*, 2426–2434. [[CrossRef](#)]
22. Sandhu, K.; Nair, A.R. A comparative study of ARIMA and RNN for short term wind speed forecasting. In Proceedings of the 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kanpur, India, 6–8 July 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–7.
23. Chen, Q.; Folly, K.A. Comparison of three methods for short-term wind power forecasting. In Proceedings of the 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, Brazil, 8–13 July 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 1–8.

24. Sim, S.K.; Maass, P.; Lind, P.G. Wind speed modeling by nested ARIMA processes. *Energies* **2019**, *12*, 69. [[CrossRef](#)]
25. Eldali, F.A.; Hansen, T.M.; Suryanarayanan, S.; Chong, E.K. Employing ARIMA models to improve wind power forecasts: A case study in ERCOT. In Proceedings of the 2016 North American Power Symposium (NAPS), Denver, CO, USA, 18–20 September 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 1–6.
26. Tsai, A.C.; Liou, M.; Simak, M.; Cheng, P.E. On hyperbolic transformations to normality. *Comput. Stat. Data Anal.* **2017**, *115*, 250–266. [[CrossRef](#)]
27. Peterson, R.A.; Cavanaugh, J.E. Ordered quantile normalization: A semiparametric transformation built for the cross-validation era. *J. Appl. Stat.* **2019**, 1–16. [[CrossRef](#)]
28. Casula, L.; D’Amico, G.; Masala, G.; Petroni, F. Performance estimation of a wind farm with a dependence structure between electricity price and wind speed. *World Econ.* **2020**, 1–25. [[CrossRef](#)]
29. Dunnett, S.; Sorichetta, A.; Taylor, G.; Eigenbrod, F. Harmonised global datasets of wind and solar farm locations and power. *Sci. Data* **2020**, *7*, 1–12. [[CrossRef](#)] [[PubMed](#)]
30. Christensen, T.S.; Pircalabu, A.; Høg, E. A seasonal copula mixture for hedging the clean spark spread with wind power futures. *Energy Econ.* **2019**, *78*, 64–80. [[CrossRef](#)]
31. Hyndman, R.J.; Athanasopoulos, G. *Forecasting: Principles and Practice*; OTexts: Clayton, Australia, 2018.
32. Hering, A.S.; Genton, M.G. Powering up with space-time wind forecasting. *J. Am. Stat. Assoc.* **2010**, *105*, 92–104. [[CrossRef](#)]
33. Guidolin, M.; Guseo, R. Modelling seasonality in innovation diffusion. *Technol. Forecast. Soc. Chang.* **2014**, *86*, 33–40. [[CrossRef](#)]
34. Benth, J.Š.; Benth, F.E. Analysis and modelling of wind speed in New York. *J. Appl. Stat.* **2010**, *37*, 893–909. [[CrossRef](#)]
35. Taylor, J.W.; McSharry, P.E.; Buizza, R. Wind power density forecasting using ensemble predictions and time series models. *IEEE Trans. Energy Convers.* **2009**, *24*, 775–782. [[CrossRef](#)]
36. Cadenas, E.; Rivera, W.; Campos-Amezcuca, R.; Heard, C. Wind speed prediction using a univariate ARIMA model and a multivariate NARX model. *Energies* **2016**, *9*, 109. [[CrossRef](#)]
37. Box, G.E.; Jenkins, G.M.; Reinsel, G.C.; Ljung, G.M. *Time Series Analysis: Forecasting and Control*; John Wiley & Sons: Hoboken, NJ, USA, 2015.
38. Pobočíková, I.; Sedláčková, Z.; Michalková, M. Application of four probability distributions for wind speed modeling. *Procedia Eng.* **2017**, *192*, 713–718. [[CrossRef](#)]
39. Liu, B.; Nowotarski, J.; Hong, T.; Weron, R. Probabilistic Load Forecasting via Quantile Regression Averaging on Sister Forecasts. *IEEE Trans. Smart Grid* **2017**, *8*, 730–737. [[CrossRef](#)]
40. Bello, A.; Reneses, J.; Muñoz, A.; Delgado, A. Probabilistic forecasting of hourly electricity prices in the medium-term using spatial interpolation techniques. *Int. J. Forecast.* **2016**, *32*, 966–980. [[CrossRef](#)]
41. Hong, T.; Pinson, P.; Fan, S.; Zareipour, H.; Troccoli, A.; Hyndman, R.J. Probabilistic Energy Forecasting: Global Energy Forecasting Competition 2014 and Beyond. *Int. J. Forecast.* **2016**, *32*, 896–931. [[CrossRef](#)]
42. Gneiting, T.; Balabdaoui, F.; Raftery, A.E. Probabilistic forecasts, calibration and sharpness. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **2007**, *69*, 243–268. [[CrossRef](#)]
43. Red Eléctrica de España. International Interconnections. Available online: <https://www.ree.es/en/activities/operation-of-the-electricity-system/international-interconnections> (accessed on 15 May 2020).
44. Entsoe. Installed Capacity per Production Type. Available online: <https://transparency.entsoe.eu> (accessed on 15 May 2020).
45. Troen, I.; Petersen, E.L. *European Wind Atlas*; Risø National Laboratory: Roskilde, Denmark, 1989.

