

Article

## An Advanced Bayesian Method for Short-Term Probabilistic Forecasting of the Generation of Wind Power

Antonio Bracale <sup>1,†</sup> and Pasquale De Falco <sup>2,†,\*</sup>

<sup>1</sup> Department of Engineering, University of Naples Parthenope, Centro Direzionale Is. C4, Naples 80143, Italy; E-Mail: antonio.bracale@uniparthenope.it

<sup>2</sup> Department of Electrical Engineering and Information Technologies, University of Naples Federico II Via Claudio 21, Naples 80125, Italy

† These authors contributed equally to this work.

\* Author to whom correspondence should be addressed; E-Mail: pasquale.defalco2@unina.it; Tel.: +39-081-768-3203.

Academic Editor: Enrico Sciubba

Received: 17 June 2015 / Accepted: 11 September 2015 / Published: 21 September 2015

---

**Abstract:** Currently, among renewable distributed generation systems, wind generators are receiving a great deal of interest due to the great economic, technological, and environmental incentives they involve. However, the uncertainties due to the intermittent nature of wind energy make it difficult to operate electrical power systems optimally and make decisions that satisfy the needs of all the stakeholders of the electricity energy market. Thus, there is increasing interest determining how to forecast wind power production accurately. Most the methods that have been published in the relevant literature provided deterministic forecasts even though great interest has been focused recently on probabilistic forecast methods. In this paper, an advanced probabilistic method is proposed for short-term forecasting of wind power production. A mixture of two Weibull distributions was used as a probability function to model the uncertainties associated with wind speed. Then, a Bayesian inference approach with a particularly-effective, autoregressive, integrated, moving-average model was used to determine the parameters of the mixture Weibull distribution. Numerical applications also are presented to provide evidence of the forecasting performance of the Bayesian-based approach.

**Keywords:** wind energy; power production; forecasting methods; probabilistic approaches

---

## 1. Introduction

Currently, there is a vast amount of research, as well as many new visions and concepts, concerning future electrical systems [1–6]. For example, super grids, smart grids, micro grids, intelligent grids, active networks, and virtual power plants are becoming the keywords related to the future development of power systems. In this context, it is expected that the penetration of distributed generation (DG) systems, especially those based on renewable energy sources, will become increasingly important in future Smart Grids due to environmental and technical reasons. In fact, the presence of DG systems in Smart Grids will result in advantages for energy-users and for the social wellness [7,8].

The foreseeable extensive use of DG systems in the future requires that distribution system engineers properly account for their impact in the system. In fact, their interconnection with the system significantly alters the characteristics of the distribution systems, traditionally designed with the assumption of a passive network. The consequence of the presence of DG systems is that the assumption of a passive network is no longer valid; instead, the network becomes active, which generates a number of new technical considerations that must be addressed, such as distribution network planning and operation, especially protection coordination, steady-state analysis, and power quality issues.

Among renewable DG systems, wind generators currently are receiving a great deal of interest; in fact, the unsubsidized cost of energy at the bus has decreased by more than 80% [9].

Wind energy is a new, emerging research field characterized by a high degree of interdisciplinary studies, and there are several related topics of interest in the relevant literature.

Increasing the quality and value of wind power generation will be one of the priorities in wind energy research in the coming years, and this requires that we improve our ability to predict the performance of wind systems [10]. In fact, accurate forecasting is needed to solve several distribution system engineers' problems, and in particular to allow for unit commitment and the provision of ancillary services in the framework of competitive electricity markets as well as for the scheduling and dispatch of the required hourly ramping and load following [11]. Then, accurate and reliable forecasts are mandatory for the optimal design and management of the Smart Grid resources.

In the relevant literature, several wind forecasting methods have been proposed with different levels of success [12]. These include physical, statistical, artificial neural network, and hybrid methods, which differ in the use of input and output data, as well as in the time horizon of their application.

In particular, deterministic and probabilistic forecasting techniques have been provided depending on the type of information on the predicted output. In deterministic forecasting, a single value is provided without any other information about the nature of the wind's uncertainties; in probabilistic forecasting, the output value is accompanied by information on the wind's random nature. A review of existing methods is reported in [12].

Recently, interest in probabilistic forecasting has been increasing because of the need to take into account the unavoidable uncertainties that characterize the availability of wind energy resources [13,14].

On the other hand, deterministic forecast does not meet various applications needs, such as power system operations where uncertainties and risks have to be quantified [14,15]. In particular, given the significant variability of the level of forecasting errors, forecast-users usually need additional information about forecast uncertainty; in fact, this additional information, given *i.e.*, in terms of risk indices or quantile or interval forecasts, can be introduced in users' decision-making processes.

In this paper, a new method for short-term probabilistic forecasting is proposed that directly supports the probabilistic representation of the predicted wind power output. The proposed method uses the classical relationships that link wind active power to wind speed, the probability density function (PDF) of which is predicted by applying the Bayesian inference (BI) approach.

The BI approaches were extensively used to improve the forecasts for economic, social and weather time series [16–18] and in relevant literature they were demonstrated to significantly improve the forecasts obtained through AutoRegressive Integrated Moving Average (ARIMA) models, by estimating the parameters from a probabilistic point of view [19–21]. Bayesian methods are being used increasingly in wind energy conversion systems due to the significant advantages they offer when uncertainty and variability are predominant concerns [17]. They have been used in several fields of interest to wind energy engineers, such as forecasting long- and short-term production, modeling extreme wind conditions, evaluating the reliability of the systems, and in the process of designing the systems' components. However, the applications of wind energy are still in the early stage, and they are limited in number, so it can be expected that their use will increase as much more attention is paid to both methodologies and applications [17].

As is well known [22], two of the key steps of a Bayesian-based method for the short-term, probabilistic forecasting of wind speed are (i) the choice of the analytical expressions of the PDF modeling the uncertainties associated with wind speed and (i) defining the best time series model to determine the PDF parameters that are not assumed to be prior random parameters of the Bayesian approach.

Some studies addressing these steps were presented in [16,19,23,24]. In [19], a Gaussian PDF was used to model wind speed, and a sixth-order autoregressive model that involved only wind speed was used. In [23], a mixture model was used with a normal distribution that fit the values around the stall speed, and a Weibull distribution was used to fit the remaining values. In [16,24], a Weibull distribution was used to model wind speed, after which a first-order autoregressive model that involved the mean value of the wind speed was used. A review of existing Bayesian applications to short-term wind forecasting is reported in Section 2.3 of Reference [17]. In this paper, a more complex PDF for modeling the uncertainties of wind speed and a particularly effective approach are used to identify the most adequate time series model. In particular, a mixture distribution of two Weibull distributions is used as the PDF analytical expression.

We considered a mixture distribution of two Weibull distributions because it was concluded in [25] that the use of the classical Weibull distribution of two parameters cannot represent all of the wind regimes encountered in nature, such as, for example, those with bimodal distributions. Therefore, a more suitable PDF must be selected for each wind regime in order to minimize errors in the estimation of the energy produced. A mixture of two Weibull distributions seemed most suitable for both unimodal and bimodal wind regimes, and it was evaluated experimentally for some actual cases.

Concerning the autoregressive, integrated, moving-average time series model, we applied the Box-Jenkins approach based on the use of the sample autocorrelation function [26]. This approach is particularly effective in determining the orders and the parameters of the model itself.

The aims of the research reported in this paper are (i) to propose a new Bayesian-based method for the short-term forecasting of wind power; (ii) to include a new probability function and improved time series models in the frame of the Bayesian method; and (iii) to conduct a critical comparison of the performances of the new Bayesian method with both a traditional Bayesian method and a reference

predictor (the probabilistic persistence method) in order to outline the advantages and disadvantages of the proposed method.

Note that there exists a large extend of literature on wind power forecasting, including several state-of-the-art papers, that evidences that numerical weather prediction (NWP) models have enabled relatively accurate wind forecasts [27–32]. However, as the operating time moves closer to the near-term (e.g., hour-ahead or 15 minute-ahead), at a high spatial resolution, the computation complexity (in terms of simulation time and memory requirements) often renders NWP models intractable [31,33]. In sharp contrast, data-driven statistical model is thought to be the most competitive method for near-term wind forecasting problems being able to capture the rapidly changing dynamics of the atmosphere and with nice model interpretation [32]. Then, our proposed probabilistic method is targeted directly at computationally-efficient, near-term wind forecasts (e.g., hour-ahead or 15 minutes-ahead forecasts). The emphasis in our work was on computational efficiency because computational complexity (in terms of simulation time and memory requirements) often makes numerical weather prediction models excessively burdensome and expensive to operate [34].

This paper is organized as follows. Section 2 describes the probabilistic method we used that was based on Bayesian theory. In Section 3, the results of the numerical applications of the proposed method are reported, and they are discussed and compared with the results provided by both a traditional Bayesian method that uses distributions of two parameters and a probabilistic extension of the persistence method in order to show the advantages and benefits of the proposed method.

## 2. A Probabilistic Approach for Forecasting Wind Power Production: The Bayesian-Based Method

In the research reported in this paper, the Bayesian-based method was used to predict the PDF of the active power generated by wind systems. In particular, a relationship linking the wind active power with the wind speed was selected. Then, using the selected relationship in the frame of a Monte Carlo simulation approach, we forecasted the PDF of the active power production at the time horizon  $t = h$  standing at the origin time  $t = h - k$ , where  $k$  is the lead time, starting from the evaluation of the PDF of the wind speed at time  $t = h$ . The forecast of the PDF of the wind speed at time step  $t = h$  was obtained by selecting an appropriate analytical expression for the wind speed PDF and evaluating the PDF parameters by applying the BI with an autoregressive, integrated, moving-average, time-series model.

Details about the various steps of this method are reported in the following subsections; for the sake of conciseness, only the results of the numerical applications with reference to the case of  $k = 1$  are shown in Section 3.

### 2.1. Description of the Relationship that Links Wind Active Power with Wind Speed

In the most general case, the wind active power depends not only on wind speed, but also on meteorological variables such as wind direction, temperature, local air density, and precipitation. Moreover, the behavior of power curves when the wind speed increases can be different from the behavior when the speed decreases. We should also consider that, in many cases, there is the problem to predict wind power for an entire wind farm so that the choice of a deterministic power curve can be complicated by the fact that the wind turbines in a wind farm can have different cut-in and rated speeds;

finally, there may be changes in the power capacity of the wind farm due to the addition of new turbines and turbine maintenance [35–39].

However, a deterministic power curve is assumed frequently in the relevant literature [31,35–37,40–44]. In fact, manufacturers typically provide information regarding the power curve assuming fixed air density and standard environmental variables.

Considering the wind active power as a random variable dependent not only on wind speed, but also on other explicative variables (such as wind direction or air density) would require complexity increase in the Bayesian inference, given the augmented number of parameters to be estimated. Therefore, since our aim was to propose a computationally-efficient forecasting tool, we decided to use the deterministic power curve furnished by manufacturers.

The following analytical relationship between the active power,  $P_{W_h}$ , and wind speed,  $w_h$ , at the time horizon  $h$  can be written:

$$P_{W_h} = \begin{cases} 0 & \text{if } 0 \leq w_h \leq W_{ci} \\ g(w_h) & \text{if } W_{ci} < w_h \leq W_r \\ P_{max} & \text{if } W_r < w_h \leq W_{co} \\ 0 & \text{if } w_h > W_{co} \end{cases} \quad (1)$$

where  $g(w_h)$  is a non-linear function usually approximated by a linear function, linear pieces, a parabolic function, or a cubic function; and  $W_{ci}$ ,  $W_r$ , and  $W_{co}$  are the cut-in, rated, and cut-off characteristic values, respectively, of the wind turbine power generation unit.

## 2.2. Selection of the Analytical Expression of the PDF of the Wind Speed

As is well known, wind speed is frequently modeled using the Weibull distribution (WB), as reported in [27]:

$$f_{0_{w_h}}(w_h | \eta_{0_h}, \beta_{0_h}) = \frac{\beta_{0_h}}{\eta_{0_h}} \left( \frac{w_h}{\eta_{0_h}} \right)^{\beta_{0_h}-1} e^{-\left( \frac{w_h}{\eta_{0_h}} \right)^{\beta_{0_h}}} \quad (2)$$

where  $\eta_{0_h}$  is the scale parameter and  $\beta_{0_h}$  is the shape parameter. The scale parameter  $\eta_{0_h}$  can be expressed in terms of the mean value  $\mu_{0_{w_h}}$  of the distribution of the wind speed, according to the following relationship:

$$\eta_{0_h} = \frac{\mu_{0_{w_h}}}{\Gamma\left(1 + \frac{1}{\beta_{0_h}}\right)} \quad (3)$$

where  $\Gamma(\cdot)$  is the Gamma function. Consequently, one can treat the PDF in Equation (2) as a function of the mean value and the shape factor.

In [25], it was concluded that the Weibull distribution of two parameters presents a series of advantages that simply its use, *i.e.*, (i) flexibility; (ii) dependence on only two parameters; (iii) the simplicity of the estimation of its parameters; and (iv) its closed form. However, the Weibull PDF cannot represent all the wind regimes encountered in nature, *e.g.*, those with bimodal distributions. The mixture of two Weibull distributions can be particularly suitable for these wind regimes.

As is well known, mixture density is a probability density function that is a convex linear combination of other probability density functions [45,46].

A two-component mixture Weibull distribution (MWB) depends on five parameters  $(\omega_h, \eta_{1h}, \beta_{1h}, \eta_{2h}, \beta_{2h})$  and is given by:

$$f_{1_{w_h}}(w_h | \omega_h, \eta_{1h}, \beta_{1h}, \eta_{2h}, \beta_{2h}) = \omega_h \left[ \frac{\beta_{1h}}{\eta_{1h}} \left( \frac{w_h}{\eta_{1h}} \right)^{\beta_{1h}-1} e^{-\left(\frac{w_h}{\eta_{1h}}\right)^{\beta_{1h}}} \right] + (1 - \omega_h) \left[ \frac{\beta_{2h}}{\eta_{2h}} \left( \frac{w_h}{\eta_{2h}} \right)^{\beta_{2h}-1} e^{-\left(\frac{w_h}{\eta_{2h}}\right)^{\beta_{2h}}} \right] \tag{4}$$

with  $0 \leq \omega_h \leq 1$ .

The scale parameter  $\eta_{1h}$  in Equation (4) can be expressed in terms of the mean value  $\mu_{1_{w_h}}$  of the distribution of the wind speed and of the other parameters  $\omega_h, \beta_{1h}, \eta_{2h}, \beta_{2h}$ , according to the following relationship [25]:

$$\eta_{1h} = \frac{\mu_{1_{w_h}}^{-(1-\omega_h)\eta_{2h}} \Gamma\left(1 + \frac{1}{\beta_{2h}}\right)}{\omega_h \Gamma\left(1 + \frac{1}{\beta_{1h}}\right)} \tag{5}$$

As a result of the analysis of Equations (4) and (5), for the time horizon  $h$ , the estimation of the mean value  $\mu_{1_{w_h}}$  and of the parameters  $\omega_h, \beta_{1h}, \eta_{2h}, \beta_{2h}$  is sufficient to unequivocally predict the probability density function  $f_{1_{w_h}}$ . In this paper, the parameters  $\omega_h, \beta_{1h}, \eta_{2h}, \beta_{2h}$  were assumed to be prior random parameters of the Bayesian approach, while the mean value  $\mu_{1_{w_h}}$  was estimated using the AutoRegressive Moving Average (ARMA) and ARIMA time-series models reported in the next subsection.

Note that we considered both relationships Equations (1) and (4) separately, because they were to be used in a Monte Carlo simulation approach that can handle them very easily.

### 2.3. ARMA and ARIMA Time-series Models

The general ARMA family for a stochastic variable  $x_t$  can be represented as [26]:

$$\Phi(B)x_t = \theta_0 + \theta(B)e_t \tag{6}$$

where:

- $B$  is the backward shift operator, defined by  $B^m x_t = x_{t-m}$ ;
- $\Phi(B)$  is the stationary autoregressive operator of order  $p$ , defined by  $\Phi(B) = 1 - \Phi_1 B - \Phi_2 B^2 - \dots - \Phi_p B^p$ , fulfilling the condition that all of the roots of the polynomial  $\Phi(B)$  must be greater than unity;
- $\theta_0$  is a constant term;
- $\theta(B)$  is the moving average operator of order  $q$ ; it is  $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$ ;
- $e_t$  is the white noise at time  $t$ , characterized by a null mean and constant variance  $\sigma_e^2$ .

Expanding Equation (6) in terms of past values of  $x_t$  and  $e_t$ , we obtain the following form of the difference equation:

$$x_t = \theta_0 + \Phi_1 x_{t-1} + \dots + \Phi_p x_{t-p} - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q} + e_t \tag{7}$$

Therefore, an ARMA model is unequivocally determined by fixing its orders  $(p, q)$ , and the  $p + q + 2$  unknown parameters  $\theta_0, \Phi_1, \dots, \Phi_p, \theta_1, \dots, \theta_q, \sigma_e$ . ARMA models represent linear, stationary stochastic processes mathematically, but these models usually perform poorly when fitting non-stationary processes.

Unfortunately, some time-series can present non-stationary characteristics. To obtain a better mathematical representation of such time-series, an extended version of the ARMA model must be used in order to take into account the past values of the stochastic variable  $x_t$  and the differences among actual and past values of the stochastic variable, *i.e.*,  $(x_t - x_{t-1})$ .

Such models belong to the ARIMA family, and, for the generic stochastic variable  $x_t$ , they can be represented as:

$$\Phi(B)\nabla^d x_t = \theta_0 + \theta(B)e_t \quad (8)$$

where  $\nabla^d$  is the backward difference operator defined by  $\nabla^d x_t = x_t - x_{t-d}$ . Note that the polynomial  $\Phi(B)$  must satisfy the condition of stationary mentioned above.

Expanding Equation (8) in terms of past values of  $x_t$  and  $e_t$ , we obtain the following form of the difference equation:

$$x_t = \theta_0 + \varphi_1 x_{t-1} + \dots + \varphi_{p+d} x_{t-p-d} - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q} + e_t \quad (9)$$

where the coefficients  $\varphi_1, \dots, \varphi_{p+d}$  are the coefficients of the operator  $\varphi(B) = \Phi(B)(1 - B)^d = 1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_{p+d} B^{p+d}$ . In practice, the polynomial  $\varphi(B)$  can be separated into two contributions, *i.e.*, the polynomial  $(1 - B)^d$  that has  $d$  solutions equal to unity and the polynomial  $\Phi(B)$  that presents the aforesaid stationary requirements consisting of all of the roots of  $\Phi(B)$  to be greater than unity.

Therefore, an ARIMA model is determined unequivocally by fixing its orders  $(p, d, q)$  and the  $p + d + q + 2$  unknown parameters  $\theta_0, \varphi_1, \dots, \varphi_{p+d}, \theta_1, \dots, \theta_q, \sigma_e$ . Note that the ARIMA family includes the ARMA family in the particular case of  $d = 0$ ; so, one can use a general methodology for the identification of an ARIMA model to represent an examined time-series presenting either stationary characteristics ( $d$  equal to 0) or non-stationary characteristics ( $d$  not equal to 0).

In [26], Box and Jenkins proposed different techniques for the identification of the orders  $(p, d, q)$  of an ARIMA model; in this paper, we used the Box-Jenkins approach based on the use of the sample autocorrelation function  $r_x(l)$ , which is an estimation of the following theoretical autocorrelation function  $\rho_x(l)$  at different lags  $l$ :

$$\rho_x(l) = \frac{E[(x_t - \mu_x)(x_{t+l} - \mu_x)]}{\sigma_x^2} \quad (10)$$

where  $\mu_x$  and  $\sigma_x^2$  are the theoretical mean and the theoretical variance of the stochastic variable  $x_t$ , respectively. Since time-series always consist of a finite number of samples,  $N$ , only an estimation  $\hat{\rho}_x(l)$  of  $\rho_x(l)$  can be provided as follows:

$$\hat{\rho}_x(l) = \frac{\sum_{t=1}^{N-l} (x_t - \hat{\mu}_x)(x_{t+l} - \hat{\mu}_x)}{\sum_{t=1}^N (x_t - \hat{\mu}_x)^2} \quad (11)$$

where  $\hat{\mu}_x$  is the sample mean of the time-series.

The first step of the Box-Jenkins approach is to identify the degree of differencing,  $d$ , exploiting the properties of the autocorrelation functions. In fact, for a stationary time-series, the sample autocorrelation function  $\hat{\rho}_x(l)$  quickly decays to zero for moderate lags  $l$ , while the non-stationary characteristics in an examined time-series can be observed by the fact that the sample autocorrelation function  $\hat{\rho}_x(l)$  decreases very slowly and does not tend to reach zero even for large lags  $l$ . This fact suggests that:

- if the sample autocorrelation function  $\hat{\rho}_x(l)$  decreases quickly for increasing values of  $l$ , the time-series can be represented by a stationary model, and therefore  $d$  is assumed to be equal to zero;
- if the sample autocorrelation function  $\hat{\rho}_x(l)$  does not decrease quickly for increasing values of  $l$ , the stochastic process is supposed to be non-stationary in  $x_t$  but stationary in  $\nabla^d x_t$  for  $d \geq 1$ . Specifically, the stochastic process  $y_t = \nabla^d x_t$  is studied iteratively for  $d = 1, 2, \dots$ ; at each iteration, the autocorrelation function  $\hat{\rho}_y(l)$  of  $y_t = \nabla^d x_t$  is investigated, and the iterative process is stopped when the autocorrelation function  $\hat{\rho}_y(l)$  decreases quickly for increasing values of  $l$ . Therefore,  $d$  is assumed to be equal to the number of the iteration that achieved this result; in practice,  $d$  is normally equal to 1 or 2, and is sufficient to inspect the first 20 estimated autocorrelation coefficients ( $l = 1, 2, \dots, 20$ ) of the original series and of its first and second differences to determine the value of  $d$ .

Once the value of the differencing order,  $d$ , is selected, the appropriately-differenced time-series  $y_t = (1 - B)^d x_t$  shows characteristics of a stationary process; therefore, it can be modeled by an ARMA process of order  $(p, q)$ . Having built the time-series  $y_t$  in such a way, the ARMA  $(p, q)$  process representing  $y_t$  and the ARIMA  $(p, d, q)$  process representing the original time-series  $x_t$  share the same orders  $p, q$ ; therefore, in the second step of the Box-Jenkins approach, one can study the differenced time-series  $y_t$ , and, by fixing the orders  $p, q$  of the correspondent ARMA model, the orders  $p, q$  of the original ARIMA model also are individuated automatically.

Specifically, in [26], it was shown that different behaviors of the autocorrelation function  $\hat{\rho}_y(l)$  for the differenced series  $y_t$  suggest different values of  $(p, q)$ , and Table 1 reports the values for the most common time-series.

**Table 1.** Behavior of the sample autocorrelation function,  $\hat{\rho}_y(l)$ , for the  $d^{\text{th}}$  difference of an ARIMA process of order  $(p, d, q)$ .

Order of the ARIMA Model				
$(1, d, 0)$	$(0, d, 1)$	$(2, d, 0)$	$(0, d, 2)$	$(1, d, 1)$
$\hat{\rho}_y$ decreases exponentially	$\hat{\rho}_y(1)$ is the only appreciable non-zero term	$\hat{\rho}_y$ is a mixture of exponential functions or sine waves	$\hat{\rho}_y(1), \hat{\rho}_y(2)$ are the only appreciable non-zero terms	$\hat{\rho}_y$ decreases exponentially after $\hat{\rho}_y(1)$

Once the three orders  $(p, d, q)$  of the ARIMA process have been determined, a consolidated estimation procedure can be used to obtain estimates of the  $p + q + 2$  unknown parameters  $\theta_0, \Phi_1, \dots, \Phi_p, \theta_1, \dots, \theta_q, \sigma_e$  in Equation (9), which unequivocally identify the time-series model.

In this paper, the parameters of the ARIMA model were evaluated by minimizing the unconditional log-likelihood function of samples of  $x_t$  via the unconditional least squares estimates reported in [26].

Here, the stochastic variable,  $x_t$ , was assumed to be the wind speed,  $w_t$ . Moreover, once the minimum mean square error forecast of the wind speed for the time horizon  $h$  was obtained by using the appropriately-estimated ARIMA model, we assumed it to be the expected value of the forecasted distribution, *i.e.*, the mean value  $\mu_{1,w_h}$  to be included in Equation (5).



#### 2.4. Evaluation of the PDFs of the Parameters $\omega_h, \beta_{1h}, \eta_{2h}$ , and $\beta_{2h}$

Once the mean value of the distribution at the time horizon  $h$  were determined as described in Section 2.3, the remaining parameters  $\omega_h, \beta_{1h}, \eta_{2h}, \beta_{2h}$  of the distribution in Equation (4) must be obtained. The BI allows the probabilistic estimation of these parameters, identifying their joint posterior probability distribution, by the inference of an array of observations upon the known (or hypothesized) prior probability distributions of each parameter.

The full procedure that is used is described as follows.

The set  $\underline{S}_{W_h} = \{w_{h-k-M+1}, \dots, w_{h-k}\}$ , composed of  $M$  measurements of wind speed observed until the origin time  $h - k$ , is provided initially. In addition, the prior distributions of the parameters are chosen.

Let  $z_h$  be the generic parameter whose prior distribution must be provided for the time horizon  $h$ ; the parameters of the prior distributions commonly are called hyperparameters. There is a great debate in the relevant literature [47] concerning how to determine the type of the prior distribution of the parameter  $z_h$  and the corresponding hyperparameters. For example, when little or no prior information is provided on the parameter  $z_h$ , an uninformative distribution, such as Jeffreys prior or uniform distribution, is commonly used. However, when some prior statistical information is provided on the parameter  $z_h$ , an informative, appropriate distribution can be used, such as the Gaussian PDF with hyperparameters  $(\hat{\mu}_{z_h}, \hat{\sigma}_{z_h}^2)$ .

The main advantage of the Gaussian distribution is the simplicity of the operations in that only the estimates of two hyperparameters are needed, and one can fix the variance  $\hat{\sigma}_{z_h}^2$  immediately on the basis of her or his confidence in the estimate of the mean value  $\hat{\mu}_{z_h}$ ; a large variance yields a larger, more uniform distribution around the mean value, while a small variance yields a distribution that is more concentrated around the mean value. Coherently, with the behavior of uninformative distributions, specifying a large variance  $\hat{\sigma}_{z_h}^2$  ensures that the historical data used for the inference determines the relevant changes in the posterior distribution of  $z_h$  to a greater extent than the prior distribution [47].

In this paper, an initial estimation  $\hat{z}_h$  was performed for each time horizon  $h$  for each parameter  $\beta_{1h}, \eta_{2h}, \beta_{2h}$ , by applying the well-known moment estimation procedure [48] on the set of observations of wind speed  $\underline{S}_{W_h}$ . Then, the resulting value of each estimate was assumed to be equal to the mean value of the corresponding prior informative Gaussian distribution, *i.e.*,  $\hat{z}_h = \hat{\mu}_{z_h}$ ; then, the variance was assumed to be very high ( $\hat{\sigma}_{z_h} = 10^4$  for each parameter  $\beta_{1h}, \eta_{2h}, \beta_{2h}$ , as in [23]) in order to ensure that the historical data used for the inference, more than the prior distribution, determines the relevant changes in the posterior distribution of parameters.

Instead, for the parameter  $\omega_h$ , a completely uninformative uniform distribution in the interval from 0 to 1 was chosen due the restricted domain in which the parameter is defined.

Now, let  $\underline{z}_h = \{\omega_h, \beta_{1h}, \eta_{2h}, \beta_{2h}\}$  be the random parameter vector to be estimated for each time horizon  $h$  in the BI approach.

Once the prior distributions are set, the BI allows the estimation of the joint posterior distribution  $p(\underline{z}_h | \underline{S}_{W_h})$ , given the set of measurements  $\underline{S}_{W_h}$ , through the extension of the Bayes' Theorem. Unfortunately, a closed-form expression of  $p(\underline{z}_h | \underline{S}_{W_h})$  cannot be provided analytically, but the expression of the un-normalized posterior distribution,  $q(\underline{z}_h | \underline{S}_{W_h})$ , which is directly proportional to  $p(\underline{z}_h | \underline{S}_{W_h})$ , is sufficient for the probabilistic estimation of the parameters.

We can calculate the un-normalized posterior distribution  $q(\underline{z}_h | \underline{S}_{W_h})$  of the random parameters by:

$$q(\underline{z}_h | \underline{S}_{W_h}) = p(\underline{S}_{W_h} | \underline{z}_h) \prod_{j=1}^4 p(z_j) \tag{12}$$

where  $p(\underline{S}_{W_h} | \underline{z}_h)$  is the likelihood function; and  $p(z_j)$  is the prior distribution of the  $j$ th prior random parameter of the vector  $\underline{z}_h$ . The likelihood function  $p(\underline{S}_{W_h} | \underline{z}_h)$  in Equation (12) is given by:

$$p(\underline{S}_{W_h} | \underline{z}_h) = \prod_{s=p+d+k}^M f_{1_{w_s}} [w_s, \eta_{1_s}(\mu_{1_{w_s}}) | \underline{z}_h] \tag{13}$$

where  $\eta_{1_s}(\mu_{1_{w_s}})$  is the Equation (5) evaluated in correspondence to parameters  $\underline{z}_h$  and  $\mu_{1_{w_s}}$ , where  $\mu_{1_{w_s}}$  is the minimum mean square error forecast [26] drawn from the selected ARIMA  $(p, d, q)$  model for the time horizon  $t = s$ , given the past  $p + d$  values of wind speed  $\{w_{s-p-d-k+1}, \dots, w_{s-k}\}$  contained in the set  $\underline{S}_{W_h}$ .

The explicit expression of the likelihood function can be provided in the following form:

$$p(\underline{S}_{W_h} | \underline{z}_h) = \prod_{s=p+d+k}^M \omega_h \left[ \frac{\beta_{1_h}}{\eta_{1_s}} \left( \frac{w_s}{\eta_{1_s}} \right)^{\beta_{1_h}-1} e^{-\left( \frac{w_s}{\eta_{1_s}} \right)^{\beta_{1_h}}} \right] + (1 - \omega_h) \left[ \frac{\beta_{2_h}}{\eta_{2_h}} \left( \frac{w_s}{\eta_{2_h}} \right)^{\beta_{2_h}-1} e^{-\left( \frac{w_s}{\eta_{2_h}} \right)^{\beta_{2_h}}} \right] \tag{14}$$

Numerical values of Equation (12) can be obtained through different methods that have been extensively used in Bayesian relevant literature [22,47,49,50]. In this paper, the Monte Carlo Markov Chain simulation method based on the Metropolis-Hasting algorithm was used to obtain samples of the posterior distributions of the parameters in  $\underline{z}_h$  from the evaluation of the un-normalized posterior distribution  $q(\underline{z}_h | \underline{S}_{W_h})$  [22,47,49,50]. Moreover, the size  $M$  of the historical data can be selected with adequate criteria, thus improving the accuracy of the forecasting method.

### 2.5. Evaluation of the Samples of the PDF $f_{P_{W_h}}$

The samples of the posterior distributions of the parameters  $\omega_h, \beta_{1_h}, \eta_{2_h}, \beta_{2_h}$  for the time horizon  $h$ , obtained as shown in Section 2.4, and the mean value  $\mu_{1_{w_h}}$  of Section 2.3, can be used together to obtain samples of the parameter  $\eta_{1_h}$  from Equation (5); then, the samples of wind speed,  $w_h$ , can be acquired from the estimated distribution  $f_{1_{w_h}}(\omega_h, \eta_{1_h}, \beta_{1_h}, \eta_{2_h}, \beta_{2_h})$  (4). In this paper, these samples were acquired by applying the random rejection sampling algorithm by Von Neumann [51].

Then, the samples of  $w_h$  can be used in a Monte Carlo procedure to obtain samples of  $P_{W_h}$  from Equation (1) and to provide a probabilistic estimation of the wind active power for the time horizon  $h$ .

## 3. Experimental Section

The procedure presented in Section 2 was firstly used on an USA. wind speed time-series (TS1) to forecast the hourly active power produced by a  $P_{max} = 75$  kW wind generator for the time horizon  $h$  with a lead time  $k = 1$  h, acquiring the hourly measurements until the origin time  $h - 1$ . The wind generator was characterized by the wind speed values of  $W_{ci} = 2.3$  m/s,  $W_r = 9$  m/s, and  $W_{co} = 16$  m/s. The non-linear part of the function  $g(w_h)$  in Equation (1) was approximated numerically through a sixth-order polynomial, interpolating the data provided by the manufacturer.

Then, to further validate the proposed approach, the results of forecasts in case of three additional time-series related to two USA sites (TS2 and TS3) and an Italian site (TS4) are reported in Section 3.3.

### 3.1. Characteristics of the Data

The dataset of available measurements consisted of 8760 hourly observations of wind speed obtained from 1 January 2013 to 31 December 2013 by the NWTC M2 Tower at latitude 39°54' north and longitude 105°14' west (USA). Table 2 provides the monthly and annual mean values of wind speed observed during the entire year.

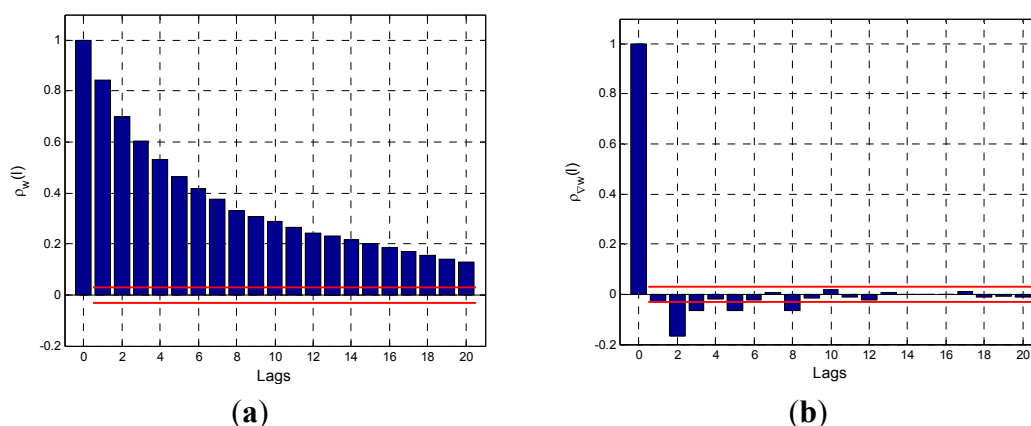
**Table 2.** Monthly and annual mean values of observed wind speed.

Month	Mean Value of Wind Speed (m/s)	Month	Mean Value of Wind Speed (m/s)	Month	Mean Value of Wind Speed (m/s)
January	4.85	May	2.89	September	7.66
February	3.47	June	7.06	October	2.39
March	2.65	July	3.80	November	1.98
April	3.83	August	3.00	December	3.14
Yearly mean value of wind speed (m/s)				4.56	

The procedure for identifying the ARIMA model and estimating the corresponding parameters, described in Section 2.3, was applied to the first 4380 measurements of wind speed, taken during the first six months of the year; then, wind speeds were predicted for the second half of the year.

Figure 1 shows the autocorrelation coefficients and the corresponding significance levels for the original time-series (Figure 1a) and for the once-differenced time-series (Figure 1b) of the observations of wind speed.

Figure 1a shows that the sample autocorrelation function decreases quickly for increasing values of lag,  $l$ ; this suggests that the original time-series was stationary and non-seasonal. The behavior of the autocorrelation function for the once-differenced time-series shown in Figure 1b supports this hypothesis.



**Figure 1.** Autocorrelation coefficients of observations of wind speed and corresponding significance levels: (a) for the original time-series; (b) for the differenced time-series.

Therefore, for the ARIMA model the order  $d = 0$  was selected. Furthermore, the decay of the autocorrelation function is pseudo-exponential, so the ARIMA model (1, 0, 0) and the ARIMA model (2, 0, 0) were used to evaluate the mean value of the distribution of wind speed; using these models in the form of difference Equation (9), we have Equations (15) and (16), respectively:

$$w_t = \theta_0 + \varphi_1 w_{t-1} + e_t \quad (15)$$

$$w_t = \theta_0 + \varphi_1 w_{t-1} + \varphi_2 w_{t-2} + e_t \quad (16)$$

Forecasts were made by using the MWB Bayesian method presented in Sections 2.4 and 2.5, in which the ARIMA (1, 0, 0) model (B-MWB-1 case) and the ARIMA (2, 0, 0) model (B-MWB-2 case) were used to individuate the mean value  $\mu_{1w_h}$  of the distribution of wind speed.

### 3.2. Evaluation of the Quality of the Forecasts

The quality of the probabilistic forecast methods was quantified in this Section by using either a single (spot) value (*i.e.*, the mean value) or the entire probability function. In particular, we define the “spot-value framework” as the case in which the quality of the probabilistic forecast is quantified using a single value of the random variable wind power production, whereas we define “distribution framework” as the case in which the quality of the probabilistic forecast is quantified using a score that uses the entire probability function.

Then, considering the spot-value framework, the mean absolute error (MAE), the root mean square error (RMSE), and their corresponding normalized indices, *i.e.*, NMAE and NRMSE, respectively were selected to quantify the quality of the forecast. These values were calculated as shown below:

$$\begin{aligned} MAE &= \frac{1}{H} \sum_{h=1}^H |P_{W_h}^{spot} - P_h^*| \\ RMSE &= \sqrt{\frac{1}{H} \sum_{h=1}^H (P_{W_h}^{spot} - P_h^*)^2} \\ NMAE &= \frac{1}{H} \sum_{h=1}^H \frac{|P_{W_h}^{spot} - P_h^*|}{P_{max}} 100 \\ NRMSE &= \sqrt{\frac{1}{H} \sum_{h=1}^H \frac{(P_{W_h}^{spot} - P_h^*)^2}{P_{max}^2}} 100 \end{aligned} \quad (17)$$

where  $P_{W_h}^{spot}$  is the spot-forecasted value (*i.e.*, the mean value);  $P_h^*$  is the observed hourly value of active power produced; and  $H$  is the total number of forecasting hours.

With reference to the distribution framework, it was assumed that the continuous ranked probability score (CRPS) and its corresponding normalized index, *i.e.*, NCRPS (probabilistic indices), can be used to quantify the quality of the forecast, and they were calculated as follows:

$$\begin{aligned} CRPS &= \frac{1}{H} \sum_{h=1}^H \int_{-\infty}^{+\infty} [\hat{F}_h(P) - \Theta(P - P_h^*)]^2 dP \\ NCRPS &= \frac{1}{H} \sum_{h=1}^H \frac{\int_{-\infty}^{+\infty} [\hat{F}_h(P) - \Theta(P - P_h^*)]^2 dP}{P_{max}} 100 \end{aligned} \quad (18)$$

where  $\Theta$  is the *Heaviside function*, and  $\hat{F}_h$  is cumulative density function (CDF) of the entire forecasted power. From the analysis of relationships Equation (18) it clearly appears that the CRPS is linked to the total area between the CDF of the forecast and the Heaviside function along all of the hours that were considered. The calculation of the CRPS will result in a value that has the units of the forecast variable, and, therefore, CRPS can be interpreted as a probabilistic version of the MAE [52,53].

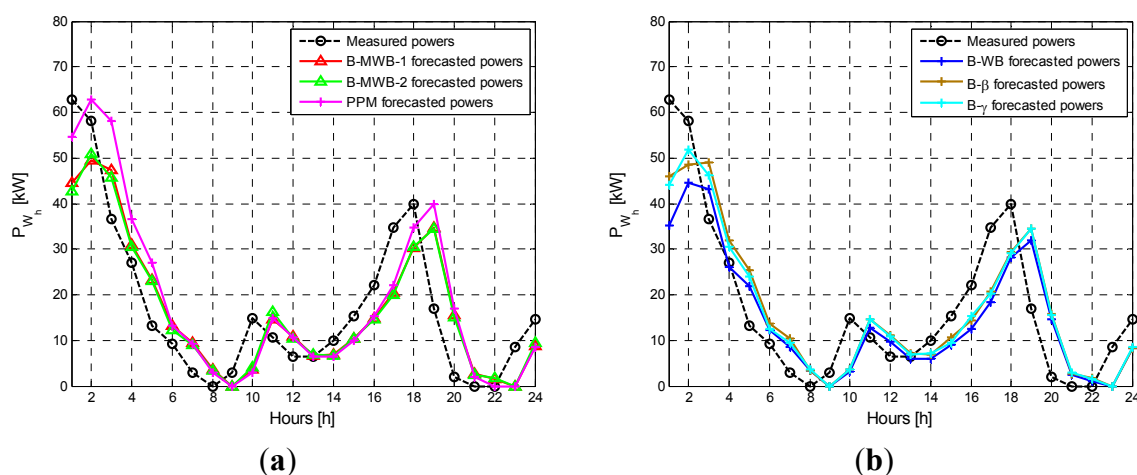
Furthermore, an evaluation of the reliability of the forecasting methods also could be obtained by following the well-known reliability diagrams approach shown in [13,14].

The values of indices Equations (17) and (18), obtained by using the Bayesian method proposed in Section 2, were compared with the values obtained using a reference method, *i.e.*, in the spot-value framework, the well-known persistence method was used as in references [54,55], and, in the distribution framework, the selected benchmark was the probabilistic extension of the persistence method (PPM) proposed in [56,57]. The probability function used in the frame of the PPM was a classical, two-parameter Weibull distribution, the mean value ( $\mu_{0_{w_h}}$ ) of which was assumed to be the observed hourly value of active power at the origin time  $h - 1$ . The shape parameter  $\beta_{0_h}$  was calculated through an iteration of the variance of samples, coherently to the model that was used for the iteration of the mean value, as shown in [56,57].

Further comparisons of the results obtained with the proposed method were provided by using the Bayesian method (B-WB case) proposed in [23], which was based on the use of a two-parameter Weibull distribution, and the Bayesian method which was based on the use of a Beta distribution (B- $\beta$  case) and a Gamma distribution (B- $\gamma$  case).

Forecasts were made for different time intervals and for several days during the period from 14 June 2013 to 31 December 2013. For sake of conciseness, only the results of the forecasts made for 90 days (2160 h), from 14 June 2013 to 12 August 2013, are reported and discussed here to validate the proposed Bayesian method.

Figure 2 shows the results of the one-hour-ahead forecasts as an example of the obtained results. Figure 2a specifically shows the results obtained using the proposed Bayesian-based approaches (B-MWB-1 and B-MWB-2) and using the PPM, and Figure 2b shows the results obtained using the Bayesian methods with unimodal distributions (B-WB, B- $\beta$  and B- $\gamma$ ); the actual hourly wind power is reported in both figures for comparison. For all of the Bayesian forecasts, we assumed that the mean value is the spot forecast value.

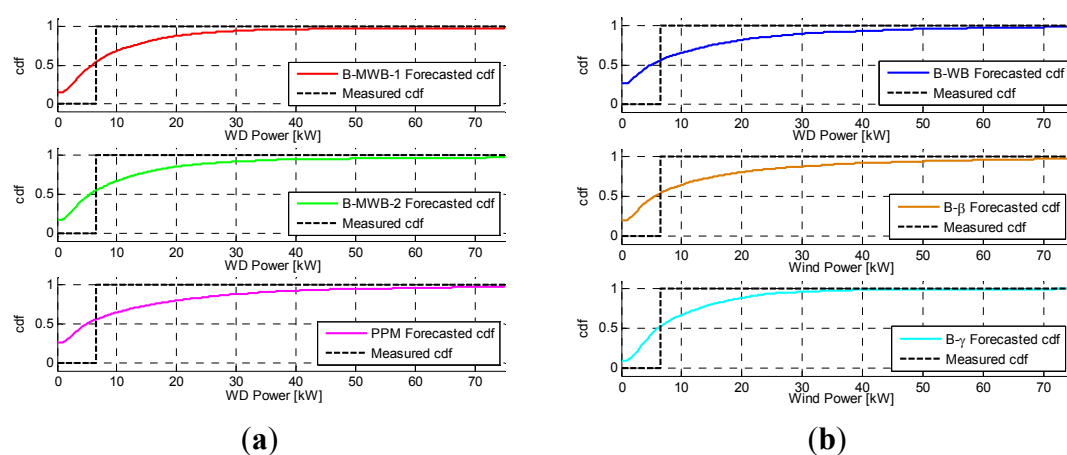


**Figure 2.** Spot value framework compared to the actual hourly wind power (dotted lines) (04/07/2013): (a) Proposed methods: B-MWB-1 and B-MWB-2 cases, and PPM; (b) Bayesian methods with unimodal distributions: B-WB, B- $\beta$  and B- $\gamma$  cases.

Figure 2 indicates that:

- the forecasts of all of the Bayesian methods were closer to the observed values of wind active power than the forecasts of PPM for almost all hours of the day. This is an interesting result, since the persistence method usually has good behavior for low values of lead time,  $k$ ;
- all of the Bayesian methods had very similar results due the similar approaches that were used to evaluate the spot value by the ARIMA models. Thus, a more comprehensive analysis was required in terms of sharpness and reliability of the forecasting methods, and it had to consider the entire forecasted PDF in the distribution framework.

Figure 3 shows the CDFs of forecasts and the CDF of a one-hour observation during the day (*i.e.*, 1:00 PM). Figure 3a compares the forecasted CDFs obtained through the proposed method (B-MWB-1 and B-MWB-2 cases) and through PPM to the CDF of the actual value of wind power. Figure 3b compares the forecasted CDFs obtained through the Bayesian methods with unimodal distributions (B-WB, B- $\beta$  and B- $\gamma$ ) to the CDF of the actual value of wind power. Figure 2 shows that, at 1:00 PM, the deviations of the forecasted spot values from the actual value of wind power were very similar for all the Bayesian forecasting methods that were considered; however, this behavior was not the same in terms of hourly contribution to the CRPS. In fact, when Figure 3a,b were compared, it was evident that the total areas between the measured CDFs and the forecasted CDF obtained by using the proposed method were smaller than the total areas obtained by using the other methods. In this case, both underproduction error area (where the measured CDF is greater than the forecasted CDF) and overproduction error area (where the measured CDF is smaller than the forecasted CDF) for the reference methods were greater than the corresponding areas for the proposed method. The best performances were obtained by the proposed methods; B- $\gamma$  performed slightly worse, leading to a greater underproduction area, while PPM, B-WB and B- $\beta$  led to greater underproduction and overproduction areas. There were no appreciable differences between the results obtained in the B-MWB-1 and B-MWB-2 cases.



**Figure 3.** Forecasted CDFs compared to the CDFs of the actual value of wind power (at 1:00 PM on 04/07/2013); (a) Proposed methods: B-MWB-1 and B-MWB-2 cases, and PPM; (b) Bayesian methods with unimodal distributions: B-WB, B- $\beta$  and B- $\gamma$  cases.

As a further example of the results that were obtained, Table 3 shows the values of the absolute and percentage error indices Equations (17) and (18) obtained for all of the methods and averaged over the time interval from 14 June 2013 to 12 August 2013.

**Table 3.** Time-series TS1: Values of absolute and percentage indices for the forecasts averaged from 14 June 2013 to 12 August 2013.

Index	Forecasting Method					
	Proposed Method		Reference Method			
	B-MWB-1	B-MWB-2	PPM	B-WB	B- $\beta$	B- $\gamma$
MAE (kW)	7.47	7.41	7.74	7.49	7.50	7.50
RMSE (kW)	13.67	13.58	14.39	13.81	13.79	13.81
CRPS (kW)	5.51	5.50	6.17	5.74	5.59	5.64
NMAE (%)	9.96	9.88	10.31	9.98	9.99	10.00
NRMSE (%)	18.23	18.10	19.18	18.41	18.39	18.41
NCRPS (%)	7.35	7.33	8.22	7.65	7.45	7.52

The values in Table 3 indicate that:

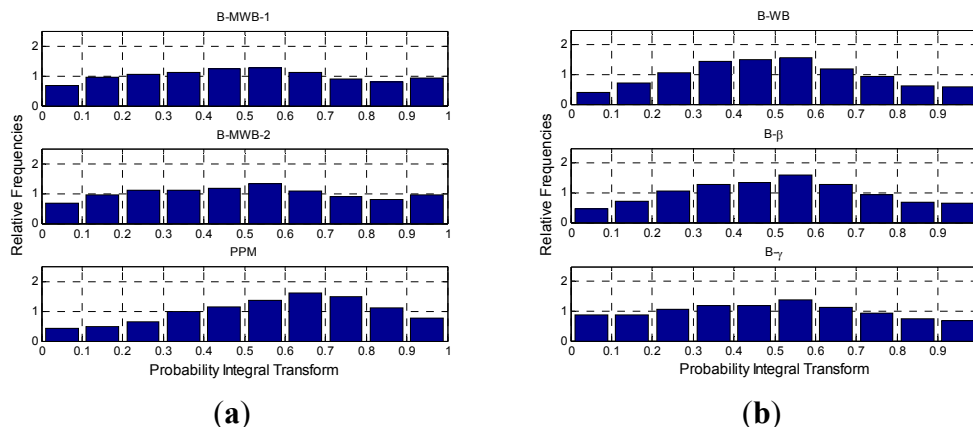
- Bayesian methods performed better than PPM in the spot-value framework; in fact, the MAE index decreased by about 5%. In addition, the proposed Bayesian method provided results that were slightly better than the B-WB, B- $\beta$  and B- $\gamma$  methods; for example, the RMSE calculated with B-MWB-2 was 2% lower than the RMSE calculated with all of the Bayesian methods with unimodal distributions;
- in the distribution framework, all of the Bayesian methods performed better than PPM; in fact, they provided CRPS values that were about 11.5% even lower than the values obtained by PPM. The proposed method had the best performances; for example, the CRPS values obtained by using either B-MWB-1 or B-MWB-2 were 4% lower than the values obtained by using B-WB, 2.5% lower than the values obtained with B- $\gamma$  and 2% lower than the values obtained with B- $\beta$ ;
- a comparison of B-MWB-1 and B-MWB-2 indicates that the latter provided slightly better performance, since the fact that all of its indices had relatively smaller values.

The CRPS is a consolidated tool to evaluate the calibration and the sharpness of forecasts. However, for sake of completeness, the calibration of forecasts can be further checked through the inspection of PIT histograms [58–60]. The PIT histogram is a visual, informal diagnostic tool; deviations from uniformity usually evidence forecast failures and model deficiencies. U-shaped histograms indicate under-dispersed predictive distributions, inverse-U shaped histograms indicate over-dispersed predictive distributions, and skewed histograms usually evidence biased forecasts.

Figure 4 shows the PIT histograms obtained through the proposed methods B-MWB-1 and B-MWB-2 and PPM (Figure 4a) and the PIT histograms obtained through the Bayesian benchmark methods (B-WB, B- $\beta$  and B- $\gamma$  in Figure 4b).

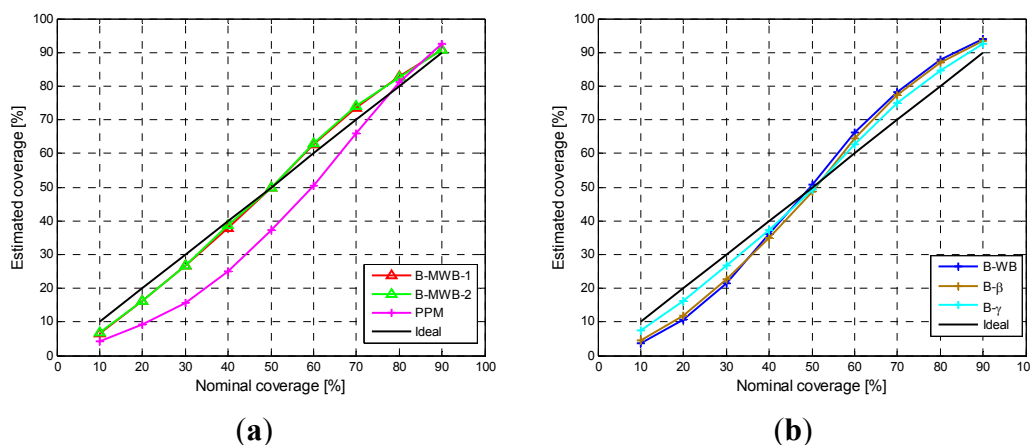
From the graphical inspection of Figure 4, the behavior of PIT histograms seems to be coherent to the corresponding values of CRPS in Table 3; the proposed method provides almost uniform histograms, coupled with the lowest values of CRPS. B- $\gamma$  also provides an almost uniform histogram, even if the

corresponding value of CRPS is higher, while PPM, B-WB and B- $\beta$  show a typical inverse-U shaped histogram, suggesting an over-dispersion in the forecasted distributions.



**Figure 4.** PIT histograms for: (a) Proposed methods: B-MWB-1 and B-MWB-2 cases, and PPM; (b) Bayesian methods with unimodal distributions: B-WB, B- $\beta$  and B- $\gamma$  cases.

The reliability of the proposed methods was evaluated by comparing the empirical coverages of the various quantiles of the forecasted PDFs to the nominal coverages for the entire interval of forecasts. Figure 5 shows the estimated coverage with respect to the nominal coverage obtained by the four methods that were considered. Figure 5 shows that the forecasts produced by either B-MWB-1 or B-MWB-2 provided the best reliabilities, with their reliability diagrams virtually overlapping each other. B- $\gamma$  performed slightly worse than the proposed method, especially in correspondence to the higher quantiles. B-WB and B- $\beta$  deviated more than the proposed methods from the ideal reliability curve, and the reliability of PPM was particularly poor.



**Figure 5.** Reliability diagrams for the proposed methods and the reference methods compared to the ideal reliability; (a) Proposed methods: B-MWB-1 and B-MWB-2 cases, and PPM; and (b) Bayesian methods with unimodal distributions: B-WB, B- $\beta$  and B- $\gamma$  cases.

### 3.3. Further Analysis

The proposed forecasting method was validated using several additional wind time-series. For sake of conciseness, in this section the results of forecasts performed for two U.S. locations (TS2 at latitude



34°59' and longitude 104°02'; TS3 at latitude 45°20' and longitude 104°25', respectively) and for an Italian location (TS4, at latitude 16°21' and longitude 40°57') are reported. The selected locations were chosen in order to test the proposed method in different conditions of wind occurrences. In the following, the results of a 1-month time interval forecasts are shown. Table 4 shows the yearly mean values of wind speed, the ARIMA model chosen through the procedure described in Section 2.3, and the characteristics of the corresponding wind turbines for the three locations.

**Table 4.** Values of yearly mean wind speed and characteristics of the wind turbines for TS2, TS3, and TS4.

Location	Yearly Mean Value of Wind Speed (m/s)	ARIMA Model	Wind Turbine Rated Power $P_{max}$ (kW)	Cut-in Wind Speed $W_{ci}$ (m/s)	Rated Wind Speed $W_r$ (m/s)	Cut-off Wind Speed $W_{co}$ (m/s)
TS2	9.98	(2,0,0)	1500	4	15	25
TS3	9.61	(2,0,0)	670	3	13	25
TS4	6.82	(4,0,0)	670	3	13	25

The quality of the forecasts was quantified by using the indices defined in Section 3.2; the values of indices are shown in Tables 5–7 for TS2, TS3, and TS4, respectively, and compared with the values obtained through the same benchmark methods used in Section 3.2.

**Table 5.** Time-series TS2: Values of absolute and percentage indices for the forecasts averaged from 1 October 2013 to 31 October 2013.

Index	Forecasting Method				
	Proposed Method	Reference Methods			
		B-MWB-2	PPM	B-WB	B- $\beta$
MAE (kW)	131.40	134.53	137.00	136.34	136.76
RMSE (kW)	232.65	243.52	241.36	240.39	240.54
CRPS (kW)	104.91	114.38	113.76	113.00	107.57
NMAE (%)	8.76	8.97	9.13	9.09	9.12
NRMSE (%)	15.51	16.23	16.09	16.03	16.04
NCRPS (%)	6.99	7.63	7.58	7.53	7.17

**Table 6.** Time-series TS3: Values of absolute and percentage indices for the forecasts averaged from 1 October 2013 to 31 October 2013.

Index	Forecasting Method				
	Proposed Method	Reference Methods			
		B-MWB-2	PPM	B-MWB-2	B- $\beta$
MAE (kW)	64.20	70.03	70.48	70.11	70.12
RMSE (kW)	100.07	109.94	108.15	107.78	107.81
CRPS (kW)	53.03	54.80	61.39	63.38	59.93
NMAE (%)	9.58	10.45	10.52	10.46	10.47
NRMSE (%)	14.94	16.41	16.14	16.09	16.09
NCRPS (%)	7.92	8.18	9.16	9.46	8.94

**Table 7.** Time-series TS4: Values of absolute and percentage indices for the forecasts averaged from 1 October 2013 to 31 October 2013.

Index	Forecasting Method				
	Proposed Method	Reference Method			
		B-MWB-4	PPM	B-WB	B- $\beta$
MAE (kW)	62.92	62.54	63.10	62.95	63.30
RMSE (kW)	111.41	114.54	113.00	112.70	112.97
CRPS (kW)	51.39	52.21	57.60	57.14	57.20
NMAE (%)	9.39	9.33	9.42	9.40	9.45
NRMSE (%)	16.63	17.10	16.86	16.82	16.86
NCRPS (%)	7.67	7.79	8.60	8.53	8.54

The values in Tables 3, 5–7 indicate that:

- the proposed method seems to be the best forecasting tool, since it provides the lowest values in all of the indices, with the only exception of MAE for TS4; in particular, it seems to be very suitable to provide sharp probabilistic forecasts, since it provides the lowest values of CRPS index;
- comparing the proposed method to the PPM benchmark, the values of MAE decrease by about 2% and 8% for TS2 and TS3, respectively, while for TS4 the value of MAE provided by the proposed method is slightly (0.5%) higher than PPM benchmark. The dispersion of forecasts around the mean value is however smaller for the proposed method, since it provides lower values of RMSE (about 5%, 9%, and 3% lower for TS2, TS2, and TS4, respectively) and CRPS (about 8%, 3% and 2% lower for TS2, TS2, and TS4, respectively);
- among the Bayesian benchmarks, B- $\gamma$  seems to provide the most accurate forecasts; however, the proposed method outperforms B- $\gamma$  in all of the considered time-series. In fact, the values of MAE decrease by about 4%, 9%, and less than 1% using the proposed method; the values of RMSE decrease by about 3%, 7%, and less than 2%; the values of CRPS decrease by about 3%, 10% and 11% for TS2, TS3, and TS4, respectively;
- the global behavior of the proposed method seem to be coherent for all of the considered time-series, even if the application is significantly different in terms of rated power. This behavior can be detected by comparing the normalized values of the indices provided by the proposed method; they are very similar for all of the four considered time-series, even if the rated power of the considered turbines vary significantly (from 75 to 1500 kW).

#### 4. Conclusions

In this paper, we proposed a new probabilistic method for forecasting the generation of wind power. This method was based on the Bayesian theory and was particularly appropriate for forecasting wind power in correspondence with wind speed regimes that vary significantly over time; this result was achieved by using a more complex probability distribution to characterize wind speed, *i.e.*, the mixture Weibull distribution, which seemed the most suitable to represent both unimodal and bimodal wind regimes.

The results obtained with the proposed method were compared with the results obtained by using two probabilistic forecasting approaches that have been used extensively, *i.e.*, the persistence method and a traditional Bayesian method using two-parameter distributions.

The numerical applications were performed with respect to various wind turbines; the proposed method proved to be useful in short-term probabilistic forecasting of wind power, performing better than the reference methods in terms of both point-value and distribution forecasts. In particular, the proposed method offered significant improvements in terms of the sharpness and reliability of forecasts.

We concluded that the proposed Bayesian method is the most suitable for representing both unimodal and bimodal wind regimes. Also, additional significant improvements of the forecasts are expected in particular regimes that are characterized by a significant number of days of bimodal wind speed distributions.

Future work will focus on the application of the Box-Jenkins approach, based on the use of the sample autocorrelation function, to forecast the photovoltaic power generation. In addition, since in this paper we assumed a deterministic power curve, studies on the probabilistic behavior of power curve depending on meteorological variables such as wind direction, temperature, local air density, and precipitation will be carried out.

### Author Contributions

Authors contributed in the same way all over the paper.

### Conflicts of Interest

The authors declare no conflict of interest.

### References

1. Fang, X.; Misra, S.; Xue, G.; Yang, D. Smart grid—The new and improved power grid: A survey. *IEEE Commun. Surv. Tutor.* **2012**, *14*, 944–980.
2. Farhangi, H. The path of the smart grid. *IEEE Power Energy Mag.* **2010**, *8*, 18–28.
3. Blumsack, S.; Fernandez, A. Ready or not, here comes the smart grid! *Energy* **2012**, *37*, 61–68.
4. Lo, C.H.; Ansari, N. The progressive smart grid system from both power and communications aspects. *IEEE Commun. Surv. Tutor.* **2012**, *14*, 799–821.
5. Zhou, L.; Rodrigues, J.J.P.C. Service-oriented middleware for smart grid: Principle, infrastructure, and application. *IEEE Commun. Mag.* **2013**, *51*, 84–89.
6. Zhou, L.; Rodrigues, J.J.P.C.; Oliveira, L.M. QoE-driven power scheduling in smart grid: Architecture, strategy, and methodology. *IEEE Commun. Mag.* **2012**, *50*, 136–141.
7. Santacana, E.; Rackliffe, G.; Tang, L.; Feng, X. Getting smart. *IEEE Power Energy Mag.* **2010**, *8*, 41–48.
8. CIGRE—International Council on Large Electrical Systems. *Impact of Increasing Contribution of Dispersed Generation on the Power System*; CIGRE WG 37-23; CIGRE: Paris, France, 1999.
9. Smith, J.C. Winds of change. *IEEE Power Energy* **2005**, *3*, 20–24.
10. Bessa, R.J.; Miranda, V.; Botterud, A.; Wang, J. Good or bad wind power forecasts: A relative concept. *Wind Energy* **2011**, *14*, 625–636.

11. Potter, C.W.; Archambault, A.; Westrick, K. Building A Smarter Smart Grid through Better Renewable Energy Information. In Proceedings of the IEEE/PES Power Systems Conference and Exposition, Seattle, WA, USA, 15–18 March 2009.
12. Costa, A.; Crespo, A.; Navarro, J.; Lizcano, G.; Madsen, H.; Feitosa, E. A review on the young history of the wind power short-term prediction. *Renew. Sustain. Energy Rev.* **2008**, *12*, 1725–1744.
13. Pinson, P.; Nielsen, H.; Møller, J.K.; Madsen, H.; Kariniotakis, G.N. Non-parametric probabilistic forecasts of wind power: Required properties and evaluation. *Wind Energy* **2007**, *10*, 497–516.
14. Pinson, P.; Juban, J.; Kariniotakis, G.N. On The Quality and Value of Probabilistic Forecasts of Wind Generation. In Proceedings of the International Conference on Probabilistic Methods Applied to Power Systems, Stockholm, Sweden, 11–15 June 2006.
15. Kariniotakis, G.N.; Pinson, P.; Siebert, N.; Giebel, G.; Bartelmie, R. State of the art in short-term prediction of wind power—From an offshore perspective. In Proceedings of Sea Tech Week, Brest, France, 20–21 October 2004.
16. Bracale, A.; Caramia, P.; Carpinelli, G.; Di Fazio, A.R.; Ferruzzi, G. A Bayesian method for short-term probabilistic forecasting of photovoltaic generation in smart grid operation and control. *Energies* **2013**, *6*, 733–747.
17. Li, G.; Shi, J. Applications of Bayesian methods in wind energy conversion systems. *Renew. Energy* **2013**, *43*, 1–8.
18. Uusitalo, L. Advantages and challenges of Bayesian networks in environmental modelling. *Ecol. Model.* **2007**, *203*, 3–4.
19. Miranda, M.S.; Dunn, R.W. One-Hour-Ahead Wind Speed Prediction using a Bayesian methodology. In Proceedings of the IEEE Power Engineering Society General Meeting, Montreal, QC, Canada, 18–22 June 2006.
20. Jiang, Y.; Song, Z.; Kusiak, A. Very short-term wind speed forecasting with Bayesian structural break model. *Renew. Energy* **2013**, *50*, 637–647.
21. Kim, C.J.; Kim, J. Bayesian inference in regime-switching arma models with absorbing states: The Dynamics of the ex-ante real interest rate under regime shifts. *J. Bus. Econ. Stat.* **2014**, doi:10.1080/07350015.2014.979995.
22. Gelman, A.; Carlin, J.B.; Stern, H.S.; Rubin, D.B. *Bayesian Data Analysis*; Chapman & Hall: London, UK, 1995.
23. Zhang, J.; Pu, J.; McCalley, J.D.; Stern, H.; Gallus, W.A., Jr. A Bayesian approach for short-term transmission line thermal overload risk assessment. *IEEE Trans. Power Deliv.* **2002**, *17*, 770–778.
24. Bracale, A.; Caramia, P.; Carpinelli, G.; Di Fazio, A.R.; Varilone, P. A Bayesian-based approach for a short-term steady-state forecast of a smart grid. *IEEE Trans. Smart Grid* **2013**, *4*, 1760–1771.
25. Carta, J.A.; Ramirez, P.; Velazquez, S. A review of wind speed probability distributions used in wind energy analysis. Case studies in the Canary islands. *Renew. Sustain. Energy Rev.* **2009**, *13*, 933–955.
26. Box, G.; Jenkins, G. *Time Series Analysis: Forecasting and Control*, 3rd ed.; Prentice-Hall: Upper Saddle River, NJ, USA, 1994.
27. Xie, L.; Gu, Y.; Zhu, X.; Genton, M.G. Short-term spatio-temporal wind power forecast in robust look-ahead power system dispatch. *IEEE Trans. Smart Grid* **2014**, *5*, 511–520.

28. Black, J. Plans for Wind Integration in ISO-NE: Progress and Challenges. In Proceedings of the 8th Annual Carnegie Mellon Conference on Electricity Industry, Pittsburgh, PA, USA, 12–14 March 2012; pp. 1–37.
29. Deep Thunder-Precision Forecasting for Weather-Sensitive Business Operations. 2012. Available online: <http://www.research.ibm.com/weather/DT.html> (accessed on 15 September 2015).
30. Mahoney, W.P.; Parks, K.; Wiener, G.; Liu, Y.; Myers, W.L.; Sun, J.; Delle Monache, L.; Hopson, T.; Johnson, D.; Haupt, S.E. A wind power forecasting system to optimize grid integration. *IEEE Trans. Sustain. Energy* **2012**, *3*, 670–682.
31. Constantinescu, E.M.; Zavala, V.M.; Rocklin, M.; Lee, S.; Anitescu, M. A computational framework for uncertainty quantification and stochastic optimization in unit commitment with wind power generation. *IEEE Trans. Power Syst.* **2011**, *26*, 431–441.
32. Genton, M.; Hering, A. Blowing in the wind. *Significance* **2007**, *4*, 11–14.
33. Giebel, G.; Brownsword, R.; Kariniotakis, G.; Denhard, M.; Draxl, C. *The State-Of-The-Art in Short-Term Prediction of Wind Power: A Literature Overview*, 2nd ed.; ANEMOS.plus Project Deliverable Report D1.2. 2011. Available online: [http://orbit.dtu.dk/fedora/objects/orbit:83397/datastreams/file\\_5277161/content](http://orbit.dtu.dk/fedora/objects/orbit:83397/datastreams/file_5277161/content) (accessed on 15 September 2015).
34. Karaki, S.H.; Chedid, R.B.; Ramadan, R. probabilistic performance assessment of wind energy conversion systems. *IEEE Trans. Energy Convers.* **1999**, *14*, 217–224.
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.
36. Hering, A.S.; Genton, M. Powering up with space-time wind forecasting. *J. Am. Stat. Assoc.* **2009**, *105*, 92–104.
37. Gneiting, T. Quantiles as optimal point forecasts. *Int. J. Forecast.* **2011**, *27*, 197–207.
38. Sánchez, I. Short-term prediction of wind energy production. *Int. J. Forecast.* **2006**, *22*, 43–56.
39. Jeon, J.; Taylor, J.W. Using conditional kernel density estimation for wind power density forecasting. *J. Am. Stat. Assoc.* **2012**, *107*, 66–79.
40. Lange, M. On the uncertainty of wind power predictions—Analysis of the forecast accuracy and statistical distribution of errors. *J. Solar Energy Eng.* **2005**, *127*, 177–184.
41. Korpaas, M.; Holen, A.T.; Hildrum, R. Operation and sizing of energy storage for wind power plants in a market system. *Int. J. Electr. Power Energy Syst.* **2003**, *25*, 599–606.
42. Matevosyan, J.; Söder, L. Minimization of imbalance cost trading wind power on the short-term power market. *IEEE Trans. Power Syst.* **2006**, *21*, 1396–1404.
43. Morales, J.M.; Conejo, A.J.; Pérez-Ruiz, J. Short-term trading for a wind power producer. *IEEE Trans. Power Syst.* **2010**, *25*, 554–564.
44. Chang, T.P.; Liu, F.J.; Ko, H.; Cheng, S.; Sun, L.; Kuo, S. Comparative analysis on power curve models of wind turbine generator in estimating capacity factor. *Energy* **2014**, *73*, 88–95.
45. Titterton, D.M.; Smith, A.F.M.; Makov, U.E. *Statistical Analysis of Finite Mixture Distributions*, 2nd ed.; Wiley: New York, NY, USA, 1995.
46. Kaylan, A.R.; Harris, C.M. Efficient algorithms to derive maximum-likelihood estimates for finite exponential and weibull mixtures. *Comput. Oper. Res.* **1981**, *8*, 97–104.
47. Gamerman, D. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*; Chapman & Hall: London, UK, 1997.

48. Carta, J.A.; Ramirez, P. Analysis of two-component mixture Weibull statistics for estimation of wind speed distributions. *Renew. Energy* **2007**, *32*, 518–531.
49. Martz, H.F.; Waller, R.A. *Bayesian Reliability Analysis*; John Wiley and Sons: New York, NY, USA, 1982.
50. Berger, J.O. *Statistical decision theory and Bayesian analysis*; Springer Science & Business Media: New York, NY, USA, 2013.
51. Von Neumann, J. *Various Techniques Used in Connection With Random Digits*; National Bureau of Standards Applied Mathematics Series; U.S. Government Printing Office: Washington, DC, USA, 1951; Volume 12, pp. 36–38.
52. Gneiting, T.; Raftery, A.E.; Westveld, A.H., III; Goldman, T. calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Mon. Weather Rev.* **2005**, *133*, 1098–1118.
53. Gneiting, T.; Raftery, A.E. Strictly proper scoring rules, prediction, and estimation. *J. Am. Stat. Assoc.* **2007**, *102*, 359–378.
54. Pinson, P.; Chevallier, C.; Kariniotakis, G. Trading wind generation from short-term probabilistic forecasts of wind power. *IEEE Trans. Power Syst.* **2007**, *22*, 1148–1156.
55. Bacher, P.; Madsen, H.; Nielsen, H.A. Online short-term solar power forecasting. *Solar Energy* **2009**, *83*, 1772–1783.
56. Pinson, P.; Reikard, G.; Bidlot, J.R. Probabilistic forecasting of the wave energy flux. *Appl. Energy* **2012**, *93*, 364–370.
57. Pinson, P. Very-short-term probabilistic forecasting of wind power with generalized logit-normal distributions. *J. R. Stat. Soc.* **2012**, *61*, 555–576.
58. Czado, C.; Gneiting, T.; Held, L. Predictive model assessment for count data. *Biometrics* **2009**, *65*, 1254–1261.
59. Diebold, F.X.; Gunther, T.A.; Tay, A.S. Evaluating density forecasts with applications to financial risk management. *Int. Econ. Rev.* **1998**, *39*, 863–883.
60. Gneiting, T.; Balabdaoui, F.; Raftery, A.E. Probabilistic forecasts, calibration and sharpness. *J. R. Stat. Soc.* **2007**, *69*, 243–268.