A Day-Ahead Wind Power Scenario Generation, Reduction, and Quality Test Tool

Ceyhun Yıldız [1], Mustafa Tekin [2], Ahmet Gani [2], Ö. Fatih Keçecioğlu [2], Hakan Açıkgöz [3] and Mustafa Şekkeli [2,*]

[1] Department of Electric and Energy, Elbistan Vocational School, University of K.Maraş Sütçü İmam, K.Maraş 46340, Turkey; cyildiz@ksu.edu.tr
[2] Department of Electrical and Electronics, Faculty of Engineering, University of K.Maraş Sütçü İmam, K.Maraş 46040, Turkey; mustafatekin@ksu.edu.tr (M.T.); agani@ksu.edu.tr (A.G.); fkececioglu@ksu.edu.tr (F.K.)
[3] Department of Electrical Science, Kilis 7 Aralık University, Kilis 79000, Turkey; hakanacikgoz@kilis.edu.tr
* Correspondence: msekkeli@ksu.edu.tr; Tel.: +90-344-300-1633

Abstract: During the last decades, thanks to supportive policies of countries and a decrease in installation costs, total installed capacity of wind power has increased rapidly all around the world. The uncertain and variable nature of wind power has been a problem for transmission system operators and wind power plant owners. To solve this problem, numerous wind power forecast systems have been developed. Unfortunately none of them can obtain absolutely accurate forecasts yet. Thus, researchers assumed that wind power generation is a stochastic process and they proposed a stochastic programming approach to solve problems arising from the uncertainty of wind power. It is well known that representing stochastic process by possible scenarios is a major issue in the stochastic programming approach. Large numbers of scenarios can represent a stochastic process accurately, but it is not easy to solve a stochastic problem that contains a large number of scenarios. For this reason scenario reduction methods have been introduced. Finally, the quality of this reduced scenario set must be at an acceptable level to use them in calculations. All of these reasons have encouraged authors to develop a wind power scenario tool that can generate and reduce the scenario set and test the quality of it. The developed tool uses historical data to model wind forecast errors. Scenarios are generated around 24 day-ahead point wind power forecasts. A fast forward reduction algorithm is used to reduce the scenario set. Two metrics are proposed to assess the quality of the reduced scenario set. Site measurements are used to test the developed wind power scenario tool. Results showed that the tool can generate and reduce the scenario set successfully and the proposed metrics are useful to assess the quality.

Keywords: wind power scenarios; day-ahead wind power; wind forecast

1. Introduction

During recent decades wind energy has been an outstanding renewable energy source. Most countries have developed policies to support wind energy deployment globally [1]. Thanks to these policies and a decrease in the installation costs, total installed capacity of wind power has increased substantially [2]. Wind farms are environmentally friendly power systems and their fuel cost is zero. However wind is an uncontrollable and fluctuating energy source so the power output of a wind farm is variable and uncertain. This uncertainty and variability has been a problem for transmission system operators and wind farm owners. System operators need an accurate generation plan for the next day to maintain and ensure a load-generation balance. On the other hand, in

competitive day-ahead markets, farm owners must, in each hour, generate exactly the amount of energy they promised to deliver much earlier. To solve these problems, researchers have developed wind power forecast systems [3]. Unfortunately none of these forecast systems can generate accurate forecasts yet. Thus, in the literature, many studies assumed that maintaining the energy balance of the wind-integrated grid and wind energy trading in day-ahead markets are stochastic problems [4–11]. To solve these problems, having a scenario set that can represent stochastic processes successfully is needed. Afterwards, a stochastic programming framework can be used to achieve optimal solutions.

In the literature, researchers proposed methods to generate short-term wind scenarios. It has been seen that several studies [11–18] have used wind speed to generate wind power scenarios. These methods are not so suitable to generate wind power scenarios because it is well known that the wind to power conversion adds substantial errors to the results [19]. Additionally, considerable amounts of studies [12–18] have used an autoregressive moving average model (ARMA) to generate wind power scenarios. The ARMA approach is actually for modeling stationary processes. However, wind power forecast errors are not stationary, different weather conditions of seasons affect uncertainty, and it is not easy to reject these complex effects from forecast error data. Thus, the ARMA approach is not so suitable to model the uncertainty of wind power. Few studies in the literature [20,21] have considered the seasonal variation in the uncertainty of forecast errors. Quevedo and Contreras [20] used seasonal cumulative distribution functions (CDF) of historical data to generate wind speed scenarios. The method proposed in [21] is based on converting probabilistic forecast series (obtained by quantile regression) into multivariate Gaussian random variables. The covariance matrix is used to represent the interdependence of these variables and this matrix is recursively updated to consider non-stationary characteristics of forecast errors. Other contribution on modeling forecast errors is assessing error distributions within forecast power levels (bins). The method based on this idea is proposed in [22] and adopted in [23]. Ma et al. [23] used the empirical cumulative distribution function (ECDF) approach to model error distributions within the bins. Thanks to the non-parametric and universal characteristic of ECDF, it is possible to model any distribution successfully.

Large scenario sets will achieve all probable realizations. However it is not easy to solve a stochastic problem that contains large scenario sets. Therefore, scenario reduction algorithms have been proposed to minimize the number of scenarios [24–27]. On the other hand, the quality of the reduced scenario set must be at an acceptable level to obtain optimal solutions. For this reason scenario quality tests are essential. After all, scenario generation, reduction, and quality test processes have been the major issues of these stochastic problems.

The aim of this study is to develop three useful methods that operate together to generate a scenario set for stochastic short-term wind power generation. The first method creates a scenario set that involves almost all possible realizations. This new method combines two powerful approaches; a non-parametric wind power forecast error modeling approach of [23], and a recursive covariance matrix updating formula of [21] that injects seasonal effects into the model. Instead of wind speed errors, wind power errors were used to increase the accuracy of the model. Additionally, forecasts are separated into power bins to add the uncertainty change by power level. The second method reduces the number of these scenarios. This method is taken from the literature [28]. The last one evaluates the quality of the reduced scenario set. For this purpose two simple metrics are used. One can generate and test scenarios through these methods. Furthermore, a detailed stepwise explanation of the introduced algorithms and some useful MATLAB (2012a, Mathworks Inc., Natick, MA, USA) functions are given to make the reutilization process easier.

The contributions of this paper are two-fold:

(1) to introduce a novel day-ahead wind power scenario tool that uses scenario generation, reduction, and scenario quality testing methods together; and

(2) to propose a new day ahead wind power scenario generation method that combines two powerful approaches.

The rest of this paper is structured as follows: Section 2 describes the theoretical background of the introduced methods and the power plant upon which experimental studies were done. Additionally, a detailed explanation of application algorithms and results are given in this section. Section 3 describes the details of the experimental studies and results. Section 4 provides the discussion and conclusion of this study.

2. Scenario Generation, Reduction, and Quality Tests

Researchers have proposed a stochastic programming approach to solve problems that are encountered in wind-integrated grids. This approach requires scenario sets that can represent stochastic wind power generation successfully. In this study we introduced day-ahead wind power scenario generation, reduction, and quality testing methods. The theoretical background of these methods and the description of the study site are given in this section. The next subsection describes the forecast system and the wind farm which was investigated in this work.

2.1. Description of Study Site and Data

In this study wind scenarios are generated for an on-shore wind farm, located in the Hatay province of Turkey. Some technical details of the power plant are summarized in Table 1 and a general view of the plant is given in Figure 1.

Table 1. Technical details of the Belen Wind Farm.

Name of Wind Farm	Belen
Turbine Model	VESTAS V90
Hub Height (m)	80
Rotor Diameter (m)	90
Wind Class	IA
Turbine Output Power (MW)	3
Number of Turbines	16
Total Installed Power (MW)	48

Figure 1. General view of the Belen Wind Farm.

Wind power and forecast data are taken from the national wind power monitoring and forecasting system [29]. This system generates point forecasts for all wind farms in Turkey. It uses outputs of numerical weather prediction (NWP) models (wind speed and direction forecasts) as wind farm model inputs. Computational fluid dynamics based software (WindSim) is used to create the wind farm model. Additionally, a clustering mechanism is used to select the best NWP outputs geographically close to the wind farm. Performance testing of the forecast system showed that forecasts have an acceptable accuracy.

The dataset used in this study comprises wind power forecasts and power measurements for a two year period (1 January 2014–31 December 2015). The power analyzer that was installed at the

powerhouse of the wind farm is used to measure the actual generation. All data is divided by the installed power of the wind farm to obtain per-unit (pu) values. The sampling time of the data is one hour. Measured and forecast data are used to model the forecast errors. This model is used to generate scenarios around 24 day-ahead point forecasts. The modeling process is described in the next section.

2.2. Modeling Forecast Errors

In this work probability distribution functions are used to model forecast errors. Ma et al. [23] introduced forecast error modelling method based on historical power measurements and forecasts. This method models historical wind power measurements within the specified forecast power levels (bins). It is well known that the distribution of wind power within a forecast power bin does not follow any theoretical distribution. The ECDF approach is proposed in [23] to model wind power distributions. On the other hand, forecast error distribution is not as complicated as power distribution. Much of the literature makes the simplifying assumption that forecast errors follow a normal distribution [30–32], while actual studies [33,34] show that forecast errors do not follow a normal distribution. Tewari et al. [32] concluded that the forecasting approach and site effects will change the shape of the distribution. Since modelling the error distribution is easier and more effective than modelling the power distribution, we proposed to model the forecast errors to generate scenarios around point forecasts. Firstly, installed power values are separated into equidistant 20 power levels (bins) as in [22]. The bin width is 0.05 pu. Point forecasts are sorted into these bins by considering the power levels. Each point forecast has a corresponding error that is assigned to the same bin. The twenty bins have 20 error distributions. ECDF is used to model the error distributions. Thanks to the universal and nonparametric approach of ECDF, it is possible to model various distributions. The mathematical background of ECDF is described in [34].

To see the results of the abovementioned method, error distributions within bins 6, 10, and 14 are obtained. Bin 6 comprises historical point forecasts between 0.25 pu and 0.30 pu. Bins 10 and 14 comprise point forecasts occurring in intervals 0.45–0.50 pu and 0.65–0.70 pu, respectively. The error distributions of these sorted point forecasts are shown in the upper panel of Figure 2. It is not clear which common family of theoretical distributions bins 6, 10, and 14 all follow.
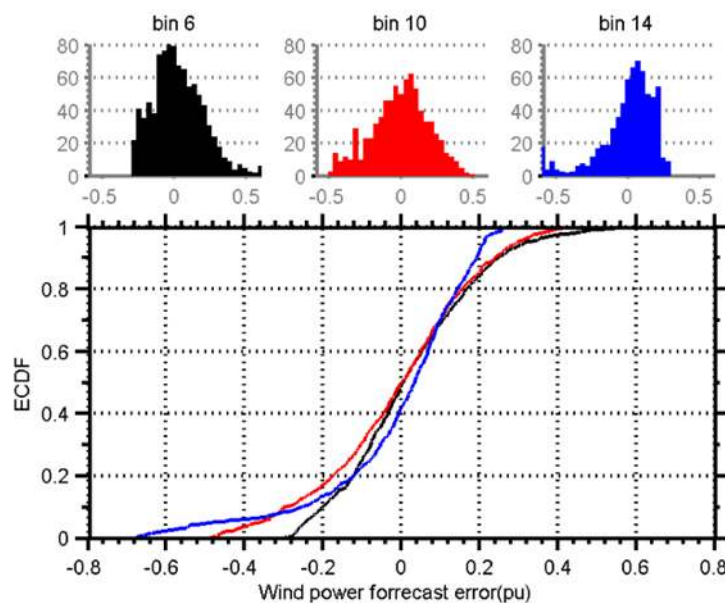

Figure 2. Error distributions within bins 6, 10, and 14 are in the upper panel of the figure. Corresponding ECDFs are given in the lower panel. The colors black, red, and blue represent bins 6, 10, and 14, respectively.

Cumulative distributions of errors within bin 6, 10, and 14 (given at the bottom of Figure 2) are quite different. Thanks to the non-parametric universal approach of the ECDF method, they are modeled successfully.

After finding the bin that a point forecast belongs to, one can generate scenarios by sampling random values from the ECDF of bin. The distribution transformation method proposed in [35] is used to sample the scenarios around point forecasts. In the literature this method is widely used to simulate wind scenarios [13,21,36,37]. A detailed explanation of this transformation method and the scenario generation procedure is given in the next section.

2.3. Scenario Generation Method

This section describes the theoretical background and application procedure of the proposed wind scenario generation method. Stochastic wind forecast errors of a wind farm can be expressed mathematically as $S = \{S_t, t \in T\}^T$. S_t is a forecast error at time t. One can find the bin number that the wind power forecast at time t (F_t) belongs to. This bin has an ECDF to sample potential error scenarios.

The first issue of our approach is sampling random scenarios from the ECDF. Inverse transform is a useful method to sample values from the ECDF. The procedure is as follows: Y_t is a generated random variable which follows a normal distribution (with a unit deviation and zero mean). The probability of random value Y_t is given in Equation (1). Then one can obtain the corresponding S_t through Equation (2).

$$\varphi(Y_t) = \int_{-\infty}^{Y_t} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy \tag{1}$$

$$S_t = f^{-1}(\varphi(Y_t)) \tag{2}$$

f^{-1} is an inverse transform function . An illustrative example of the sampling procedure for the 10th bin is given in Figure 3. Arrows indicate the direction of transformation. The generated random value Y_t is equal to 0.5. The probability ($\varphi(Y_t)$) of Y_t is 0.69. This value is equal to the probability of sampled value S_t. The sampled S_t is equal to 0.09.
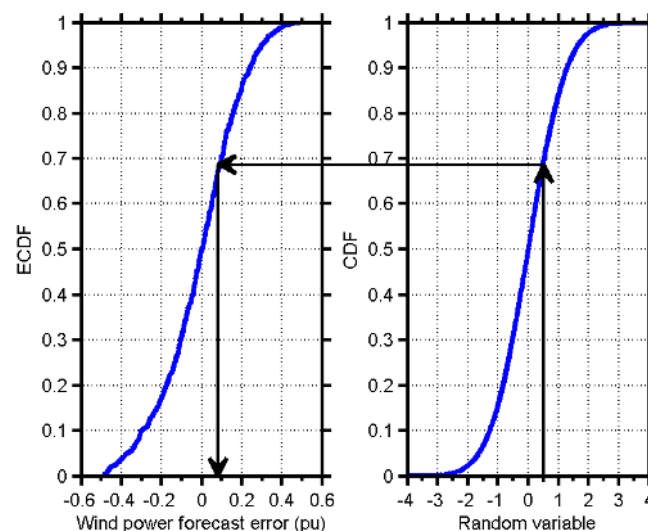

Figure 3. Distribution transformation example. The CDF of a normal distribution (**right**) and the ECDF of the error distribution within the 10th bin (**left**) are shown.

In this study, scenarios are generated around 24 point forecasts of the next day ($t = 1, \ldots., 24$). Y_t is a random variable for t^{th} point forecast. A random vector $\mathbf{Y} = \{Y_t, t \in T\}^T$ is introduced for 24 point forecasts. We assumed that if Y_t follows a normal distribution with unit deviation and zero

mean, Y will follow a multivariate normal distribution, $Y \sim N(\mu_0, \Sigma)$. The theoretical background of this joint distribution is given in [38]. It is a basic assumption that one can make about Y, and a more complex assumption could be made by using the theory of copulas [21]. Diagonal elements of the covariance matrix (Σ) are variances and they are very close to 1. The mean vector (μ_0) is a vector of zeros. A simple expression of Σ matrix is given in Equation (3):

$$\Sigma = \begin{bmatrix} \sigma_{1,1} & \cdots & \sigma_{1,24} \\ \vdots & \ddots & \vdots \\ \sigma_{24,1} & \cdots & \sigma_{24,24} \end{bmatrix} \tag{3}$$

Historical data can be used to obtain the covariance matrix Σ. The values of elements in the matrix varies with the period of the data. It is generally assumed that there is a seasonal effect on forecast system performance [39]. For this reason matrix Σ is updated recursively.

The second issue in our approach is calculating a suitable covariance matrix. For this purpose a simple recursive estimation formula proposed in [21] is preferred. This method estimates the Σ matrix recursively with a forgetting factor. The simple idea of this method is based on updating Σ recursively. The unbiased Σ that depends on historical data can be written as in Equation (4). Σ^t is a covariance matrix at time t. Y^t is a vector that denotes Y at time t:

$$\Sigma_t \ = \ \frac{1}{t-1} \sum_{k=1}^{t} Y^k Y^{k^T} \tag{4}$$

One can rewrite Equation (4) as a normalized sum of two terms. This new equation (Equation (5)) is a recursively-updating formula for Σ_t:

$$\Sigma_t = \frac{t-2}{t-1} \Sigma_{t-1} + \frac{1}{t-1} Y^t Y^{t^T} \tag{5}$$

Forgetting factor λ is introduced to catch the effect of the time-varying forecast system performance on the covariance matrix. For this reason Equation (5) is rewritten as Equation (6):

$$\Sigma_t = \lambda \left(\frac{t-2}{t-1} \right) \Sigma_{t-1} + \left(1 + \lambda \left(\frac{1}{t-1} - 1 \right) \right) Y^t Y^{t^T} \tag{6}$$

Equation (6) is a recursively-updating formula for Σ_t that includes forgetting factor λ. When t converges to infinity, Equation (6) transforms to the following form:

$$\Sigma_t = \lambda \Sigma_{t-1} + (1 - \lambda) Y^t Y^{t^T} \tag{7}$$

If the covariance matrix is obtained as explained above, one can generate the random variable set Y via a multivariate random number generator. Distribution transformation is a useful method to convert these jointly-distributed variables (Y) to forecast error scenarios (S). A detailed explanation of this process and overall algorithms are given in the next section.

2.4. Scenario Generation Algorithm

This section describes proposed wind power scenario generation algorithm. The algorithm generates scenarios around 24 point forecasts of the next day. Thus, each scenario vector has 24 elements. D is the number of scenarios. The flow diagram of the algorithm is given in Figure 4. A stepwise explanation of the algorithm is as follows.
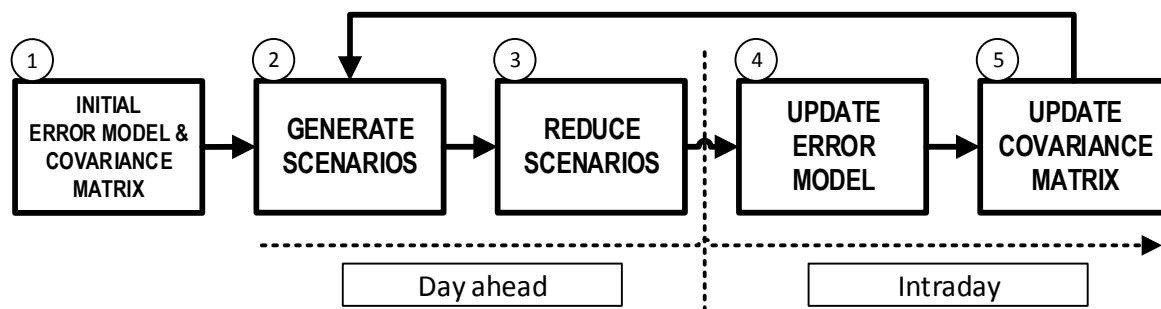
Figure 4. Flow diagram of the wind scenario generation algorithm.

- Step (1): If historical forecast data and actual error data are available, use them to obtain the initial error model and covariance matrix. The error model consists of 20 ECDFs obtained from historical errors occurring within 20 power levels (power bins). The error modelling procedure is given in Section 2.2. The MATLAB function ecdf is used in this procedure. Use Equation (7) to update Σ. If historical data is not available, one can set Σ to be equal to a unit matrix.
- Step (2): A multivariate random number generator is used to generate D scenarios. Each scenario $\boldsymbol{Y}$ has 24 elements. The MATLAB function mvnrnd is used as a random number generator. Then, bin numbers of point forecasts are obtained by using the MATLAB function ceil. These bin numbers indicate ECDFs which are used to transform each Y_t to S_t. The distribution transformation that was explained in the previous section is used to transform jointly-distributed numbers ($\boldsymbol{Y}$) to the scenario of forecast error $\boldsymbol{S}$.
- Step (3): D scenarios are obtained in Step 2. The scenario reduction algorithm mentioned in Section 2.4 is used to reduce number of scenarios D to d.
- Step (4): The ECDFs of the error model are updated after error realizations. Simply, all ECDFs are calculated again.
- Step (5): Use Equation (7) to update Σ. The update process is applied after realization of the forecast errors. Each measured forecast error (S_t) is transformed into a normally-distributed value (Y_t) via distribution transformation. A new $\boldsymbol{Y}^t$ is used in Equation (7). The algorithm goes back to Step 2.

The proposed method has five simple steps, as mentioned above. One must select values of two parameters λ and d before implementation of this algorithm. λ is a forgetting factor which takes a value between 0 and 1. This parameter affects the covariance matrix updating equation (Equation (7)). There is an inverse relationship between the λ value and the amount of forgotten historical data. Lower values close to 0 result in a large amount of forgotten historical data. On the contrary, larger values close to 1 result in a small amount of forgotten historical data.

It is possible to generate all probable scenarios by selecting very large D values. In this case accuracy of the stochastic process model increases, but it is very difficult to solve a stochastic model that includes a large number of scenarios. For this reason a scenario reduction algorithm is used to reduce the number of scenarios to d. The next section discusses the scenario reduction method.

2.5. Scenario Reduction Method

A large number of scenarios can successfully represent the stochastic process. One can use the aforementioned algorithm to generate D scenarios. However, it is not easy to solve stochastic decision-making problems that contain large numbers of scenarios. As discussed in [12,24,26,27], a reduction of the initial scenario set is required. The reduced scenario set should be very close to the initial one. If one can obtain the probability distances of the scenarios, it is possible to obtain the scenario set close to the original one. A scenario reduction method using the Kantorovich probability distance metric is proposed in [24]. Two different Kantorovich distance-based wind power scenario

set reduction algorithms are proposed in [28]. In this study, a "Fast Forward Selection Algorithm" is preferred to reduce the initial scenario set. A detailed explanation of this algorithm is given in [28]. The last issue of our study is with respect to assessing the quality of the reduced scenario set. For this purpose we introduced simple metrics to assess the quality of scenarios. The proposed metrics are explained in the next section.

2.6. Scenario Quality Metrics

A researcher must be sure that the quality of a generated scenario set is good enough to use them in calculations. As mentioned before, the number of a scenarios in the reduced scenario set affects the overall quality of the scenario set. For this reason a researcher must decide how many scenarios should be in the reduced scenario set. In this study we introduced two metrics to assess the overall quality. One can use them to select suitable parameters that results in good quality. The following section describes the calculation of the proposed quality metrics.

2.6.1. Mean Absolute Error Metric

The mean absolute error metric is the mean absolute difference between the mean of wind power scenarios and the measured values. Stochastic wind power values can be expressed mathematically as $P = \{P_t, t \in T\}^T$. P_t is the wind power at time t. P_t^i is the ith power scenario at time t. The following equations give the mean of wind power scenarios for time t:

$$P_t^i = F_t + S_t^i \tag{8}$$

$$\overline{P_t} = \sum_{i=1}^{d} \rho_i P_t^i \tag{9}$$

S_t^i is the ith error scenario and F_t is the wind power forecast for time t. d is the number of scenarios in the reduced scenario set. ρ_i is the probability of the ith scenario. The following equation gives the *mae* of the scenario set for a single day:

$$mae = \frac{1}{24} \sum_{k=1}^{24} \left| \overline{P_{t+k}} - P_{t+k} \right| \tag{10}$$

2.6.2. Sum of Deviation Metric

The sum of deviation metric is the sum of the distances between the scenario set region and measured values out of the scenario set region. Deviation under and over the scenario set is obtained by Equation (11). P^d is the scenario set that has has d scenarios for a forecast horizon. WP_t is the actual wind power measured at time t. de_t is the deviation for time t.

$$de_t = \begin{cases} \min\left(P^d\right) - WP_t, \mid WP_t < \min\left(P^d\right) \\ WP_t - \max\left(P^d\right), \mid WP_t > \max\left(P^d\right) \\ \qquad\qquad 0, \mid \min\left(P^d\right) \leq WP_t \leq \max\left(P^d\right) \end{cases} \tag{11}$$

The sum of deviation (*sde*) for the investigated day is equal to the sum of under and over deviations, as in Equation (12):

$$sde = \sum_{k=1}^{24} de_{t+k} \tag{12}$$

2.7. Practical Application

The simple practical application is performed to clarify the benefits of the proposed wind power scenario generation methodology. For this purpose, two day-ahead generation offering approaches for the Belen wind farm are investigated. In the first approach, day-ahead generation offers (x_h) are equal to day ahead point forecasts (F_t). As mentioned before, none of the forecast systems can obtain absolute accuracy yet. Forecast errors cause penalties in this approach. In the second approach all possible generations are represented with scenarios. The proposed methodology is used to obtain these scenarios. Then a stochastic optimization model is proposed and solved to obtain the best offers for all possible scenarios. An objective function and a constraint of the model are given in Equations (13) and (14):

$$max \sum_{s=1}^{S} \rho^s \sum_{h=1}^{H} [P.w_h^s - c.P.|w_h^s - x_h|] \tag{13}$$

$$\begin{array}{c} \text{s.t} \\ 0 \le x_h \le w^{max} \end{array} \tag{14}$$

The objective function maximizes the total income. Total income is equal to the sum of possible incomes minus the sum of possible penalties. A possible income of scenario *s* in hour *h* is equal to the multiplication of the electricity price P, generation scenario w_h^s, and the probability of occurrence ρ^s. The possible penalty of scenario *s* in hour *h* is equal to the multiplication of the penalty factor *c*, electricity price P, probability of occurrence ρ^s, and the absolute difference between w_h^s and offer x_h. x_h are "here and now" decisions in this stochastic model. They are expected to give optimal solutions when considering all possible generation scenarios. There is only one constraint that limits x_h to the installed power of wind farm, w^{max}.

Linear programming formulation is used to solve the proposed stochastic optimization model. A method proposed in [40] is used to linearize the absolute value of the objective function. Electricity price is intentionally kept constant in the model. The effect of price uncertainty is not in the scope of this study. One can add price scenarios to this model. The results of the application are given in the next section.

3. Tests and Results

Results and tests of the proposed scenario generation, reduction, quality assessment methods, and practical application are given in this section. The methods are tested on the dataset, which was described in Section 2.1. A personal computer with an AMD 2.0 Ghz processor (Advanced Micro Devices Inc., Sunnyvale, CA, USA) with 4 GB memory and MATLAB is used to implement the algorithms. The CPLEX solver (Version 12.7, IBM Corporation, Armonk, NY, USA) is used to solve the stochastic optimization model.

3.1. Application of the Proposed Scenario Generation Method

One can use the algorithm mentioned in Section 2.4 to generate a very large number of scenarios. In this study we generated 400 scenarios around 24 point forecasts. Each scenario is a probable realization, so a larger initial scenario set comprises a large number of probable realizations. On the other hand, it is not easy to solve a stochastic problem that contains too many scenarios. For this reason a scenario reduction algorithm called "fast forward selection" is used to reduce the number of scenarios. One can find a detailed application procedure for this algorithm in [28]. Another important issue regarding scenario generation and the reduction process is the quality of the reduced set. We introduced two metrics to evaluate quality. They are the *sde* and *mae* of the scenarios. A detailed explanation of them is given in Section 2.5. To find the suitable parameters, we performed quality tests on a one year period (2014). According to the test results, we decided to reduce the number of scenarios from 400 to 50. The next section describes the quality tests and test results.

3.2. Scenario Quality Assessment Tests

As mentioned before, the generated scenarios must have good quality. The number of scenarios in the initial scenario set is 400. They were reduced to (10, 20, 30, 40, 50) scenarios respectively. Scenario sets are generated for every single day of the test year. Values of the two introduced metrics are summarized in Figures 5 and 6.
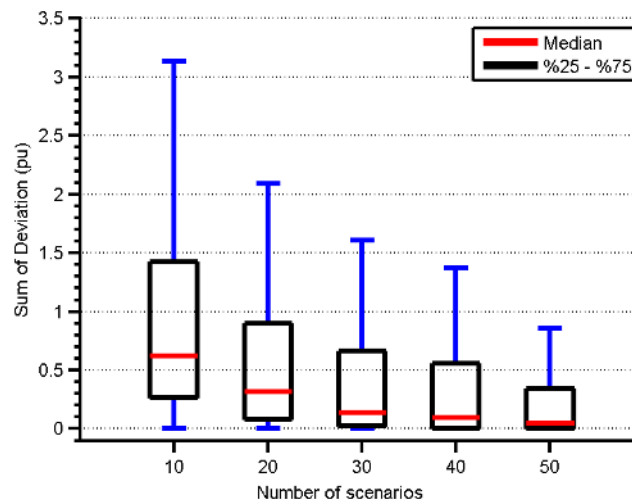

Figure 5. The black boxes comprise *sde* values between the 25th and 75th percentiles. The minimum adjacent value is equal to $q_1 - 1.5(q_3 - q_1)$ and the maximum adjacent value is equal to $q_3 + 1.5(q_3 - q_1)$, where q_1 and q_3 are the 25th and 75th percentiles, respectively.
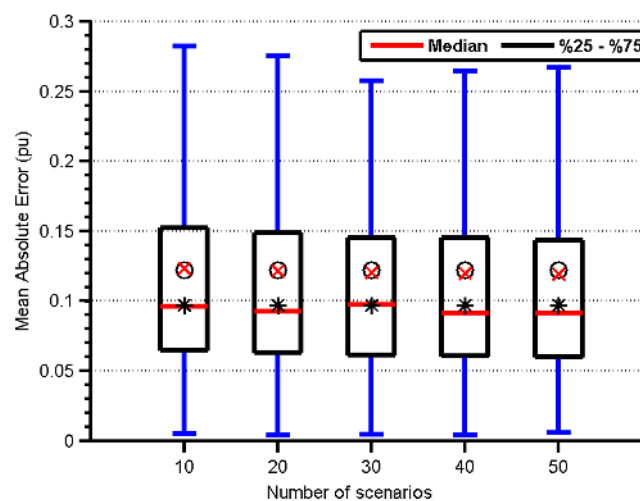

Figure 6. The black boxes comprise *mae* values between the 25th and 75th percentiles. The black circles are equal to the actual mean absolute forecast errors. The black stars are equal to the median of the actual forecast errors. The red x and red lines are the *mae* and median of the reduced scenario sets, respectively. The minimum adjacent value is equal to $q_1 - 1.5(q_3 - q_1)$ and the maximum adjacent value is equal to $q_3 + 1.5(q_3 - q_1)$, where q_1 and q_3 are the 25th and 75th percentiles, respectively.

Some actual measurements can be out of the scenario set region, which is an unwanted situation. The sum of the deviation metrics is equal to the sum of the distances between the scenario set and the actual measurements which are out of the scenario set. Thus, the smallest value in Figure 5 is the best result. Formulas for this metric are given in Section 2.5. As we expected, the test results showed that if the number of the reduced scenarios in the reduced scenario set increases, the value of this metric

decreases. When the number of scenarios reached 50, the obtained metric values are very close to zero. Thus, we concluded that 50 scenarios are enough to represent the stochastic wind power generation. The second test evaluates the *mae* of the scenario set. The results of this test are given in Figure 6.

The mean absolute error (*mae*) of a reliable scenario set must be less than, or close to, the actual mean absolute forecast error. The formula of the *mae* metric is given in Section 2.5. Our test results showed that the *mae* of all reduced scenario sets is close to the actual mean absolute forecast error. Being in accordance with the previous test results, the number of scenarios is set to 50.

3.3. Sample Scenario Sets

After investigating the quality test results, we generated scenario sets for eight selected days. They are low- and high-wind days of four seasons. The generated initial scenario set has 400 scenarios. The reduced scenario set has 50 scenarios. The results are given in Figure 7. These sample scenario sets show that the generated scenarios comprise almost all actual wind power generations.
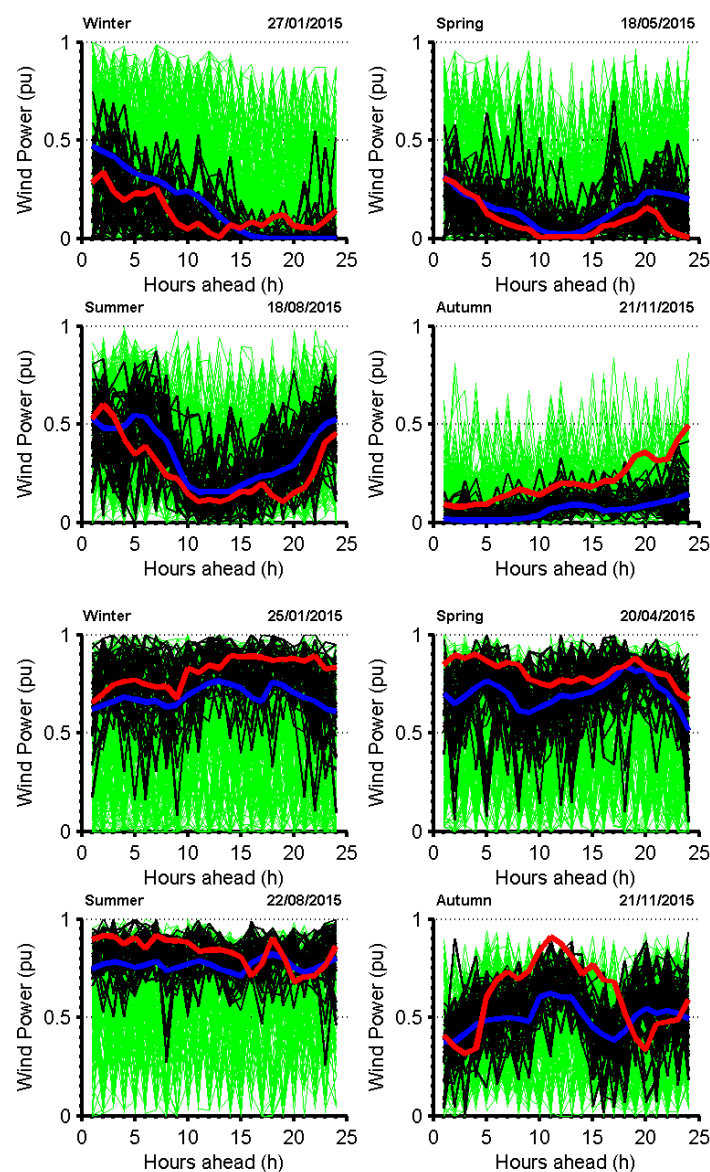

Figure 7. Green lines are the 400 initial scenarios, black lines are the 50 reduced scenarios, the red line is the actual power, and the blue line is the day-ahead forecast power. Scenarios are for high and low wind-occurring days of four seasons.

3.4. Results of Practical Application

The results of the two proposed approaches are compared in this section. Generation offers are equal to the day-ahead forecasts in Approach 1. The developed scenario generation tool is used in Approach 2. The penalty factor *c* is 0.15. It is the mean percentage of positive and negative imbalance penalties for the test year. As mentioned before, electricity price *P* is kept constant. It is equal to the mean market clearing price, 138.0415 TL (Turkish Lira, currency). The installed power of the wind farm w^{max} is equal to 48 MW. Two approaches are applied to eight selected days. The actual Turkish day-ahead market income calculation formula (Equation 15) is used to calculate the incomes of Approach 1 and Approach 2 offers. The results are shown in Table 2.

Table 2. Comparison of incomes for eight representative days.

Season	Wind Level	Approach1	Approach2	Increase
Winter	High	134,326.902	137,974.862	3647.961
	Low	17,512.958	17,727.418	214.459
Spring	High	74,887.819	76,305.077	1417.258
	Low	10,702.941	10,848.907	145.966
Summer	High	170,449.988	172,864.430	2414.442
	Low	43,831.233	44,395.968	564.735
Autumn	High	100,672.182	102,901.536	2229.354
	Low	40,005.274	40,519.178	513.904
Total Income (TL)		592,389.297	603,537.376	11,148.079

CP_h and MP_h are the actual clearing and marginal prices at hour h, respectively. w_h^g is the actual wind generation. The results showed that the proposed scenario generation tool is successful. Farm owners can use it to increase their income.

$$\boldsymbol{income} = x_h.CP_h + (w_h^g - x_h) \begin{cases} 1.03 \ .\max(CP_h, MP_h), \mid w_h^g < x_h \\ 0.97 \ .\min(CP_h, MP_h), \mid w_h^g > x_h \\ 0, \mid w_h^g = x_h \end{cases} \tag{15}$$

4. Discussion and Conclusions

This paper presents scenario generation, reduction, and quality testing methods to obtain a reduced number of wind power scenarios that have acceptable quality. Methods are not site-specific, so one can use them at any site and with any data. Additionally, the quality testing method can be used to confirm the quality of generated scenarios.

The scenario generation method uses historical forecast error distributions to obtain scenarios for stochastic wind power generation. In the literature there is a conflict of selecting a distribution function that the historical error data fits [22,30–33]. Our method uses the ECDF approach to model the historical data distribution. This approach is nonparametric and universal, so there is no need to select a specific distribution function (Normal, Beta, etc.). In the application process of the scenario generation method, historical power forecasts are sorted into power bins (intervals). It has been seen that performance of the forecast system varies with the level of the forecast bin. Each bin has an ECDF that represents the forecast error distribution in its interval. A scenario reduction algorithm called "fast forward selection" [33] is preferred to reduce the number of scenarios. This reduction makes it easy to solve stochastic problems that involve large scenario sets.

Another important issue of this study is assessing the quality of the reduced scenario set. Our quality testing method uses two proposed quality metrics. These metrics are *sde* and **mae**. *sde* represents the number of measurements out of the scenario set. **mae** indicates the difference between measurements and mean of scenario set. The values close to, or smaller than, forecast errors are acceptable. Test results showed that the proposed metrics are useful to assess the quality of the

generated scenarios. A simple practical application showed that the developed scenario generation tool is successful. Wind farm owners will use it to increase incomes.

It must be pointed out that the proposed scenario generation method needs sufficient historical data to obtain ECDFs. Thus, this method is suitable for sites where a sufficient number of historical generation measurements and forecasts are available. On the other hand, in this study and in the literature, it is assumed that the performance of the forecast system varies by the power bin. However, the effect of wind direction on performance has not been considered yet. The authors are currently working on including the direction effect into the scenario generation procedure.

Author Contributions: M.S. managed the research; C.Y. developed algorithms and models; F.K. and A.G. researched the background literature; and H.A. and M.T. wrote and revised the manuscript. All authors read and approved the final manuscript.

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

References

1. Cox, S.; Tegen, S.; Gould, I.B.; Oteri, F.A.; Esterly, S.; Forsyth, T.; Baranowski, R. Policies to Support Wind Power Deployment Key Considerations and Good Practices. Available online: http://www.nrel.gov/docs/fy15osti/64177.pdf (accessed on 27 September 2016).
2. World Energy Resources 2016. Available online: https://www.worldenergy.org/wp-content/uploads/2016/10/World-Energy-Resources_SummaryReport_2016.pdf (accessed on 27 September 2016).
3. Foley, A.M.; Leahy, P.G.; Marvuglia, A.; McKeogh, E.J. Current methods and advances in forecasting of wind power generation. *Renew. Energy* **2011**, *37*, 1–8. [CrossRef]
4. Baringo, L.; Conejo, A.J. Offering Strategy of Wind-Power Producer: A Multi-Stage Risk-Constrained Approach. *IEEE Trans. Power Syst.* **2016**, *31*, 1420–1429. [CrossRef]
5. Mestre, V.G.; Nieta, A.A.S.; Contreras, J.; Catalão, J.P.S. Optimal bidding of a group of wind farms in day-ahead markets through an external agent. *IEEE Trans. Power Syst.* **2016**, *31*, 2688–2700. [CrossRef]
6. Yang, S.; Zeng, D.; Ding, H.; Yao, J.; Wang, K.; Li, Y. Stochastic security-constrained economic dispatch for random responsive price-elastic load and wind power. *IET Renew. Power Gen.* **2016**, *10*, 936–943. [CrossRef]
7. Uçkun, C.; Botterud, A.; Birge, J.R. An improved stochastic unit commitment formulation to accommodate wind uncertainty. *IEEE Trans. Power Syst.* **2016**, *31*, 2507–2517. [CrossRef]
8. Ji, L.; Huang, G.H.; Huang, L.C.; Xie, Y.L.; Niu, D.X. Inexact stochastic risk-aversion optimal day-ahead dispatch model for electricity system management with wind power under uncertainty. *Energy* **2016**, *109*, 920–932. [CrossRef]
9. González, J.G.; Muela, R.M.R.; Santos, L.M.; González, A.M. Stochastic joint optimization of wind generation and pumped-storage units in an electricity market. *IEEE Trans. Power Syst.* **2008**, *23*, 460–468. [CrossRef]
10. Bouffard, F.; Galiana, F.D. Stochastic security for operations planning with significant wind power generation. *IEEE Trans. Power Syst.* **2008**, *23*, 306–316. [CrossRef]
11. Meibom, P.; Barth, R.; Hasche, B.; Brand, H.; Weber, C.; O'Malley, M. Stochastic optimization model to study the operational impacts of high wind penetrations in Ireland. *IEEE Trans. Power Syst.* **2011**, *26*, 1367–1379. [CrossRef]
12. Sharma, K.C.; Jain, P.; Bhakar, R. Wind power scenario generation and reduction in stochastic programming framework. *Electr. Power Compon. Sys.* **2013**, *41*, 271–285. [CrossRef]
13. Morales, J.M.; Mínguez, R.; Conejo, A.J. A methodology to generate statistically dependent wind speed scenarios. *Appl. Energy* **2010**, *87*, 843–855. [CrossRef]
14. Oskouei, M.Z.; Yazdankhah, A.S. Scenario-based stochastic optimal operation of wind, photovoltaic, pump-storage hybrid system in frequency-based pricing. *Energy Convers. Manag.* **2015**, *105*, 1105–1114. [CrossRef]
15. Matevosyan, J.; Söder, L. Minimization of imbalance cost trading wind poweron the short-term power market. *IEEE Trans. Power Syst.* **2006**, *21*, 1396–1404. [CrossRef]

16. Hosseini, F.M. Optimal offering strategy considering the risk management for wind power producers in electricity market. *Int. J. Eelectr. Power* **2013**, *49*, 359–368. [CrossRef]

17. Parastegari, M.; Hooshmand, R.A.; Khodabakhshian, A.; Zare, A.H. Joint operation of wind, photovoltaic, pump-storage and energy storage devices in energy and reserve markets. *Int. J. Eelectr. Power* **2015**, *64*, 275–284. [CrossRef]

18. Sharma, K.C.; Bhakar, R.; Tiwari, H.P. Strategic Bidding for Wind Power Producers in Electricity Markets. *Energy Convers. Manag.* **2014**, *86*, 259–267. [CrossRef]

19. Wan, Y.H.; Ela, E.; Orwig, K. Development of an Equivalent Wind Plant Power-Curve. In Proceedings of the WindPower 2010, Dallas, TX, USA, 22–26 May 2010.

20. Quevedo, P.M.; Contreras, J. Optimal placement of energy storage and wind power under uncertainty. *Energies* **2016**, *9*, 528. [CrossRef]

21. Pinson, P.; Madsen, H.; Nielsen, H.A.; Papaefthymiou, G.; Klöckl, B. From probabilistic forecasts to statistical scenarios of short-term wind power production. *Wind Energy.* **2008**, *12*, 51–62. [CrossRef]

22. Bludszuweit, H.; Domínguez-Navarro, J.A.; Llombart, A. Statistical analysis of wind power forecast error. *IEEE Trans. Power Syst.* **2008**, *23*, 983–991. [CrossRef]

23. Ma, X.Y.; Sun, Y.Z.; Fang, H.L. Scenario generation of wind power based on statistical uncertainty and variability. *IEEE Trans. Sustain. Energy* **2013**, *4*, 894–904. [CrossRef]

24. Dupaˇcov´a, J.; Growe-Kuska, N.; Romisch, W. Scenario reduction in stochastic programming an approach using probability metrics. *Math. Program.* **2003**, *95*, 493–511. [CrossRef]

25. Sumaili, J.; Keko, H.; Miranda, V.; Zhou, Z.; Botterud, A.; Wang, J. Finding Representative Wind Power Scenarios and their Probabilities for Stochastic Models. In Proceedings of the 2011 16th International Conference on Intelligent System Applications to Power Systems, Hersonissos, Greece, 25–28 September 2011.

26. Heitsch, H.; Römisch, W. Scenario reduction algorithms in stochastic programming. *Comput. Optim. Appl.* **2003**, *24*, 187–206. [CrossRef]

27. Growe-Kuska, N.; Heitsch, H.; Romisch, W. Scenario Reduction and Scenario Tree Construction for Power Management Problems. In Proceedings of the 2003 IEEE Bologna PowerTech Conference, Bologna, Italy, 23–26 June 2003.

28. Sarfati, M. Modeling the Diversification Benefit of Transmission Investments in the Presence of Uncorrelated Generation Sources. Master's Thesis, KTH Royal Inst. of Technology, Stockholm, Sweden, 2012.

29. Terciyanlı, E.; Demirci, T.; Küçük, D.; Buhan, S.; Özkan, B.M.; Köksoy, C.E.; Koç, E.; Kahraman, C.; Haliloğlu, A.B.; Demir, T.; et al. The Architecture of a Large-Scale Wind Power Monitoring and Forecast System. In Proceedings of the 4th International Conference on Power Engineering, Energy and Electrical Drives, Istanbul, Turkey, 13–17 May 2013.

30. Pappala, V.S.; Erlich, I.; Rohrig, K.; Dobschinski, J. A Stochastic model for the optimal operation of a wind-thermal power system. *IEEE Trans. Sustain. Energy.* **2009**, *24*, 940–950. [CrossRef]

31. Makarov, Y.V.; Etingov, P.V.; Ma, J. Incorporating uncertainty of wind power generation forecast into power system operation, dispatch, and unit commitment procedures. *IEEE Trans. Sustain. Energy* **2011**, *2*, 433–442. [CrossRef]

32. Tewari, S.; Geyer, C.J.; Mohan, N.; Huang, Z.; Subbarao, K. A statistical model for wind power forecast error and its application to the estimation of penalties in liberalized markets. *IEEE Trans. Power Syst.* **2011**, *26*, 2031–2039. [CrossRef]

33. Pinson, P. Estimation on the Uncertainty in Wind Power Forecasting. Ph.D. Thesis, Mines Paris Tech. Graduate School, Paris, France, 23 March 2006.

34. Spanos, A. *Probability Theory and Statistical Inference: Econometric Modeling with Observational Data.*, 2nd ed.; Cambridge University Press: New York, NY, USA, 2003; pp. 229–253.

35. Liu, P.L.; Kiureghian, A.D. Multivariate distribution models with prescribed marginals and covariances. *Probabilist. Eng. Mech.* **1986**, *1*, 105–112. [CrossRef]

36. Conejo, A.J.; Carrión, M.; Morales, J.M. *Decision Making Under Uncertainty in Electricity Markets*; Springer: New York, NY, USA, 2010; pp. 74–75.

37. Villanueva, D.; Feijóo, A.; Pazos, J.L. Simulation of correlated wind speed data for economic dispatch evaluation. *IEEE Trans. Sustain. Energy* **2012**, *3*, 142–149. [CrossRef]

38. Bernard, F. *A First Course in Multivariate Statistics*, 1st ed.; Springer: New York, NY, USA, 1997; pp. 171–207.

39. Wyszogrodzki, A.A.; Liu, Y.; Jacobs, N.; Childs, P.; Zhang, Y.; Roux, G.; Warner, T.T. Analysis of the surface temperature and wind rorecast errors of the NCAR-AirDat operational CONUS 4-km WRF forecasting system. *Meteorol. Atmos. Phys.* **2013**, *122*, 125–143. [CrossRef]
40. Pousinhoa, H.M.I.; Mendesc, V.M.F.; Catalãoa, J.P.S. A stochastic programming approach for the development of offering strategies for a wind power producer. *Electr. Power Syst. Res.* **2012**, *89*, 45–53. [CrossRef]