

Multi-Stage Statistical Approach to Wind Power Forecast Errors Evaluation: A Southern Sulawesi Case Study

Dhany Harmeidy Barus^{a,*}, Rinaldy Dalimi

^a *Electrical Engineering Department, Universitas Indonesia, Jl. Margonda Raya, Depok, 16424, Indonesia*

*Corresponding author: *dhany.harmeidy@ui.ac.id*

Abstract— Wind Power Plant (WPP) is part of renewable energy sources, with rapid expansion worldwide. It has the advantages of clean and green energy, but its uncertainty leads to an additional grid integration cost. The uncertainty of wind power output is much dependent on the accuracy of the wind power forecast (WPF) result. Since there is no perfect wind power forecast, understanding the current system's forecast accuracy characteristics is essential in expecting typical errors faced in the future. This paper proposed a new algorithm of the statistical approach method to evaluate characteristics of wind power forecast errors (WPFE) from an observed power system with high-penetration WPP. This method combined the approach of scatter diagram, statistical distribution, standard error performance, and score weighting in a multi-stage algorithm. It consists of serial and parallel processes to check the consistency of the results. In this study, a comprehensive analysis was made of various scenarios based on location and timescale. This proposed algorithm has been successfully tested on statistical data of Sidrap WPP and Jeneponto WPP in the Southern Sulawesi power system. The result showed that the scenario with the aggregation of both WPPs in hour-ahead timescale has the most accurate and consistent performance among all scenarios. It demonstrated specific characteristics of WPFE in the observed power system that can be used as an essential starting point in conducting future wind integration expansion studies.

Keywords—Statistical approach; multi-stage; wind power forecast errors.

*Manuscript received 30 Jun. 2020; revised 2 Dec. 2020; accepted 25 Jan. 2021. Date of publication 30 Apr. 2021.
IJASEIT is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.*



I. INTRODUCTION

The integration of a Wind Power Plant (WPP) into a power system is often constrained by wind power resources' uncertain and variable characteristics. As the increased contribution of WPP in the power system also impacts increasing the uncertainty of the WPP output power. It affects the increased need for operating reserves from the load follower power plants, the provision of energy storage that eventually will cause an increase in the integration cost of the WPP. One cause of the uncertainty of WPP is the low accuracy of the wind power forecast (WPF). The error of the day-ahead forecast can be as high as 25% up to 40% of the installed capacity of WPP [1].

Statistical methods can be used for a short-term forecast. However, for the longer-term one, it must be collaborated with Numerical Weather Prediction (NWP) models to provide better wind power forecast accuracy [2], [3]. Typically, operational forecasting of a utility uses NWP to obtain wind power forecast data within its working area based on global and local wind speed forecasts using specific statistical

techniques [4]–[6]. Furthermore, some studies try to explore the results of the NWP model from the perspective of wind power forecast errors (WPFE) [7], [8]. The characteristics of WPFE can be used to optimize the reserve management of a power system, especially during the phase of wind integration expansion studies [9], [10]. Han *et al.* [11] have explored multi-time scale rolling economic dispatch using WPFE, while Yu *et al.* [12] optimize WPFE compensation for energy storage sizing optimization.

Some statistical approaches are used to describe WPFE phenomena. In some studies, the statistical distribution is used to examine the WPFE characteristics [13]–[16]. The Normal Distribution is the most common model used to describe the WPFE phenomena [17]–[20]. The studies were carried out in various locations and several forecast timescales. In the research of Miettinen *et al.* [21], [22], the characteristics of day ahead WPFE in Nordic countries were carried out. They used some standard error performance parameters as statistical tools. It explains the characteristics of various locations and geographical conditions in terms of aggregation. The uncertainty of WPF was diminishing as the wider

geographical area of the WPP. Several studies used the scatter diagram in their visualization to illustrate the dispersion of WPF data [23]–[25]. It helps to explain the random nature of the existing dataset. Spiliotis *et al.* [26] use the wind speed and wind direction parameters in the scatter diagram, while Sun *et al.* [27] use wind power and time as the parameters. However, most of the studies used only one or two statistical approaches to the WPF evaluation in their experiments or simulations.

This paper proposes a new and more comprehensive method of WPF evaluation using a statistical approach. Instead of using one or two statistical perspectives, this breakthrough method is using a multi-stage algorithm. It combines the scatter diagram's depiction, the power of statistical distribution, the accuracy of standard error performance, and the score weighting method's orderliness. The last-mentioned method is brand new in evaluating the characteristic of WPF. It is proposed in this multi-stage method to summarize the result of the previous stages and sort them proportionally. Complemented by radar chart as a visualization tool, the score weighting method will systematically describe the WPF evaluation results, primarily when related to aggregation and timescale aspects. By this method, the evaluation processes are more comprehensive and cover more aspects of the statistical approach.

The question to be answered from the research is: *How to optimize statistical approach tools in evaluating the characteristics of WPF, primarily when related to the aspects of aggregation and timescales?* The contribution of this manuscript is to offer a comprehensive method for understanding the WPF characteristics of observed data that can be used as a reference in future wind integration expansion studies. This multi-stage method has been effectively tested on the Southern Sulawesi power system in Indonesia, part of the South East Asia Region that has a tropical climate. Furthermore, as a pioneer and the only large-scale WPP currently in Indonesia. This study determines the development direction of WPP expansion in other major power systems in Indonesia.

The next sections of the article are organized as follows. Section-2 presents the dataset used and explores methods with a detailed algorithm to process the data. Section-3 describes the results in various curves and graphics as a representation of the WPF characteristics. Then proceed with a discussion of the results and offers consequences for future studies. Section-4 gives a conclusion from the discussion and suggests further works to be done.

II. MATERIAL AND METHODS

A. Dataset of Observed Power System

In Indonesia, there are five major interconnected power systems, that all are operated by PLN (Perusahaan Listrik Negara) as the only state-owned electricity company in Indonesia. One of the Southern Sulawesi interconnected power systems [28] that PLN UIKL Sulawesi manages as the operational unit. This power system covers four provinces: South Sulawesi, West Sulawesi, Central Sulawesi, and Southeast Sulawesi. This power system's highest peak load is around 1411MW, with a total power plant capacity of

1790MW. Table 1 shows the composition of all types of power plants connected to the Southern Sulawesi power system. It shows that more than 50% of the generation comes from Coal Steam PPs commonly used as based load power plants. There are 20% of Gas PP for load followers, 15% of Hydro PP for peakers, and 5% of Diesel PP that standby. The WPP (7% of total) is operated as a must-run unit. Currently, it is the first, and the only large-scale WPP installed and operated in Indonesia.

TABLE I
COMPOSITION OF POWER PLANTS IN SOUTHERN SULAWESI POWER SYSTEM

Type of Power Plant	Capacity (MW)	Portion (%)
Coal Steam PP	940	53%
Hydro PP	260	15%
Gas PP	363	20%
Diesel PP	97	5%
Wind PP	130	7%
TOTAL	1790	100%

As seen in Table I [28], there are 130MW of WPP that connected to the Southern Sulawesi power system. Since the contribution is more than 5% of the total capacity, it is categorized as high-penetration WPP. An overview of the technical specifications of both WPP can be seen in Table II. The Sidrap WPP has 70MW of contract capacity while the Jenepono WPP has 60MW and 6MW as an additional option. The Sidrap WPP has a smaller size but with more units than the Jenepono WPP. It started to operate in January 2018 and March 2019, respectively. The Jenepono WPP has more modern technology of dispatch that is already equipped with Automatic Generation Control (AGC).

TABLE III
TECHNICAL SPECIFICATION OF THE WPP

Technical Spec.	Sidrap WPP	Jenepono WPP
Size of Unit (MW)	2.5	3.6
Number of Units	30	20
Capacity (MW)	70	60
Operation Date	Jan-18	March-19

This study's forecast database was obtained from the wind power forecast provider's output in Sidrap WPP and Jenepono WPP. Coincidentally, both databases are produced by the same provider, Meteologica, to attain a fair comparison since they come from the same forecast provider. The data used has been normalized for comparison purposes.

The scenarios of this study use a combination of 2 main categories as follows:

- Timescale: Day-Ahead Forecast and Hour-Ahead Forecast
- Location: Sidrap WPP, Jenepono WPP, and South Sulawesi WPP (aggregation of the two WPP).

There are six scenarios obtained from the combinations as follows:

- Sidrap WPP with Day-Ahead Forecast (SWDAF)
- Sidrap WPP with Hour-Ahead Forecast (SWHAF)
- Jenepono WPP with Day-Ahead Forecast (JWDAF)
- Jenepono WPP with Hour-Ahead Forecast (JWHAF)
- South Sulawesi WPP with Day-Ahead Forecast (SSWDAF)
- South Sulawesi WPP with Hour-Ahead Forecast (SSWHAF)

The range of observed data is from March to December 2019 with a time resolution of 30 minutes, both for Day-Ahead Forecast and Hour-Ahead Forecast. It means there are 14,688 data rows in every column (forecast and actual) of each scenario.

The forecast and actual data sets [29] are taken from dispatcher log sheets in the Southern Sulawesi Load Dispatch Center (SSLDC). The data is sent from each WPP to SSLDC through the existing Supervisory Control and Data Acquisition (SCADA) system. In this case, wind forecasting is using the Decentralized Forecasting method. This method's advantage is higher accuracy for individual WPP since it will encourage the WPP operator to be more innovative to minimize the penalty for the error result. But the disadvantage of this method is that it will increase the production cost of the WPP and cause more effort to do aggregation of wind power forecasting.

B. Algorithm of Multi-Stage Statistical Approach Method

The schema of the Multi-Stage Statistical Approach Algorithm is shown in Figure 1. It shows that the algorithm consists of 4 main stages. Stage 1 is the initial process, stage 2 and stage 3 are processed in parallel, while stage 4 is the final process that takes the results of the previous stages as input.

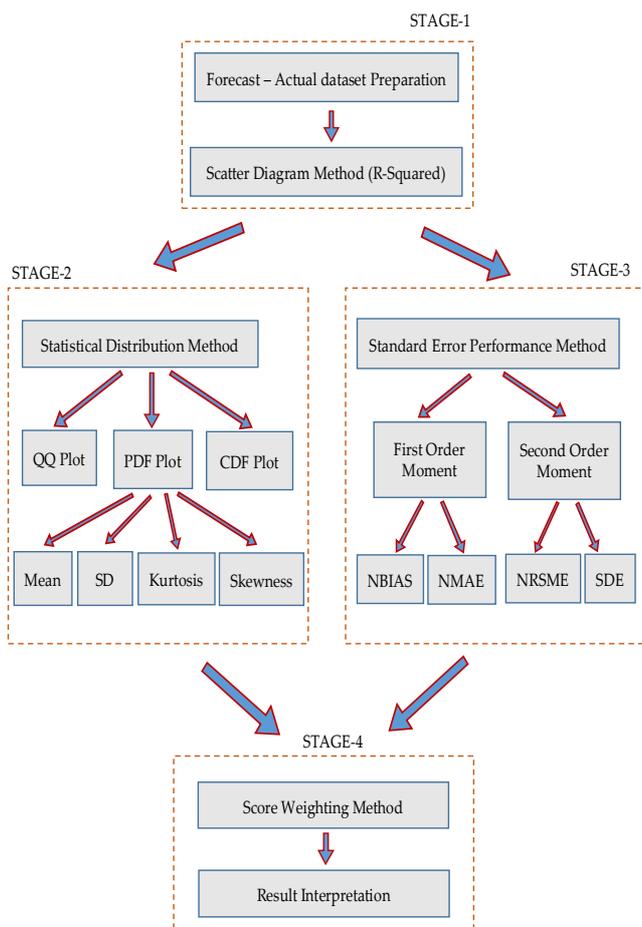


Fig. 1 Schema of Multi-Stage Statistical Approach Algorithm

Stage 1 starts with preparing the dataset for analysis purposes. The forecast and actual datasets are the primary source database. The data is arranged according to

predetermined scenarios. Then a mapping is done between forecast and actual data for each predetermined scenario with a scatter diagram. Stage 2 contains the Statistical Distribution Method, which evaluates the observed data's distribution characteristics with some indicators represented by various curves. Three kinds of graphs cover this stage. Stage 3 covers a Standard Error Performance Method, which aims to calculate the accuracy factor of existing data. This method consists of two orders of moments that evaluate different deviation factors. Stage 4 uses the Score Weighting Method, which summarizes the results of stage 1 until stage 3. The results of this last stage will be the conclusions of the whole process. The purpose and illustration of each method will be explained thoroughly in the next sub-sections.

1) *Scatter Diagram Method*: The scatter diagram is used to perceive the correlation or comparison of two pieces of information in a problem. It will check whether a variable can replace another one. In the context of WPF accuracy, a comparison is made between the forecast and the actual dataset. The distribution map of a scatter diagram can describe the dispersion level of deviation throughout the dataset. A linear trendline with the R-Squared (R^2) value of the distributed data that compared with the ideal trendline will describe the dispersion level of the WPF.

2) *Statistical Distribution Method*: For this method, we use some standard statistical distribution tools such as Histograms of Probability Density Function (PDF) Plots, Quantile-Quantile (QQ) Plots, and Cumulative Distribution Function (CDF) Plots [30]. The PDF Plot illustrates the range of values that can be assumed by random variables, as well as the probability of samples occurring within a specific range. The QQ Plot is a normal one comparing the observed distribution with the Normal Distribution with averages and standard deviations as observed distribution. It is represented by lines in the first and third quantiles of the observed data. If two distributions have the same characteristics, all points on the observed line must be reference distribution points. The CDF plot shows the significant random errors that might occur from the observed distribution against the reference.

In terms of PDF Plots, four standard parameters are used in the study. Mean (μ), Standard Deviation (σ), Skewness (γ), and Kurtosis (κ) are commonly used to analyze the observed distribution. Mean (μ) and Standard Deviation (σ), the first two statistical moments, are often used to characterize the distribution of wind power forecast errors. Mean is a measure of central tendency either from a possible distribution or from random variables that are marked by that distribution. Standard deviation is a parameter to measure the dispersion of data around the mean.

Skewness (γ) is the third moment, which is a measure of a distribution asymmetry. If skewness=0, the data are perfectly symmetrical. Positive skewness shows that more data is spread on the right-hand side. It is indicated by the existence of a long right tail of the distribution. Otherwise, negative skewness shows that more data is spread on the left-hand side. It is indicated by the existence of a long-left tail of the distribution. Table III shows the rule of thumb suggested by Bulmer [31].

TABLE III
DISTRIBUTION TYPE OF SKEWNESS RANGE

Type Skewness	Range of Skewness (γ)	Distribution Type
Positive & Negative	$\gamma < -1.0,$ $\gamma > +1.0$	Highly Skewed
Negative	$-1.0 \leq \gamma \leq -0.5$	Moderately Skewed
Positive	$+0.5 \leq \gamma \leq +1.0$	Moderately Skewed
Positive & Negative	$-0.5 < \gamma < +0.5,$ $\gamma \neq 0$	Approximately Symmetric
Neutral	$\gamma = 0$	Perfectly Symmetrical

Kurtosis (κ) is the fourth moment, which is a measure of distribution peak. However, kurtosis may be used to explain the fatness of the distribution tail. Table 4 shows the standard reference suggested by Westfall [32]. In this study, the excess kurtosis is always presented, simply mention as kurtosis.

TABLE IV
STANDARD REFERENCE OF KURTOSIS RANGE

Type Kurtosis	Range of Kurtosis (κ)	Type of Tails / Peak
Platykurtic	$\kappa < 3,$ excess < 0	Shorter and Thinner / Lower and Broader
Mesokurtic	$\kappa \approx 3,$ excess ≈ 0	Normal
Leptokurtic	$\kappa > 3,$ excess > 0	Longer and Fatter / Higher and Sharper
Type Kurtosis	Range of Kurtosis (κ)	Type of Tails / Peak

3) *Standard Error Performance Method:* Based on Madsen [33], the forecast error (e) is interpreted as the deviation between the observed (P) and the forecasted value (\hat{P}) with time function (t), which defined as:

$$e(t) = P(t) - \hat{P}(t) \quad (1)$$

The methods commonly used in evaluating the performance level of WPFE are Bias, Mean Absolute Error (MAE), Root Mean Square Error (RSME), and Standard Deviation of Errors (SDE). Bias, which refers to systematic errors, is defined as the average errors for the entire observed duration and calculated for each time function.

$$Bias(t) = \frac{1}{14688} \sum_{t=1}^{14688} e(t) \quad (2)$$

MAE, which refers to absolute systematic errors, is defined as the average absolute errors for the entire observed duration and calculated for each time function.

$$MAE(t) = \frac{1}{14688} \sum_{t=1}^{14688} |e(t)| \quad (3)$$

RMSE, which refers to root square systematic errors, is defined as the average root square error for the entire observed duration and calculated for each time function.

$$RMSE(t) = \left(\frac{1}{14688} \sum_{t=1}^{14688} e^2(t) \right)^{\frac{1}{2}} \quad (4)$$

SDE, which refers to the error distribution's standard deviation, is defined as the average root square error deviation for the entire observed duration and calculated for each time function.

$$SDE(t) = \left(\frac{1}{14688 - 1} \sum_{t=1}^{14688} (e(t) - \bar{e}_t)^2 \right)^{\frac{1}{2}} \quad (5)$$

Statistically, the Bias and MAE values are related to the first-order moments of forecast error as indications of the measurement that directly referred to the energy produced by the WPP. The RMSE and SDE values are related to second-order moments, as forecast error variances. For these moments, large forecast errors have the most significant effect. All parameters of errors presented are calculated using normalized forecasted errors to the installed capacity, $\mathcal{E}(t)$. The purpose of using normalized error data is to produce results that are independent of the size of the WPP. The new parameters are then referred to as Normalized BIAS (NBIAS), Normalized MAE (NMAE), Normalized RMSE (NRMSE), and Normalized SDE (NSDE). In the next chapter, NBIAS, NMAE, NRMSE, and NSDE will be used to process the data and analyze the results.

4) *Score Weighting Method:* This method is used to equalize the result parameter weights from the previous processes. The results are arranged in the form of a score-weighting table. This method is equipped with a radar chart that describes the score weighting table in the form of a two-dimensional graph quantitatively represented by an axis with the same reference. It is a graph or plot consisting of radii that explain the value of a variable. The length of the radius is equal to the value of the variable. Then a line is drawn connecting the data values. It forms a radar or star-shaped plot. In the context of WPFE with a multi-stage method, the results of stage-1 until stage 3 are weighted based on each variable to gain a final score in percentage scales. Those variables will be a plot in the radar chart to describe the characteristics of each scenario. The outputs of this method will be the final result of the multi-stage method. Result interpretation is made based on each scenario's final score as a conclusion of the whole multi-stage method.

III. RESULT AND DISCUSSION

Figure 2 to 5 show the result of stages 1 and 2. Each figure consists of 6 scenarios, which are a combination of three entities and two timescales. Figure 6 to 8 display the results of stage 3, which are bar charts of 4 parameters in standard error performance. Figure 9 shows the result of stage 4. It represents a summary of all variables in the multi-stage series.

A. Result of Stage 1

Figure 2 shows the scattered data of forecast versus actual wind power output. The blue dots represent the dispersion of the observed data. There are 14,688 dots in each diagram from 9 months of the statistical dataset. The black line shows the ideal trendline of a perfect forecast result (with R-Squared (R^2) value equal to 1). The red line is a representation of observed data, with an R^2 value, which scores the dispersion level. Stage 1 shows an initial overview of the WPFE.

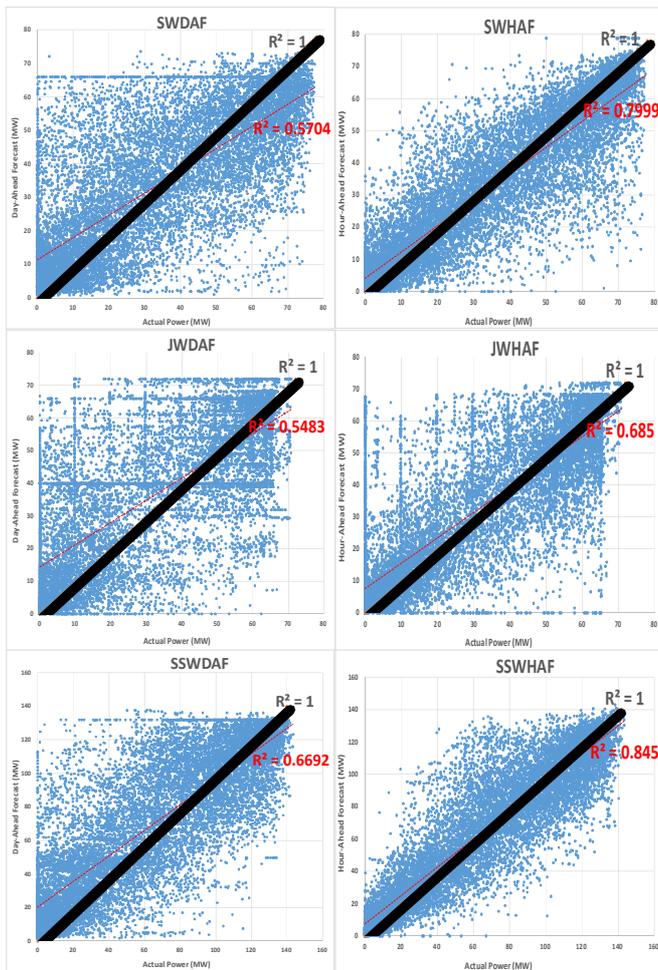


Fig. 2 Dispersion Characteristics of Wind Power Forecast (WPF)

This method demonstrates the distribution characteristics of all predefined scenarios. It can be seen that in the context of WPF, Sidrap WPP has a better dispersion index compared to Jenepono WPP. But the aggregation of the two WPP has the best dispersion index among the three entities. From a timescale perspective, it can be seen that Hour-Ahead has a better dispersion index for all three entities than Day-Ahead. The scatter diagram outlines that the South Sulawesi WPP with Hour-Ahead Forecast (the aggregation one) has the best dispersion index with 0.845 of R^2 value.

B. Result of Stage 2

Figure 3 shows the PDF Plot histogram of normalized data. The blue shaded histogram curve represents the normalized errors of the observed data. The red line represents a Normal Distribution that has the same average and standard deviation as observed errors. Analysis in Stage 2 begins by displaying a histogram of the PDF plot for all existing scenarios. Figure 3 shows that there are various characteristics measured by four indicators, namely Mean, Standard Deviation (SD), Skewness, and Kurtosis. Based on the Mean indicator, it can be seen that the SWDAF has the most distant value from the ideal conditions, while the SSWHAF almost has the ideal value. For the SD indicator, JWDAF has the highest value, while SWHAF has the smallest one. This indicator shows that SWHAF has the lowest distribution compared to other scenarios.

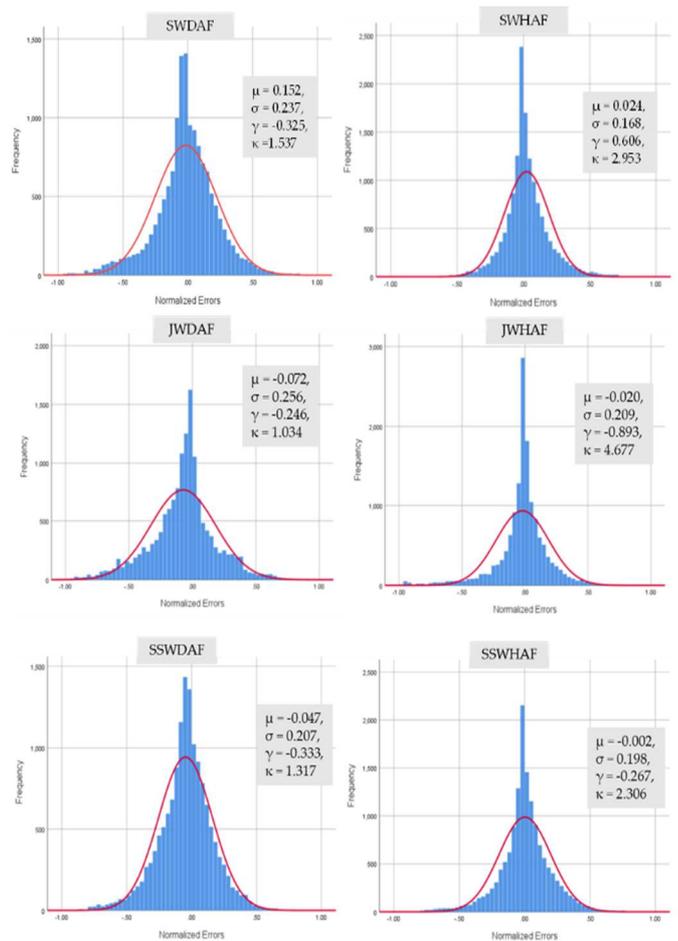


Fig. 3 PDF Plot Histogram of all Scenarios

Based on the skewness indicator, only SWHAF has a positive value, while other scenarios negatively value. It means that only SWHAF is indicated under-estimated, while five other scenarios are over-estimated. SWHAF and JWHAf are moderately skewed, while the other four scenarios are approximately symmetric. JWDAF has the smallest skewness, while JWHAf has the largest skewness. It shows Jenepono WPP has extreme characteristics in terms of skewness. Based on kurtosis indicators, all scenarios are leptokurtic. Similar to the nature of skewness, Jenepono WPP has the most extreme values. It is indicated by the JWDAF kurtosis value, which has the smallest kurtosis value, while JWHAf has the most significant kurtosis value.

Figure 4 shows the QQ Plot of normalized data. The blue line represents the normalized errors of the observed data. The red line represents QQ Plot of a normal distribution with the same average and standard deviation as observed errors. QQ Plot is used to perceive the trend of the data distribution to the Normal Distribution.

Among all scenarios, it appears that JWDAF has the closest trend, while JWHAf has the most distant tendency. It is seen that Jenepono WPP has the most extreme value compared to other entities. Judging from the balance, SWDAF, JWDAF, JWHAf, SSWDAF, and SSWHAF are left QQ Plots. Only SWHAF has the right QQ plot. In general, from the QQ Plot perspective, all scenarios are still close to Normal Distribution trends with varying degrees of distribution.

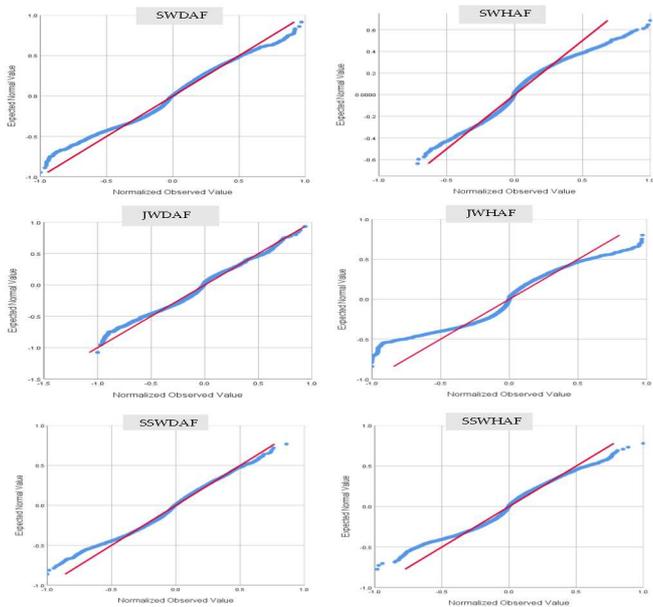


Fig. 4 Quantile-Quantile (QQ) Plot of all Scenarios

Figure 5 shows the CDF Plot of normalized data. The blue line represents the normalized errors of the observed data. The red line represents CDF Plot of a Normal Distribution with the same average and standard deviation as observed errors. Similar to QQ Plot, CDF Plot can also be used to check the distribution of datasets towards normal distribution. Figure 6 shows that, based on the CDF plot, all scenarios are close to normal distribution. The most significant deviation is at JWHAF, while the smallest deviation is at SSWDAF.

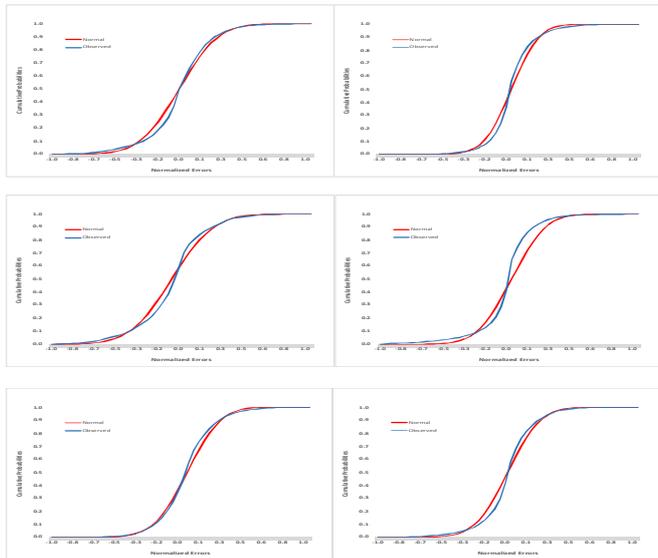


Fig. 5 Cumulative Distribution Function (CDF) Plot of all Scenarios

C. Result of Stage 3

In this stage, we use normalized metrics to analyze the performance of the six scenarios that have been defined in the previous section. Figure 6 shows the error performance of the day-ahead data. It is depicted that the NBIAS values of WPF data for all entities are negative. It means that the day-ahead forecast results are over-estimated. Among all entities, the SWDAF has the smallest systematic errors, and the JWDAF has the biggest. However, it is not in line with the NMAE

result. The SSWDAF has the best performance in terms of absolute systematic errors.

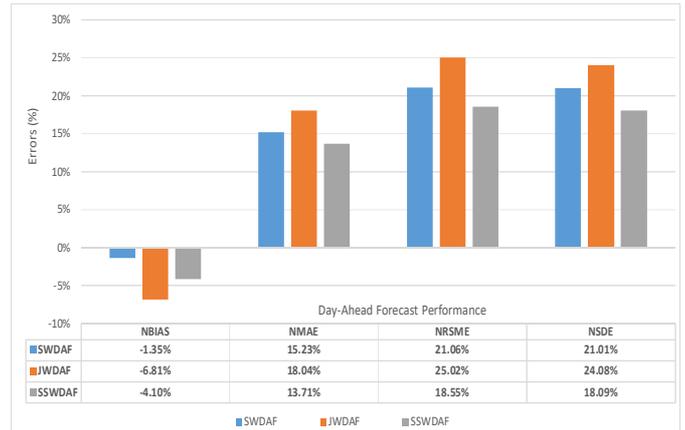


Fig. 6 Day-Ahead Forecast Performance for WPPs in Southern Sulawesi Power System

This means that the systematic errors of SWDAF have an almost balanced accumulation of over-estimated and under-estimated forecast but has more prominent accumulation in the NMAE parameter. The South Sulawesi WPP, as the aggregation of Sidrap WPP and Jeneponto WPP, has the advantage of reducing the errors. The MAE result of the SSWDAF is not a summation of the SWDAF and JWDAF values. It is even smaller than the average value of the two scenarios. Similar situations occur for large errors. The RSME results show that the error performance of the aggregation WPP has relatively better error accuracy than the individual ones. The variance of errors also indicates that the aggregation WPP produces a relatively narrower spread of errors. Overall, it is depicted in Figure 6 that Jeneponto WPP has the worst error indications for day-ahead data.

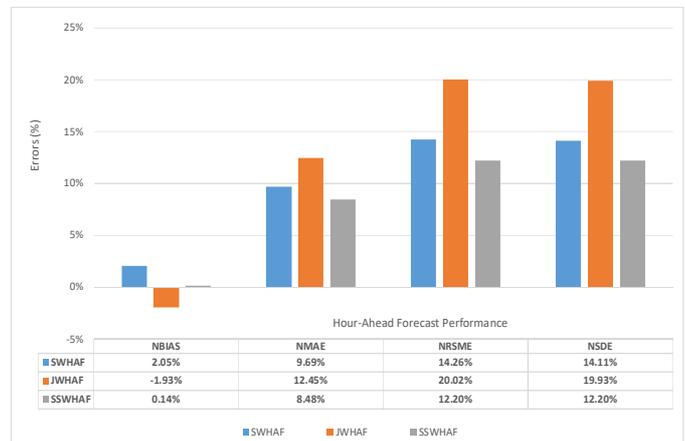


Fig. 7 Hour-Ahead Forecast Performance for WPPs in Southern Sulawesi Power System

The error performance of the hour-ahead dataset is depicted in Figure 7. It shows that the SWHAF has a positive NBIAS value in terms of systematic errors, but JWHAF has the negative one. It means that for the hour-ahead dataset, the SWHAF is under-estimated, but JWHAF is over-estimated. On the other hand, the South Sulawesi WPP has a positive value close to zero, which balances the total forecast results. Like day-ahead data, the SSWHAF has the best performance

for small, large, and variance errors in hour-ahead data. It is shown by the NMAE, NRSME, and NSDE results seen in Figure 7. Similar to day-ahead data, the JWDAF has the worst performance among all.

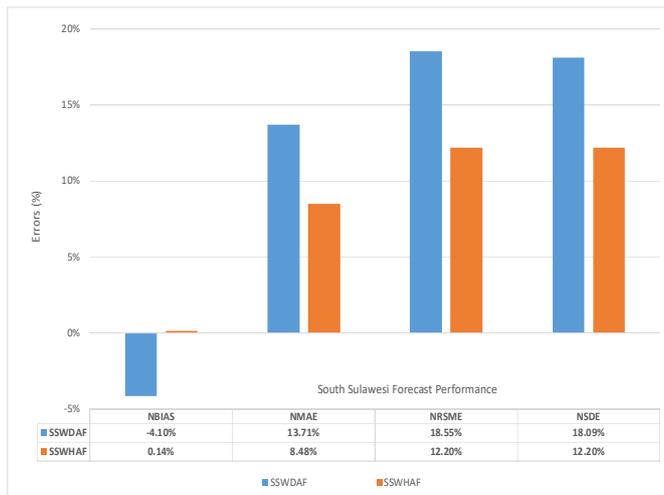


Fig. 8 Comparison of Time-Based Forecast Performance for South Sulawesi WPP

Furthermore, Figure 8 shows the comparative performance of day-ahead and hour-ahead data for South Sulawesi WPP. With the change in the forecast's time scale from day-ahead to hour-ahead, it has an error rate reduction of 96.5% for NBIAS, 38.09% for NMAE, and 34.21% NRSME, and 32.55% for NSDE. It is noticed that the hour-ahead forecast has a better result for all error criteria than the day-ahead one. With almost zero systematic errors, better error accuracy for small and large errors and dispersion of error range shows that the hour-ahead forecast has stable improvement compared to the day-ahead one. It means that, in all criteria, the hour-ahead forecast has better accuracy than the one made at the day-ahead time scale. It might be due to the decrease in uncertainty that arises between the day-ahead forecast to hour-ahead one.

D. Result of Stage 4

In this case, the score weighting method does not include the parameters of skewness and kurtosis. The two parameters do not specifically refer to the quantity of accuracy but rather to the symmetry and peak height, and a tail length of the existing distribution. Table 5 shows the weighting scores of all existing scenarios for the results of the three previous stages' overall parameters.

TABLE V
WEIGHTING SCORES OF VARIOUS SCENARIOS AND PARAMETERS

Parameter	SW	SW	JW	JW	SSW	SSW
	DAF (%)	HAF (%)	DAF (%)	HAF (%)	DAF (%)	HAF (%)
R-Squared	57.04	79.99	54.83	68.50	66.92	84.50
NBIAS	98.65	97.95	93.19	98.07	95.90	99.86
NMAE	84.77	90.31	81.96	87.55	86.29	91.52
NRSME	78.94	85.74	74.98	79.98	81.45	87.80
NSDE	78.99	85.89	75.92	80.07	81.91	87.80
Mean	84.80	97.60	92.80	98.00	95.30	99.80
SD	76.30	83.20	74.40	79.10	79.30	80.20
Scores	79.93	88.67	78.30	84.47	83.87	90.21

The final scores are added in the last row of this table. They are calculated from the average value of all available parameters. With the same standard of weighting score, SSWHAF has the highest value for almost all parameters, while JWDAF has the lowest value. Only on the Mean parameter, SWDAF has the lowest value, but the highest value remains in the SSWHA scenario.

From the final score, there is a similar pattern in all scenarios based on the timescale. Hour-Ahead Forecast has a higher score than Day-Ahead Forecast for all entities. It shows that the forecast's accuracy will be higher if done at a time closer to the target time. It also appears that, based on the final scores, JWDAF has the lowest score of 78.30 while the highest score on SSWHAF with 90.21.

As seen in Figure 9, the bold colored lines of the radar chart represent the interrelationships of the statistical parameters, while the gray spider web represents the standard score for each parameter. It is expressed as a percentage scale for each parameter. It shows similar characteristics from a different perspective for each scenario represented by the spider web. With an apparent visual description, each scenario's characteristics can be seen obviously by various bold colors of the spider web on the radar chart. JWDA, with a bold gray line, has the most inner loop spider web. In line with Table 5, this refers to the lowest final score. Whereas SSWHA, represented by a bold green line, is in the most outer loop. This condition refers to the highest final score.

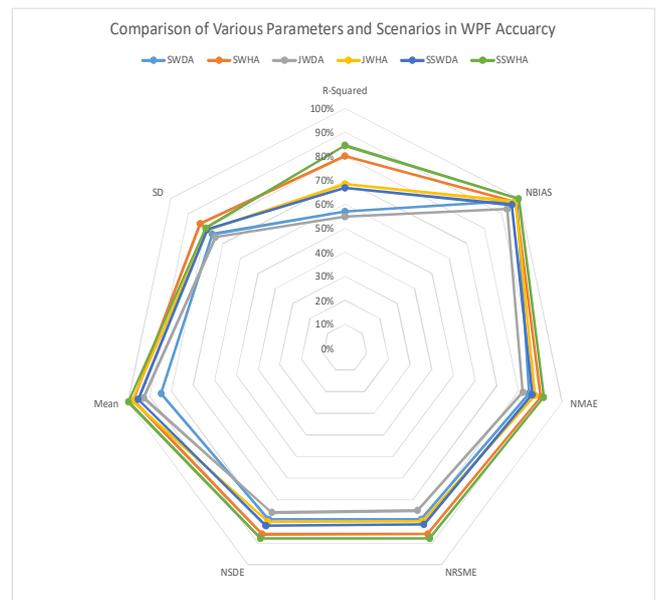


Fig. 9 Radar Chat of Various Parameters and Scenarios

E. Discussion

The previous section has described in detail all the stages that exist in the multi-stage statistical approach method. Stage 1 describes the random spreading between forecast and actual data for each predefined scenario. A similar method was also used partly in some previous studies, although with marginally different parameters [23]-[25], [34]. The result of this stage (Figure 2) is the initial description of existing datasets dispersion. It may represent a depiction of the WPF accuracy of the observed data. Additional information in the form of R-Squared (R^2), a trendline of data distribution,

provides a more quantitative perspective of each predefined scenario.

Stage 2 analyzes the data that was prepared beforehand in stage 1. At this stage, the data is processed more comprehensively by taking the normal distribution as a reference. From some previous studies, Normal Distribution is often used as an initial approach in modeling WPF Accuracy from existing data [17]-[20]. It aims to facilitate modeling specifically for the stage of wind integration planning studies [13]. Using PDF plot, QQ plot, and CDF plot (Figure 3 to 5), additional information is obtained about the dataset's characteristics being reviewed. The used parameters such as Mean, Standard Deviation, Skewness, and Kurtosis were also used separately in previous studies [20]. The use of QQ Plots and CDF Plots reinforces the picture of the results for various scenarios.

Stage 3 explores additional characteristics of the dataset from different statistical perspectives compares to the previous stages. NBIAS, NMAE, NRSME, and NSDE give a new statistical approach that focused on the observed data's accuracy (Figure 6 to 8). This stage results are in line with those in the previous two stages for all observed scenarios. These parameters were also used as a single method in other studies [21], [22], [33].

Stage 4 is the final process of this study. Radar charts are a new style used to evaluate WPF accuracy characteristics that have not been found in any previous studies. Using this method is based on the desire to comprehensively explore the characteristics of the dataset with a multi-stage process. The radar chart (Figure 9) succeeds in summarizing the characteristics obtained from the previous stages. Using the weighting score as a standard assessment, the final score of all scenarios can be attained. Something interesting about this multi-stage process is that all the series methods provide harmonious results even though using a different approach. So, there is no contradiction, and the accumulative results tend to be mutually reinforcing. Overall, this breakthrough method provides a more comprehensive with a different way of solution that may support future expansion studies.

The dataset is grouped into six scenarios based on location and timescale combination to produce a more systematic analysis regarding the case study. By using a balanced scenario concept, we can find out more detailed and comparable characteristics. It is seen that individually, Sidrap WPP (SWDAF and SWHAF) has better WPF accuracy than Jeneponto WPP (JWDAF and JWDAF), both for Day-Ahead and Hour-Ahead Forecast. It might be due to Sidrap WPP has more statistical data than Jeneponto WPP since it has been operating for one year longer. But the aggregation of the two WPP (SSWDAF and SSWHAF) has the highest accuracy of WPF. The more significant dispersion of the WPP locations, the smaller the aggregation result's dispersion. It is related to variations in geographical conditions so that many WPP operate in different weather conditions simultaneously. The outcomes of this research prove that the nature of aggregation can improve the accuracy of WPF. This situation is illustrated by almost all statistical parameters used in this study.

Next, we return to the research question written in the Introduction section: *How to optimize statistical approach tools in evaluating the characteristics of WPF, primarily when related to the aspects of aggregation and timescales?*

Based on the results and discussion described above, it is seen that there are a lot of significant benefits after applying this method. With a multi-stage process, WPF accuracy characteristics are comprehensively described based on location and timescale combination. It covers both quantitative and qualitative aspects through the existence of schemes, tables, and graphs. This result will be beneficial as a reference in future WPP expansion studies.

To get a more comprehensive picture of WPF characteristics, this case study's development has to include the observed areas' climate factors. Since the case occurs in South Sulawesi, the tropics' seasonal climate will become a significant factor in conducting expansion studies. Based on the literature, there are not so many studies that have been done in tropical areas. Lledo *et al.* [35] conducted seasonal forecasts of wind power generation in Europe, and Arslan *et al.* [36] examined modeling wind power potential on seasonal timescales with impact surfaces in Turkey. Therefore, it is widely open to this study's continuation, especially the impact of a seasonal factor on the WPF characteristics in the tropical climate area.

IV. CONCLUSION

This paper has demonstrated a multi-stage statistical approach to evaluate the WPF in the Southern Sulawesi power system. The research question in the Introduction section has been answered clearly since this innovative method is conclusively proven to give a more comprehensive description of WPF profiles in the observed power system that can be used as a basis for conducting future expansion studies.

The multi-stage method algorithm consists of four stages with different perspectives but have inline results and mutually reinforcing. It proves that the selection of methods in the process algorithm has been appropriate and effective since it provides more accurate and comprehensive results.

Further work to be done is to include the impact of seasonal climate factors on the case study's WPF evaluation. It will enrich the references for wind expansion studies, especially in tropical climate areas. In closing, as the pioneer and the only large-scale current WPP operations in Indonesia, the evaluation result of WPF characteristics in the Southern Sulawesi power system will significantly determine WPP development's direction in other major power systems in Indonesia.

ACKNOWLEDGMENT

The authors are obliged to Suroso Isnandar, Nurdin Pabi, Titi Yustiana, Goklas Batapanmuda, and Tommy Kaendo from PLN UIKL Sulawesi for providing the operational data. The authors are also grateful to Universitas Indonesia (UI) to fund this research through PUTI Q2 grant 2020 launched by DRPM UI.

REFERENCES

- [1] Y. Hao and C. Tian, "A novel two-stage forecasting model based on error factor and ensemble method for multi-step wind power forecasting," *Applied Energy*, vol. 238, pp. 368–383, Mar. 2019.
- [2] Y. Wang, Q. Hu, D. Srinivasan, and Z. Wang, "Wind power curve modeling and wind power forecasting with inconsistent data," *IEEE*

- Transactions on Sustainable Energy*, vol. 10, no. 1, pp. 16-25, Jan. 2019.
- [3] Y. Kim and J. Hur, "An ensemble forecasting model of wind power outputs based on improved statistical approaches," *Energies*, vol. 13, pp. 1071-1086, Mar. 2020.
- [4] J. Yan, H. Zhang, Y. Liu, S. Han, L. Li, and Z. Lu, "Forecasting the high penetration of wind power on multiple scales using multi-to-multi mapping," *IEEE Transactions on Power Systems*, vol. 33, pp. 3276-3284, May 2018.
- [5] M. B. Ozkan and P. Karagoz, "Data mining-based upscaling approach for regional wind power forecasting: regional statistical hybrid wind power forecast technique (regional SHWIP)," *IEEE Access*, vol. 7, pp. 171790-171800, 2019.
- [6] V. Hoolohan, A. S. Tomlin, and T. Cockerill, "Improved near surface wind speed predictions using Gaussian process regression combined with numerical weather predictions and observed meteorological data," *Renewable Energy*, vol. 126, pp. 1043-1054, Oct. 2018.
- [7] L. Zhang, Y. Dong and J. Wang, "Wind speed forecasting using a two-stage forecasting system with an error correcting and nonlinear ensemble strategy," *IEEE Access*, vol. 7, pp. 176000-176023, 2019.
- [8] L. Han, M. Li, X. Wang, and Y. Cheng, "Real-time wind power forecast error estimation based on eigenvalue extraction by dictionary learning," *Chinese Journal of Electronics*, vol. 28, pp. 349-356, Mar. 2019.
- [9] T. Ji, D. Hong, J. Zheng, Q. Wu and X. Yang, "Wind power forecast with error feedback and its economic benefit in power system dispatch," *IET Generation, Transmission & Distribution*, vol. 12, pp. 5730-5738, Nov. 2018.
- [10] L. Han, C. E. Romero, X. Wang, and L. Shi, "Economic dispatch considering the wind power forecast error," *IET Generation, Transmission & Distribution*, vol. 12, no. 12, pp. 2861-2870, Jul. 2018.
- [11] L. Han, R. Zhang, X. Wang, and Y. Dong, "Multi-time scale rolling economic dispatch for wind/storage power system based on forecast error feature extraction," *Energies*, vol. 11, pp. 2124-2140, Aug. 2018.
- [12] X. Yu, X. Dong, S. Pang, L. Zhou, and H. Zang, "Energy storage sizing optimization and sensitivity analysis based on wind power forecast error compensation," *Energies*, vol. 12, pp. 4755-4771, Dec. 2019.
- [13] A. Banik, C. Behera, T. V. Sarathkumar and A. K. Goswami, "Uncertain wind power forecasting using LSTM-based prediction interval," *IET Renewable Power Generation*, vol. 14, no. 14, pp. 2657-2667, Oct. 2020.
- [14] C. Ning and F. You, "Data-Driven Adaptive Robust Unit Commitment Under Wind Power Uncertainty: A Bayesian Nonparametric Approach," *IEEE Transactions on Power Systems*, vol. 34, no. 3, pp. 2409-2418, May 2019.
- [15] W. Xie, P. Zhang, R. Chen, and Z. Zhou, "A Nonparametric Bayesian Framework for Short-Term Wind Power Probabilistic Forecast," vol. 34, no. 1, pp. 371-379, Jan. 2019.
- [16] E. Oh and H. Wang, "Reinforcement-learning-based energy storage system operation strategies to manage wind power forecast uncertainty," *IEEE Access*, vol. 8, pp. 20965-20976, 2020.
- [17] J. Y. Shin, C. Jeong, and J. H. Heo, "A novel statistical method to temporally downscale wind speed Weibull distribution using scaling property," *Energies*, vol. 11, pp. 633-647, 2018.
- [18] M. B. H. Kumar, S. Balasubramanian, P. Sanjeevikumar, and J. B. H. Nielsen, "Wind energy potential assessment by Weibull parameter estimation using multiverse optimization method: A case study of Tirumala region in India," *Energies*, vol. 12, pp. 2158-2172, 2019.
- [19] C. Wang et al., "Optimal sizing of energy storage considering the spatial-temporal correlation of wind power forecast errors," *IET Renewable Power Generation*, vol. 13, no. 4, pp. 530-538, Mar. 2019.
- [20] F. Ge, Y. Ju, Z. Qi, and Y. Lin, "Parameter estimation of a Gaussian mixture model for wind power forecast error by Riemann L-BFGS optimization," *IEEE Access*, vol. 6, pp. 38892-38899, 2018.
- [21] J. J. Miettinen and H. Holttinen, "Characteristics of day-ahead wind power forecast errors in Nordic countries and benefits of Aggregation," *Wind Energy*, vol. 20, pp. 959-972, 2017.
- [22] J. Miettinen and H. Holttinen, "Impacts of wind power forecast errors on the real-time balancing need: a Nordic case study," *IET Renewable Power Generation*, vol. 13, pp. 227-233, 2019.
- [23] L. M. Putranto, Sarjiya, and A. N. Widiastuti, "Optimizing the preventive and corrective control scheme in integrated variable renewable energy generation," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 8, no. 5, pp. 2031-2038, 2018.
- [24] A. A. Kadhem, N. I. A. Wahab, I. Aris, J. Jasni, A. N. Abdalla and Y. Matsukawa, "Reliability assessment of generating systems with wind power penetration via BPSO," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 7, no. 4, 2017.
- [25] M. A. M. Idris, M. R. Hao, and M. A. Ahmad, "A data driven approach to wind plant control using moth-flame optimization (MFO) algorithm," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 9, no.1, pp. 18-23, 2019.
- [26] E. Spiliotis, F. Petropoulos, K. Nikolopoulos, "The impact of imperfect weather forecasts on wind power forecasting performance: evidence from two wind farms in Greece," *Energies*, vol. 13, pp. 1880-1895, 2020.
- [27] S. Sun, J. Fu, and A. Li, "A compound wind power forecasting strategy based on clustering, two-stage decomposition, parameter optimization, and optimal combination of multiple machine learning approaches," *Energies*, vol. 12, pp. 3586-3599, Sep. 2019.
- [28] D. H. Barus and R. Dalimi, "The integration of wind power plants to southern Sulawesi power system," in Proc. *ICECCE'20*, 2020, paper 290. p. 643.
- [29] "LODSSPS 2019 data sheet," SSLDC, Makassar, Indonesia.
- [30] M. B. Wilk and R. Gnanadesikan, "Probability plotting methods for the analysis of data," *Biometrika Trust*, vol. 55, no. 1, pp. 1-17, Mar. 1968.
- [31] M. G. Bulmer, *Principles of Statistics*, 3rd ed., K. M. Earnshaw, Ed. New York, USA: Dover Publication, 1979.
- [32] P. H. Westfall, "Kurtosis as peakedness, 1905-2014. R.I.P.," *The American Statistician*, vol. 68, no. 3, pp. 191-195, Jul. 2014.
- [33] H. Madsen, P. Pinson, and G. Kariniotakis, "Standardizing the performance evaluation of short-term wind power prediction models," *Wind Engineering*, vol. 29, no. 6, pp. 475-489, Dec. 2005.
- [34] M. Vahidzadeh, C. D. Markfort, "An induction curve model for prediction of power output of wind turbines in complex conditions," *Energies*, vol. 13, no. 4, pp. 891-906, 2020.
- [35] L. Lledo, V. Torralba, A. Soret, J. Ramon, and F. J. Doblas-Reyes, "Seasonal forecasts of wind power generation," *Renewable Energy*, vol. 143, pp. 91-100, Dec. 2019.
- [36] H. Arslan, H. Baltaci, B. O. Akkoyunlu, S. Karanfil, and M. Tayanc, "Wind speed variability and wind power potential over Turkey: Case studies for Çanakkale and İstanbul," *Renewable Energy*, vol. 145, pp. 1020-1032, Jan. 2020.