Analysis of Wind Speed Distribution at the Mopah Airport in Merauke

Maria Fransina Veronica Ruslau¹
Mathematics Education Department
Universitas Musamus
Merauke, Indonesia
¹maria_fktp@unmus.ac.id

Helga Charolina Antonia Silubun²
Physics Education Department
Universitas Musamus
Merauke, Indonesia
²helga@unmus.ac.id

Abstract—Modeling of wind speed characteristic in a particular location is often conducted, one of which is to find out the potential of wind energy in the interest of wind power plants. Of the many probability distributions, the following distributions are distributions that are quite commonly used to model the characteristic of wind speed data, namely Gamma Distribution, Weibull Distribution, Normal Distribution, Lognormal Distribution, and Rayleigh Distribution. The results of the analysis and test of wind speed data in the Mopah Airport area shows that the Distribution of 3 Parameters-Gamma is the most suitable distribution to model the wind speed characteristic in the region with the shape parameters \( \alpha = 26,585 \), scale parameters \( \beta = 0,2283 \) and location parameters \( \gamma = -0,7061 \). The average wind speed per month ranges from 5.2 to 5.5 knots.

Keywords—fitting distribution; goodness of fit test; wind speed

I. INTRODUCTION

Wind is moving air which occurs because of the temperature difference between hot air and cold air. This is due to differences in air on the earth's surface. Wind data analysis and accurate assessment of wind energy potential are important factors for the development of suitable wind power plants applications in certain locations [1]. Assessment of wind resources available at certain locations and time periods based on the calculated wind power density. To obtain a convincing wind power density, this method requires measurement of wind speed data or based on the wind PDF (Probability Density Function), which is used to calculate the Wind Density Power Distribution (WPDD) [2].

The study of the probability of wind speed is more often conducted for wind energy analysis. However, this paper is not intended for this purpose. In this paper, we will identify the pattern of wind distribution in the Merauke based on the maximum wind speed every month in 2017. A large number of scientific researches have been conducted related to renewable energy which produces various forms of PDF to describe the distribution of wind speeds [3]. Some studies state that wind speed follows the probability distribution pattern of Weibull, Rayleigh, Gamma, Lognormal, Exponential and Gaussian [4], [5]. In this study, wind speed data will be matched with the probability distributions of Gamma, Weibull, Normal, Lognormal and Rayleigh.

The main purpose of this paper is to present the PDF and parameters of the distributions used to model the wind speed characteristic at Mopah Airport in Merauke. The data used is the wind speed measurement data recorded during 2017.

There are several PDFs that can model wind characteristic in a particular place and it is important to choose the right PDF for wind speed [5]. The goodness of fit test is used to select PDF that best represents wind speed characteristic in that location. More than one goodness of fit test is generally used to guarantee the best PDF selection to describe the characteristic of wind speed in a particular place [2]. Zhou, et al. [2] used the test of goodness of fit chi-square (\( \chi^2 \)), coefficient of determination (R2), Root Mean Square Error (RMSE), log likelihood (Log-LH), Kolmogorov-Smirnov (KS), and Index of Agreement (Index_A) to determine which PDF shows the best match with data from the location provided to avoid biased results. In this study, the goodness of fit test used is the Anderson-darling test, the Kolmogorov-Smirnov and RMSE tests.

II. METHOD

The goodness of fit test measures the suitability of random samples with theoretical probability functions. In other words, this test aims to show how well a probability distribution matches the data. This probability is known as the level of trust. The small probability shows that it is very inappropriate while the large probability shows very good compatibility. Fitting distribution is conducted using the easyfit application. Most of the formulas of probability distributions used in this study are taken from the easyfit application website [6].

A. Kolmogorov-Smirnov Test

This test is used to decide whether a sample comes from a continuous distribution based on an empirical cumulative distribution (CDF) function. For example, a random sample \( x_1, \ldots, x_n \) comes from a continuous distribution with CDF \( F(x) \). Empirical CDF is defined as follows:

\[
F_n(x) = \frac{1}{n} \left[ \text{number of sample} \leq x \right]
\]

Definition:
Kolmogorov-smirnov’s (D) statistic is based on the biggest difference between \( F(x) \) and \( F_n(x) \), which is defined by:

\[
D_n = \sup_x |F_n(x) - F(x)|
\]

With the testing hypothesis as follows:
- \( H_0 \): Data follow a certain distribution pattern
- \( H_1 \): Data do not follow a certain distribution pattern

The null hypothesis (\( H_0 \)) is rejected at the level of significance \( \alpha \) determined if the D test statistic is greater than the critical value obtained from the table.

**B. Anderson-Darling Test**

The Anderson-Darling procedure is a general test to compare the suitability of the cumulative distribution function observed with the expected cumulative distribution function.

Anderson-Darling statistics \( (A^2) \) are defined as:

\[
A^2 = -n - \sum_{i=1}^{n} (2i - 1) \left[ \ln F(x_i) + \ln(1 - F(x_{n-i+1})) \right]
\]

With the testing hypothesis as follows:
- \( H_0 \): Data follow a certain distribution pattern
- \( H_1 \): Data do not follow a certain distribution pattern

The hypothesis of the form of distribution is rejected at the chosen level of significance (alpha) if the test statistic is greater than the critical value obtained from the table.

**C. Root Mean Square Error**

The formulation to calculate RMSE [7] is:

\[
RMSE = \left( \frac{1}{N} \sum_{i=1}^{n} \left( F_n(x_i) - \hat{F}(x_i) \right)^2 \right)^{\frac{1}{2}}
\]

Where \( F_n(x_i) \) is the data relative frequency and \( \hat{F}(x_i) \) is the value of probability distribution. The smallest RMSE value indicates that the distribution is good or suitable for modeling wind speed.

**D. Weibull Distribution**

The weibull distribution was introduced by a physicist named Walodi Weibull in 1939. The recent distribution of weibull is widely used in the probability distribution of wind speed [4]. In the Weibull function, variations in wind speed are described in the shape parameters and scale parameters.

Weibull Distribution Density Function (2 Parameters) is defined as:

\[
f(x) = \frac{\alpha}{\beta} \left( \frac{x}{\beta} \right)^{\alpha-1} \exp \left( -\left( \frac{x}{\beta} \right)^\alpha \right)
\]

Whereas, the Cumulative Distribution Function of Weibull Distribution (2 Parameters) is defined as:

\[
F(x) = 1 - \exp \left( -\left( \frac{x}{\beta} \right)^\alpha \right)
\]

Weibull Distribution Density Function (3 Parameters) is defined as:

\[
f(x) = \frac{\alpha}{\beta} \left( \frac{x}{\beta} \right)^{\gamma-1} \exp \left( -\left( \frac{x}{\beta} \right)^\gamma \right)
\]

The Weibull Cumulative Distribution Function (3 Parameters) is defined as:

\[
F(x) = 1 - \exp \left( -\left( \frac{x}{\beta} \right)^\gamma \right)
\]

Where \( x \) is the wind speed, \( \alpha \) is the shape parameters \( (\alpha > 0) \), \( \beta \) is the scale parameters and \( \gamma \) is the location parameters \( (\gamma \leq x < +\infty) \). \( \gamma = 0 \) for the 2 parameters weibull distribution.

**E. Rayleigh Distribution**

In the field of physics, rayleigh distribution is often used for modeling the radiation process of sound and light, wave height and wind speed. The wind probability function and the gust of time in one year can be approached with this distribution. When this distribution is graphically presented, the area under the curve is equal to 1. Rayleigh distribution is a special case of the Weibull distribution with \( (\alpha = 2) \) shape parameters [8].

The Rayleigh Distribution Density Function is defined as follows:

\[
f(x) = \frac{\pi v}{2 Vm^2} \exp \left( -\frac{\pi v}{4Vm^2} \right)
\]

Whereas, the Cumulative Distribution Function of the Rayleigh Distribution is defined as:

\[
F(x) = 1 - \exp \left( -\frac{\pi v}{4Vm^2} \right)
\]

The Rayleigh Distribution Density Function can be written in the form of an equation that only depends on the scale parameters where the scale parameters is always equal to two.

\[
f(x) = \frac{2v}{c} \exp \left( -\frac{v^2}{2c^2} \right)
\]

With

\[
c = \frac{2}{\pi} \left( \frac{Vm}{c} \right)
\]

**F. Gamma Distribution**

Gamma distribution is one of continuous probability distributions that can be used to solve many engineering and scientific subjects. One of the data that can be applied using
gamma distribution is wind speed data. Gamma distribution has the following probability density functions:

\[ f(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} \]  

(13)

Gamma Cumulative Distribution Function is defined as:

\[ F(x) = \frac{\Gamma(x/\beta, \alpha)}{\Gamma(\alpha)} \]  

(14)

Whereas, the Probability Distribution Function of 3 Parameters-Gamma Distribution is defined as:

\[ f(x) = \frac{(x-\gamma)^{\alpha-1}}{\beta^\alpha \Gamma(\alpha)} e^{-\frac{x-\gamma}{\beta}} \]  

(15)

And the Cumulative Probability Distribution Function of 3 Parameters-Gamma Distribution is defined as:

\[ F(x) = \frac{\Gamma(x/\beta, \alpha)}{\Gamma(\alpha)} \]  

(16)

Where \( \Gamma \) is the gamma function. \( \alpha (\alpha > 0) \) is a shape parameters, \( \beta (\beta > 0) \) is a scale parameters and \( \gamma \) is a location parameters (\( \gamma \leq x < \infty \)).

**G. Lognormal Distribution**

Lognormal distribution is one of the probability distributions that are widely applied in various fields such as agriculture, economics, geology, industry and quality control. The probability density function for the 3 parameters-lognormal distribution is as follows [9]:

\[ f(x) = \frac{1}{(x-\gamma)\sigma\sqrt{2\pi}} \exp \left[ -\frac{\ln \left( \frac{x-\gamma}{\mu} \right)^2}{2\sigma^2} \right] \]  

(17)

Dimana, \( x \geq 0 \) is wind speed (knots)  
\( \sigma > 0 \) is shape parameters  
\( \mu > 0 \) is scale parameters  
\( \gamma > 0 \) is location parameters

The cumulative distribution function of the 3 parameters-Lognormal Distribution is:

\[ F(x) = \Phi \left( \frac{\ln \left( \frac{x-\gamma}{\mu} \right)}{\sigma} \right) \]  

(18)

If the location parameters \( \gamma = 0 \) then 2 parameters-lognormal distribution with a probability density function as follows:

\[ f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp \left[ -\frac{\ln \left( \frac{x}{\mu} \right)^2}{2\sigma^2} \right] \]  

(19)

The cumulative distribution function of the Lognormal distribution (2 Parameters) is defined as:

\[ F(x) = \Phi \left( \frac{\ln \left( \frac{x}{\mu} \right)}{\sigma} \right) \]  

(20)

Where, \( \Phi \) is a Laplace integral:

\[ \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{0}^{x} e^{-\frac{t^2}{2}} dt \]  

(21)

**H. Normal Distribution**

Normal Distribution is the probability distribution that is most widely used in various statistical analyzes. The normal distribution probability density function is presented as follows:

\[ f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left( -\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2} \right) \]  

(22)

And the cumulative distribution function is defined as:

\[ F(x) = \Phi \left( \frac{x-\mu}{\sigma} \right) \]  

(23)

Where, \( \sigma \) is the scale parameters (\( \sigma > 0 \)), and \( \mu \) is the location parameters.

**III. FINDING AND DISCUSSION**

Wind speeds that occur in Merauke's Mopah Airport area vary from time to time. Calculation of the frequency of wind speed becomes quite important, especially if the highest frequency of occurrence of a particular wind speed is different from the average wind speed. The description of wind speeds recorded during 2017 on BMKG is presented as follows:

Table.1 shows that minimum wind speeds occur in April and maximum wind speeds occur in January and February. However, on average, at a 99% significance level, the average wind speed per month ranges from 5.2 to 5.5 knots. Table.1 can be presented in graphical form as shown in Figure 1. It appears that the minimum wind speed during 2017 is quite volatile. While the maximum wind speed is quite stable, especially during April to September.

Monthly wind speed data is then presented in several intervals. The division of class intervals is determined by the number of wind speed distributions. From the analysis data, it is obtained the distribution of wind speed during 2017 is as presented in Figure 2. Furthermore, the pattern and characteristics of wind speed distribution will be identified in the Mopah Airport area during 2017.
Table I. Average Wind Speed in MOPAH BMKG Station 2017

| Month     | Vmin | Vmax | Vaverage |
|-----------|------|------|----------|
| January   | 3    | 10   | 6        |
| February  | 3    | 10   | 5        |
| March     | 3    | 7    | 5        |
| April     | 2    | 8    | 4        |
| May       | 4    | 8    | 5        |
| June      | 5    | 8    | 6        |
| July      | 4    | 8    | 5        |
| August    | 4    | 8    | 6        |
| September | 5    | 8    | 6        |
| October   | 4    | 7    | 5        |
| November  | 4    | 7    | 6        |
| December  | 3    | 7    | 5        |

Fig. 1. Wind Speed in 2017

Several distributions are used as a wind speed distribution approach from the existing data, including: Weibull Distribution, Rayleigh Distribution, Gamma Distribution, Lognormal Distribution and Normal Distribution.

Fig. 2. Wind Speed Distribution in 2017

The distribution of wind speed data as presented in Figure 2 is then matched with the selected probability distributions. The results of the approach or fitting of probability distribution curves as mentioned above with the wind speed distribution obtained from the wind speed data for 2017 are presented as follows:

Fig. 3. 3 Parameters-Gamma Distribution

Fig. 4. 3 Parameters-Lognormal Distribution

Fig. 5. 3 Parameters-Weibull Distribution

Fig. 6. Normal Distribution

Fig. 7. 2 Parameters-Rayleigh Distribution

Fig. 8. Gamma Distribution
A. Curve Fitting with Kolmogorov-Smirnov Statistical Test

The results of fitting wind speed data with probability distribution using the Kolmogorov-Smirnov statistical test are presented in Table II. The significance level used is $\alpha = 0.01$. The goodness of fit based on the Kolmogorov-Smirnov test statistic shows that of all probability distributions, no distribution is exactly the same as the data pattern being tested. However, it can be seen that the probability distribution that is closest to represent the characteristic of wind speed data tested is 3 Parameters-Gamma Distribution.

| Rank | Distribution          | Kolmogorov-Smirnov Statistic | Critical Value | Decision       |
|------|-----------------------|------------------------------|----------------|----------------|
| 1    | 3 Parameters- Gamma   | 0.1753                       |                | Ho Rejected    |
| 2    | 3 Parameters- Lognormal | 0.17644                     | 0.0855         | Ho Rejected    |
| 3    | 3 Parameters- Weibull  | 0.17885                      |                | Ho Rejected    |
| 4    | Gamma                 | 0.17925                      |                | Ho Rejected    |
| 5    | Normal                | 0.19632                      |                | Ho Rejected    |
| 6    | 2 Parameters- Rayleigh | 0.28212                     |                | Ho Rejected    |

B. Curve Fitting with Anderson-Darling Statistical Test

Table III shows the results of wind speed data fitting using the Anderson-Darling statistical test at a $\alpha = 0.01$ significance level. Unlike the results of the Kolmogorov-Smirnov statistical test, the results of the goodness of fit with Kolmogorov-Smirnov obtained Gamma distribution as a probability distribution that closest to the characteristic of the data being tested. The second rank is Parameters Gamma-3 distribution.

| Rank | Distribution          | Anderson Darling Statistic | Critical value | Decision       |
|------|-----------------------|-----------------------------|----------------|----------------|
| 1    | Gamma                 | 11.717                      |                | Ho Rejected    |
| 2    | 3 Parameters- Gamma   | 11.717                      | 3.9047         | Ho Rejected    |
| 3    | 3 Parameters- Lognormal | 11.741                      |                | Ho Rejected    |
| 4    | 3 Parameters- Weibull  | 11.992                      |                | Ho Rejected    |
| 5    | Normal                | 12.185                      |                | Ho Rejected    |
| 6    | 2 Parameters- Rayleigh | 27.407                      |                | Ho Rejected    |

C. Curve Fitting with RMSE

The goodness of fit with Kolmogorov-Smirnov and Anderson-Darling statistics gave different results. We can choose the parameters Gamma-3 distribution as the probability distribution that is closest to represent the characteristic of the data, but in this study to determine the best distribution, the test continued by looking at the RMSE value.

| Distribution | Parameters | RMSE  |
|--------------|------------|-------|
| Gamma        | $\alpha=20.5$ $\beta=0.26163 \quad \gamma=0$ | 0.0201 |
| 3 Parameters - Gamma | $\alpha=26.585$ $\beta=0.2283 \quad \gamma=0.7061$ | 0.0073 |
| 3 Parameters - Lognormal | $\alpha=0.1391 \quad \mu=2.1219 \quad \gamma=-3.0645$ | 0.0248 |
| 3 Parameters - Weibull | $\alpha=3.1351 \quad \mu=5.926 \quad \gamma=1.8414$ | 0.0091 |
| Normal       | $\alpha=1.1846 \quad \mu=5.3636 \quad \gamma=1.9854$ | 0.0213 |
| 2 Parameters - Rayleigh | $\alpha=2.531 \quad \mu=5.3636 \quad \gamma=1.9854$ | 0.0195 |

Based on Table IV, it can be seen that the probability distribution that has the smallest RMSE value is 3 Parameters-Gamma Distribution with shape parameters $\alpha = 26.585$, scale parameters $\beta = 0.2283$ and location parameters $\gamma = -0.7061$.

IV. CONCLUSION

Based on the results of the analysis by comparing the results of the goodness of fit test from Kolmogorov-Smirnov Statistics, Anderson-Darling did not have a single distribution that exactly matched the data distribution of the wind speed tested. Looking at the RMSE value, it can be concluded that 3 Parameters-Gamma Distribution are the closest or the most appropriate probability distributions to represent the characteristic of wind speed data compared to Weibull, Normal, Lognormal and Rayleigh distributions. 3 Parameters-Gamma Distribution can be used as an alternative distribution to model wind speed distribution in the area of Mopah Airport in Merauke.

ACKNOWLEDGMENT

We thank Universitas Musamus for facilities, supports, and publications.

REFERENCES

[1] D. K. Kidno, R. Danwe, S. Y. Doka, N. Djongyang, Y. S. Doka, and N. Djongyang, “Statistical analysis of wind speed distribution based on six Weibull Methods for wind power evaluation in Garoua , Cameroon,” Rev. des Energies Renouvelables, vol. 18, pp. 105–125, 2015.
[2] P. A. Amaya-Martínez, A. J. Saavedra-Montes, and E. I. Arango-Zuluaga, “A statistical analysis of wind speed distribution models in the Aburrá Valley, Colombia,” CTyF - Ciencia, Tecnol. y Futur., vol. 5, no. 5, pp. 121–136, 2014.
[3] S. Carta, J.A; Ramírez, P; Velázquez, “A review of wind speed probability distributions used in wind energy analysis: Case studies in the Canary Islands,” Renewable and Sustainable Energy Reviews, .
[4] W. Widiyanto, “Analisis Probabilitas Kecepatan Angin untuk Pesisir Cilacap dengan Menerapkan Distribusi Weibull dan Rayleigh Probability Analysis of Wind Speed for Cilacap Coast by Applying Weibull and Rayleigh Distribution,” Din. Rekayasa, vol. 9, no. 1, 2013.
[5] Z. Liu, Sanming; Wang, “Probabilistic model of Wind Power Ramp Events Based On Moving Ratio Analysis,” no. Icep, pp. 126–133, 2016.
[6] M. Technologies, “Introduction to EasyFit 5.5.” [Online]. Available: http://www.mathwave.com/help/easyfit/html/menus/analyze/fit_distributions.html.

[7] I. Pobočíková, Z. Sedliačková, and M. Michalková, “Application of Four Probability Distributions for Wind Speed Modeling,” *Procedia Eng.*, vol. 192, pp. 713–718, 2017.

[8] Y. Arikan, Ö. P. Arslan, and E. Çam, “The analysis of wind data with rayleigh distribution and optimum turbine and cost analysis in Elmadag, Turkey,” *Istanbul Univ. - J. Electr. Electron. Eng.*, vol. 15, no. 1, pp. 1907–1912, 2015.

[9] A. Zaharim, “Fitting of Statistical Distributions to Wind Speed Data in Malaysia,” *Eur. J. Sci. Res.*, vol. 26, no. 1, pp. 6–12, 2009.