Indonesian Journal of Pure and Applied Mathematics
http://journal.uinjkt.ac.id/index.php/inprime
Vol. 1, No. 2 (2019), pp. 110-121
ISSN 2686-5335
doi: 10.15408/inprime.v1i2.13676

Estimation Parameter d in Autoregressive Fractionally Integrated Moving Average Model in Predicting Wind Speed

Devi Ila Octaviyani, Madona Yunita Wijaya and Nina Fitriyati
Department of Mathematics, Faculty of Sciences and Techonlogy
Universitas Islam Negeri Syarif Hidayatullah Jakarta
Email: devi.octaviyani14@gmail.com, {madona, nina.fitriyati}@uinjkt.ac.id

Abstract

Wind speed is one of the most important weather factors in the landing and takeoff process of airplane because it can affect the airplane's lift. Therefore, we need a model to predict the wind speed in an area. In this research, the wind speed forecast using the ARIMA model is discussed which has differencing parameters in the form of fractions. This model is called the ARFIMA model. In estimating differencing parameters two methods are considered, namely parametric and semiparametric methods. Exact Maximum Likelihood (EML) is used under parametric method. Meanwhile, four methods semiparametric estmation are used, i.e Geweke and Porter-Hudak (GPH), Smooth GPH (Sperio), Local Whittle and Rescale Range (R/S). The result shows the best estimation method is GPH with the selected model is ARFIMA (2,0.334,0).
Keywords: ARFIMA, Parametric Method, Semiparametric Method.

Abstrak

Kecepatan angin merupakan salah satu faktor cuaca yang penting dalam proses pendaratan dan tinggal landas pesawat karena dapat mempengaruhi daya angkat pesawat. Oleh karena itu, diperlukan suatu model untuk memprakirakan kecepatan angin di suatu wilayah. Artikel ini membahas prakiraan kecepatan angin dengan menggunakan model ARIMA yang memiliki parameter *differencing* berupa bilangan pecahan. Model ini disebut model ARFIMA. Pada estimasi parameter *differencing* terdapat dua metode yang digunakan pada penelitian ini, yaitu metode parametrik dan metode semiparametrik. Metode parametrik yang digunakan adalah *Exact Maximum Likelihood* (EML) dan empat metode semiparametrik yang digunakan adalah Geweke and Porter-Hudak (GPH), *Smooth* GPH (Sperio), *Local Whittle* dan *Rescale Range* (R/S). Hasil analisis menunjukkan pada kasus ini metode estimasi terbaik adalah GPH dengan model terpilih adalah ARFIMA(2,0.334,0).
Kata kunci: ARFIMA, Metode Parametrik, Metode Semiparametrik.

1. INTRODUCTION

Wind is one of the weather elements that has important role in determining the weather and climate conditions in a particular area. Wind energy benefits can be obtained depending on the wind speed and geographical conditions of an area. Several studies has been conducted to determine the effect of wind speed in various aspects of life and the importance of predicting wind speed in an area such as predicting short-term wind speed to get input for the wind turbin controller [1]. In addition, it is also needed to estimate the wind speed on the airport runway when the plane is going to land and takeoff. Information regarding wind speed on the runaway surface is one of the important factors in the process of aircraft's landing and takeoff as it can affect the aircraft's lift and prevent the aircraft from slipping. Several studies using the ARIMA Box-Jenkins methods have been proposed to predict wind speed, including Ulinnuha [2] and Desvina [3]. In general, the ARIMA(p, d, q) was introduced by Box and Jenkins [4] to model non-stationary time series data. Non-stationary series shows a slow-decaying autocorrelation function (ACF). The order d in ARIMA(p, d, q) is used to model a series that is not stationer in mean, where d represents differencing that takes positive integer numbers. For d that can take any fraction numbers, ARFIMA (Autoregressive Fractional Integrated Moving Average) can be utilized a generalization of ARIMA model [5]. In ARFIMA model, the series has long term dependency properties.

Estimating the appropriate d value will yield a good model fit. Estimation methods for d parameter can be divided into classes, i.e. parametric and semiparametric methods. Parametric method estimates all parameters in ARFIMA model in one step by using parametric approaches. The most commonly parametric method used is Exact Maximum Likelihood (EML) [6]. On the other hand, semiparametric methods is carried out in two steps. The first step is estimating the d value and the second step is estimating the AR and MA parameters. In semiparametric methods, the most commonly used methods are Geweke dan Porter-Hudak [7], Reisen dan Lopes [8], Kunsch [9] dan Robinson [10].

In this study, we predict wind speed at Soekarno-Hatta airport using ARFIMA model where paremeter d is estimated using parametric and semiparametric methods. Geweke dan Porter-Hudak (GPH), *Smooth GPH* (*Sperio*), R/S dan *Local Whittle* are considered for the semiparametrc approaches while EML is considered for the parametric approach. We use wind speed daily data over the period of December 1st, 2017 to November 30th, 2018. It is obtained from the NNDC *Climate Data Online* [11].

2. METHODS

2. 1. Long Memory Process

A time series is said to be a process with long-term memory if the autocorrelation function decays slowly to zero, showing that between far apart observations are still strongly correlated [12]. This condition of long-term memory can be seen from the value of Hurst (H) which can be obtained from the statistic R/S [12]. The Hurst value is determined by computing the mean $\bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t$, adjusted mean $y_t^{adj} = y_t - \bar{y}$, cumulative deviation $y_t^* = \sum_{t=1}^{T} y_t^{adj}$, range of cumulative deviation $R_t = \max(y_1^*, y_2^*, \dots, y_t^*) - \min(y_1^*, y_2^*, \dots, y_t^*)$, and standard deviation $s_t = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(y_t - \bar{y})^2}$ from time series data where $t = 1,2, \dots T$. The value of H can be calculated by the following formula:

$$H = \frac{\log(R/S)_t}{\log(t)}.$$

If the computed H is equal to 0.5 then the series are random, if $0 < H < 0.5$ then the series shows short-term memory, and if $0.5 < H < 1$ then the series shows long-term memory.

2. 2. ARFIMA model

Autoregressive Fractionally Integrated Moving Average (ARFIMA) model is one of the most appropriate model for time series data with long-term memory that has been developed by Granger and Joyeux [9], and also Hoskings [7]. ARFIMA (p, d, q) can be expressed as follows [13]:

$$\phi_p(B)(1-B)^d Y_t = \theta_q(B) e_t \,,$$

where $\{e_{LK'0«««t}\}$ is white noise process, $\phi_p(B)$ is AR polynomial equation of order p, $\theta_q(B)$ is MA polynomial equation of order q, and $(1-B)^d$ is fractional difference operator.

According to Hoskings [7], fractional difference operator on ARFIMA(p, d, q) is a generalization from an infinite binomial series [14]:

$$\nabla^d = (1-B)^d = \sum_{j=1}^{\infty} \binom{d}{j} (-1)^j B^j \,,$$

Where *B* is a backward shift operator, $\Gamma(x)$ is a gamma function, and $\binom{d}{j} = \frac{d!}{(d-j)!j!} = \frac{\Gamma(d+1)}{\Gamma(j+1)\Gamma(d-j+1)}$ is a binomial coefficient. Several characteristic of fractionally integrated series for various values of d are as follow [15]:

a. If $d = 0$, then the process shows autocorrelation function with exponential decay as an ARMA process,
b. If $d \in (0\,, 0.5)$, then the series is correlated with long memory having positive dependency between distant observations denoted by positive autocorrelation and slow-decaying and also have moving average representation of infinite order,
c. If $d \in (-0.5\,, 0)$, then the series is correlated with long memory having negative dependency denoted by negative autocorrelation and slow-decaying and also have autoregressive representation of infinite order,
d. If $|d| \geq 0.5$, maka proses panjang tidak stasioner.

2. 3. Estimation of Fractional Difference Parameter with Parametric Method

Parametric method is able to estimate all parameters in the ARFIMA model in one step [16]. In this study, the parametric method used is Exact Maximum Likelihood (EML) method introduced by Sowell (1992). This method uses the likelihood principal to estimate $d, \phi,$ dan θ in the ARFIMA model. Given the general form of ARFIMA (p, d, q) model as follows:

$$\left(1-\phi_1 B-\phi_2 B^2-\cdots-\phi_p B^p\right)(1-B)^d (Z_t-\mu) = \left(1+\theta_1 B+\theta_2 B^2+\cdots+\theta_q B^q\right) e_t,$$

$$e_t = \frac{\left(1-\phi_1 B-\phi_2 B^2-\cdots-\phi_p B^p\right)(1-B)^d (Z_t-\mu)}{\left(1+\theta_1 B+\theta_2 B^2+\cdots+\theta_q B^q\right)},$$

where $e_t \sim N(0, \sigma^2)$. The probability density function of $e = (e_1, e_2, .., e_n)$ is defined as:

$$P(e|d, \phi, \mu, \theta, \sigma_e^2) = (2\pi\sigma_e^2)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma_e^2}\sum_{t=1}^{n} e_t^2\right],$$

The likelihood function can be written as follows:

$$\ln L(d, \phi, \mu, \theta, \sigma_e^2) = -\frac{n}{2}\ln(2\pi\sigma_e^2) - \frac{1}{2\sigma_e^2}\sum_{t=1}^{n}\left(\frac{\left(1-\phi_1 B-\phi_2 B^2-\cdots-\phi_p B^p\right)(1-B)^d (Z_t-\mu)}{\left(1+\theta_1 B+\theta_2 B^2+\cdots+\theta_q B^q\right)}\right)^2. \quad (1)$$

Estimation of d, ϕ, μ, θ can be obtained by maximizing equation (1) and this is referred as maximum likelihood estimation [4].

2. 4. Estimation of Fractional Difference Parameter with Semiparametric Methods

Estimation of fractional difference parameter with semiparametric methods is carried out through two steps. The first step is estimating the fractional difference parameter (d) and the second step is estimating AR and MA parameter [16]. The most popular semiparametric method used is Geweke dan Porter-Hudak (GPH). GPH method is performed by forming spectral density function or spectral equation of ARFIMA model through spectral regression equation $(f(\omega))$ with log-periodogram as the dependent variable and the series of autocovariance γ_k as pair of Fourier transformation:

$$\ln\left|I(\omega_j)\right| = \beta_0 + \beta_1 ln\left[4\sin^2(\omega_j^2)\right] + v_j,$$

where $\omega_j = \frac{2\pi j}{T}, j = 1,2,..,m$. The estimation of d is $\hat{\beta}_1$, ω_j represents $m = \sqrt{T}$ Fourier frequency, and $I(\omega_j)$ denotes the sample periodogram defined as $I(\omega_j) = \frac{1}{2\pi T}\left|\sum_{t=1}^{T} y_t e^{-\omega_j t}\right|^2$. The second step of GPH method is build ARMA model by using Box-Jenkins method after the estimated fractional difference parameter is obtained from the GPH method ($\hat{d}_{gph}$).

The next semiparametric method is called Sperio method introduced by Reisen and Lopes (1999). It is a modification from GPH method by replacing the periodogram with the smoothed spectral density. Reisen and Lopes (1999) proposed to use Blackman-Tukey type of estimation for the spectral density [17]:

$$f_m(x) = \frac{1}{2\pi}\sum_{s=-m}^{m} k\left(\frac{s}{m}\right)\hat{p}\,(s)\cos(sx)\ .$$

This estimated smoothed periodogram is denoted by $\hat{d}_{Sperio}$.

The third semiparametric method is Local Whittle estimation that is also commonly used for estimation of fractional difference parameter. This method was proposed by Kuensch (1987) and was modified by Robinson (1995). Local Whittle estimation of fractional difference parameter, denoted by $\hat{d}_{Whittle}$, is obtained by maximizing the likelihood of log Local Whittle on Fourier frequency that goes to zero [18]:

$$\Gamma(d) = -\frac{1}{2\pi m}\sum_{j=1}^{m} f(\omega_j; d).$$

The last semiparametric method considered in this study is Rescaled Range Statistic (R/S) or often called as Hurst statistic test. The last semiparametric method is *Rescaled Range Statistic* (R/*S*) or *Hurst* test. Besides being used to see indication of long-term memory in time series data, R/S statistic can also be used to estimate the fractional difference parameter with the following equation:

$$d = H - 0.5.$$

2. 5. Model Diagnostic Checking

Diagnostic checking is carried out to check the adequacy of fitted model to the observed data in order to reveal model inadequacies and to achieve model improvement. The diagnostic checking is done by observing if the model residual follows a white noise process or not, that is checking if the residuals are independent by using Ljung Box-Pierce test [4] and also checking if the residuals are normally distributed by using Jarque-Bera test [16].

2. 6. Selection of Best ARFIMA Model

Selection of best fitted model can be determined by Akaike Information Criteria (AIC) [19]. The AIC values takes into account how well the model fits the observed data and the number of parameters used the fitted model. It can be computed by using the following formula:

$$AIC = -2\log(maximum\ likelihood) + 2k.$$

where $k = p + q + 1$ if the model contains intercept and $k = p + q$ if the model does not cointain intercept [19].

A good model is considered and expected to be the best model for fitting data in sample and at the same time it is also a good model for forecasting out sample data. MAPE (Mean Absolute Percentage Error) is one of many criteria to test for the validity of the fitted model and will be used in this study. It is defined as the mean of the sum absolute deviation of predicted and observed value dividing by the observed value [20]:

$$MAPE = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{Y_t-\widehat{Y}_t}{Y_t}\right| \times 100,$$

where Y_t is the actual series, $\widehat{Y}_t$ is the predicted series, and N is the number of data sample.

4. RESULTS

Figure 1 displays the trend of wind speed at Soekarno-Hatta airport on a daily basis. It can be seen that the series are not stationer in variance as the fluctuations of the data tend to change over time or are not constant. A formal test is performed by using Box-Cox transformation to evaluate if transformation is needed to make the variance stationary in time.

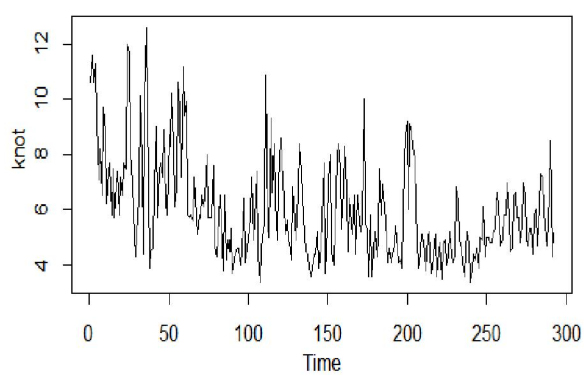

Figure 1. Plot of wind speed at Soekarno-Hatta airport.

Figure 2 indicate that the rounded value of optimal λ is not close to 1 and the range of lower and upper limit do not contain 1. According to this plot, the data needs to be transformed using square root transformation of Y_t ($\sqrt{Y_t}$). Afterwards, the stationary test in the mean is also performed by using ADF (Augmented Dickey Fuller) test. The result shows that we have strong evidence to reject the null hypothesis of non-stationary data since the p-value is less than 0.05 (p=0.01). Therefore, we can conclude that the wind speed data is already stationary in mean.

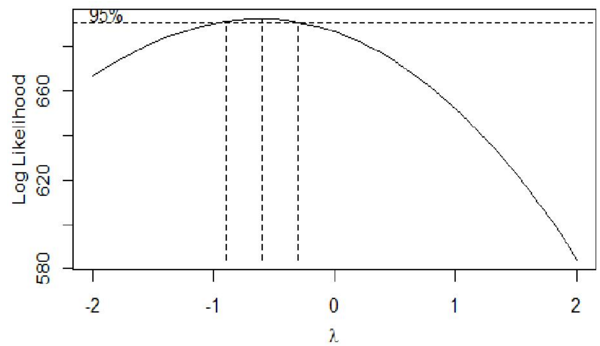

Figure 2. Box-Cox Profile Log Likelihood.

To identify if there is a long-term dependency, Hurst (H) statistic is calculated to the observed data. The computed $H = \frac{\log(R/S)_t}{\log(t)} = 0.738$ indicates that the transformed wind speed data has long-term dependency, thus ARFIMA*(p,d,q)* is the most appropriate model to be fitted to the observed data.

4.1. ARFIMA(p, d, q) Model Building with Parametric Method

In building ARFIMA model with parametric approach, the candidate models can be identified from the plot of ACF and PACF of the differenced series. A temporary *d* value is obtained by fitting ARFIMA$(0, d, 0)$ model. The estimated *d* is 0.397 (se=0.045). To identify the order of p and q as ARFIMA model, the value of d is set to 0.397. According to the plot of ACF and PACF, the model candidates are ARFIMA $(2, d, 0)$, ARFIMA $([7], d, 0)$, ARFIMA $([2,7], d, 0)$, ARFIMA $(0, d, 2)$, ARFIMA $(0, d, [7])$, ARFIMA $(0, d, [2,7])$, ARFIMA $(2, d, 2)$, ARFIMA $([7], d, [2])$, ARFIMA $([2,7], d, [2])$, ARFIMA $([2], d, [7])$, ARFIMA $([7], d, [7])$, ARFIMA $([2,7], d, 7)$, ARFIMA $([2], d, [2,7])$, and ARFIMA $([2,7], d, [2,7])$. Next, the parameters (ϕ, d, θ) for each candidate model are then estimated simultaneously by using EML method. Table 1 summarizes the estimated parameters for each model.

According to Table 1, ARFIMA$([7], 0.409, 0)$, ARFIMA$([2,7], 0.45, 0)$, ARFIMA$(0, 0.41, [7])$, A RFIMA$(0, 0.45, [2,7])$, ARFIMA$(2, 0.439, 2)$, ARFIMA$([7], 0.449, [2])$, dan ARFIMA$([2], 0.452, [7])$ models have all the parameters significant in the model. Table 2 summarizes the comparison of these 7 models based on AIC values. It shows that ARFIMA$(2, 0.439.2)$ has the lowest AIC value.

Table 1. Estimated parameter of ARFIMA(p, d, q) Using EML Method.

No.	ARFIMA(p, d, q) Model	Parameter	Coefficient	Standard Error	Sig.
1	$(2, d, 0)$	ϕ_1	0.082	0.089	No
		ϕ_2	-0.148	0.066	Yes
		d	0.399	0.070	Yes
2	$([7], d, 0)$	ϕ_7	-0.132	0.059	Yes
		d	0.409	0.043	Yes

Table 2. Estimated parameter of ARFIMA(p, d, q) Using EML Method (continued)

No.	ARFIMA(p, d, q) Model	Parameter	Coefficient	Standard Error	Sig.
3	$([2,7], d, 0)$	ϕ_2	-0.160	0.061	Yes
		ϕ_7	-0.125	0.058	Yes
		d	0.450	0.039	Yes
4	$(0, d, 2)$	θ_1	-0.089	0.112	No
		θ_2	0.112	0.077	No
		d	0.385	0.088	Yes
5	$(0, d, [7])$	θ_7	0.132	0.060	Yes
		d	0.410	0.043	Yes
6	$(0, d, [2,7])$	θ_2	0.139	0.058	Yes
		θ_7	0.126	0.061	Yes
		d	0.450	0.040	Yes
7	$(2, d, 2)$	ϕ_1	0.540	0.204	Yes
		ϕ_2	-0.796	0.116	Yes
		θ_1	0.541	0.265	Yes
		θ_2	-0.671	0.151	Yes
		d	0.439	0.055	Yes
8	$([7], d, [2])$	ϕ_7	-0.126	0.059	Yes
		θ_2	0.142	0.059	Yes
		d	0.449	0.040	Yes
9	$([2,7], d, [2])$	ϕ_2	-0.261	0.215	No
		ϕ_7	-0.120	0.059	Yes
		θ_2	-0.105	0.217	No
		d	0.447	0.040	Yes
10	$([2], d, [7])$	ϕ_2	-0.161	0.061	Yes
		θ_7	0.128	0.060	Yes
		d	0.452	0.039	Yes
11	$([7], d, [7])$	ϕ_7	-0.443	0.722	No
		θ_7	-0.323	0.769	No
		d	0.406	0.044	Yes
12	$([2,7], d, 7)$	ϕ_2	-0.159	0.062	Yes
		ϕ_7	-0.150	0.449	No
		θ_7	-0.026	0.466	No
		d	0.449	0.040	Yes
13	$([2], d, [2,7])$	ϕ_2	-0.331	0.209	No
		θ_2	-0.172	0.211	No
		θ_7	0.125	0.059	Yes
		d	0.448	0.040	Yes
14	$([2,7], d, [2,7])$	ϕ_2	-0.432	0.232	No
		ϕ_7	0.227	0.369	No
		θ_2	-0.277	0.242	No
		θ_7	0.348	0.359	No
		d	0.447	0.040	Yes

4.2. ARFIMA(p, d, q) Model Building with Semiparametric Methods

In semiparametric method, the fractional difference parameter is estimated separately from the A R and MA coefficients. Table 3 presents the estimated value of d by using the four semiparametric m ethods.

Tabel 3. AIC values of ARFIMA model fitted with EML method.

$\hat{d}_{EML}$	ARFIMA(p,d,q) Model	AIC
0.409	$([7],d,0)$	-1773.942
0.450	$([2,7],d,0)$	-1780.581
0.410	$(0,d,[7])$	-1773.884
0.450	$(0,d,[2,7])$	-1779.35
0.439	$(2,d,2)$	-1787.438
0.449	$([7],d,[2])$	-1775.627
0.452	$([2],d,[7])$	-1776.487

Table 4. Estimated value of *d* using semiparametric method.

Method	d	Standard Error
Geweke dan Porter-Hudak (GPH)	0.334	0.076
Smoothed GPH (Sperio)	0.359	0.033
Local Whittle	0.352	0.039
R/S	0.238	0.048

Model identification as ARFIMA(*p,d,q*) is conducted by using the plot of ACF dan PACF. With GPH method, the candidate models are ARFIMA$(2,d,0)$, ARFIMA$([7],d,0)$, ARFIMA$([1,2,7],d,0)$, ARFIMA$(0,d,1)$, ARFIMA$(0,d,[7])$, ARFIMA$(0,d,[1,7])$, ARFIMA$(2,d,1)$, ARFIMA$([7].d,1)$, ARFIMA$([1,2,7],d,1)$, ARFIMA$(2,d,[7])$, ARFIMA$(2,d,[1,7])$, ARFIMA$([1,2,7],d,[7])$, ARFIMA$([1,2,7],d,[1,7])$, and ARFIMA$([7],d,[7])$. With Sperio method, the candidate models are ARFIMA$(2,d,0)$, ARFIMA$([7],d,0)$, ARFIMA$([2,7],d,0)$, ARFIMA$(0,d,2)$, ARFIMA$(0,d,[7])$, ARFIMA$(0,d,[2,7])$, ARFIMA$(2,d,2)$, ARFIMA$([7],d,[2])$, ARFIMA$([2,7],d,[2])$, ARFIMA$([2],d,[7])$, ARFIMA$([7],d,[7])$, ARFIMA$([2,7],d,7)$, ARFIMA$([2],d,[2,7])$, and ARFIMA$([2,7],d,[2,7])$. With Local Whittle method, the candidate models are ARFIMA$(2,d,0)$, ARFIMA$([7],d,0)$, ARFIMA$([2,7],d,0)$, ARFIMA$(0,d,2)$, ARFIMA $(0,d,[7])$, ARFIMA $(0,d,[2,7])$, ARFIMA $(2,d,2)$, ARFIMA $([7],d,[2])$, ARFIMA $([2,7],d,[2])$, ARFIMA$([2],d,[7])$, ARFIMA$([7],d,[7])$, ARFIMA$([2,7],d,7)$,ARFIMA$([2],d,[2,7])$, ARFIMA $([2,7],d,[2,7])$. Lastly with R/S method, the candidate models are ARFIMA$(1,d,0)$,ARFIMA$(1,d,1)$, ARFIMA$(0,d,1)$, ARFIMA$(0,d,4)$, ARFIMA$(0,d,[1,4])$, ARFIMA$(1,d,[4])$, and ARFIMA$(1,d,[1,4])$.

Based on these candidate models, not all parameters are significant at 5% level. Those models with insignificant parameter are excluded from the candidates. Table 4 summarizes the comparison for all models with their corresponding AIC values. According to Table 4, the best fitted model with the lowest AIC value with GPH method is ARFIMA$(2,d,0)$. Based on Sperio method, ARFIMA$(2,d,0)$ has the smallest AIC value as compared to ARFIMA$([7],d,0)$,ARFIMA$(0,d,[7])$ and ARFIMA $(2,d,2)$ with $\hat{d}_{Sperio} = 0.359$. Based on Local Whittle method, ARFIMA$(2,d,2)$ model has the smallest AIC value as compared to ARFIMA$([7],d,0)$, ARFIMA$([2,7],d,0)$, ARFIMA$(0,d,[7])$, ARFIMA$(0,d,[2,7])$, ARFIMA$([7],d,[2])$, and ARFIMA$([2],d,[7])$ with $\hat{d}_{Whittle} = 0.352$. Based on R/S method, the smallest AIC values is for ARFIMA$(0,d,1)$ where $\hat{d}_{R/S} = 0.238$.

Diagnostic model is performed to the selected model by evaluating the assumption of the residuals that follow normal distribution and whether they are independent. The Jarque-Bera test as indicated in Table 5 shows that the residuals do not violate the normality assumption since the p-values are greater than 0.05. The Ljung-Box test to examine the assumption of independent indicates that the residuals do not correlate since the p-value is greater than 0.05. Thus, the residuals follow white noise process.

Table 4. Comparison of ARFIMA model using semiparametric method.

Semiparametric Method	Model ARFIMA	AIC
GPH ($\hat{d}_{gph} = 0.334$)	ARFIMA$(2, d, 0)$	-1786.023
	ARFIMA$([7], d, 0)$	-1770.698
	ARFIMA$(0, d, 1)$	-1785.356
	ARFIMA$([1,2,7], d, [7])$	-1767.478
	ARFIMA$([7], d, [7])$	-1757.494
Sperio ($\hat{d}_{Sperio} = 0.359$)	ARFIMA$(2, d, 0)$	-1786.023
	ARFIMA$([7], d, 0)$	-1770.698
	ARFIMA$(0, d, 1)$	-1785.356
	ARFIMA$([1,2,7], d, [7])$	-1767.478
	ARFIMA$([7], d, [7])$	-1757.494
Local Whittle ($\hat{d}_{Whittle} = 0.352$)	ARFIMA$([7], d, 0)$	-1773.893
	ARFIMA$([2,7], d, 0)$	-1779.591
	ARFIMA$(0, d, [7])$	-1773.836
	ARFIMA$(0, d, [2,7])$	-1778.372
	ARFIMA$(2, d, 2)$	-1787.139
	ARFIMA$([7], d, [2])$	-1774.698
	ARFIMA$([2], d, [7])$	-1775.391
R/S ($\hat{d}_{R/S} = 0.238$)	ARFIMA$(0, d, 1)$	-1784.631
	ARFIMA$(1, d, 0)$	-1783.05
	ARFIMA$(0, d, [4])$	-1766.393
	ARFIMA$(1, d, [4])$	-1779.315

Table 5. Normality and independent test.

d	ARFIMA Model	P-value from Normality test	P-value from Ljung-Box test
$\hat{d}_{gph} = 0.334$	$(2, d, 0)$	0.939	0.885
$\hat{d}_{Sperio} = 0.359$	$(2, d, 0)$	0.956	0.857
$\hat{d}_{Whittle} = 0.352$	$(2, d, 2)$	0.944	0.672
$\hat{d}_{R/S} = 0.238$	$(0, d, 1)$	0.823	0.863
$\hat{d}_{EML} = 0.439$	$(2, d, 2)$	0.975	0.725

Diagnostic model checking reveals that the candidate model based parametric and semiparametric methods show a good fit model since none of the assumptions are violated. Next, we examine all these five models in terms of accuracy by using MAPE. Table 6 shows that the smallest MAPE value is for ARFIMA$(2, d, 0)$ with $\hat{d}_{gph} = 0.334$. This model has MAPE of 17.760, showing that the model has relatively good forecasting ability.

Table 6. Forecasting accuracy.

d	ARFIMA Model	MAPE
$\hat{d}_{gph} =$ 0.334	$(2, d, 0)$	17.760
$\hat{d}_{Sperio}$ = 0.359	$(2, d, 0)$	17.791
$\hat{d}_{Whittle}$ = 0.352	$(2, d, 2)$	18.029
$\hat{d}_{R/S}$ = 0.238	$(0, d, 1)$	17.838
$\hat{d}_{EML}$ = 0.439	$(2, d, 2)$	18.242

ARFIMA$(2, d, 0)$ model with $\hat{d}_{gph} = 0.334$ can be expressed as follows:

$$\phi_2(B)\nabla^d Y_t = e_t,$$
$$\Leftrightarrow (1 - \phi_1 B - \phi_2 B^2)(1 - B)^{0.334} Y_t = e_t$$
$$\Leftrightarrow (1 + 0.148B - 0.117B^2)(1 - B)^{0.334} Y_t = e_t.$$

The value of $(1 - B)^{0.334}$ expresses the long-term memory in the seires. If $(1 - B)^{0.334} Y_t$ is denoted by W_t showing a long-term memory, then:

$$(1 + 0.148B - 0.117B^2)W_t = e_t,$$
$$\Leftrightarrow W_t + 0.148BW_t - 0.117B^2 W_t = e_t.$$

where $(1 - B)^{0.334}$ can be written as:

$$(1 - B)^{0.334} = 1 - 0.334B - (0.334)(1 - 0.334)B^2 - \frac{1}{6}(0.334)(1 - 0.334)(2 - 0.334)B^3 - \cdots,$$
$$\Leftrightarrow (1 - B)^{0.334} = 1 - 0.334B - 0.112B^2 - 0.062B^3 - \cdots.$$

Thus, ARFIMA $(2{,}0.334{,}0)$ model can be expressed as follows:

$$\Leftrightarrow W_t + 0.148BW_t - 0.117B^2 W_t = e_t,$$
$$\Leftrightarrow (1 - 0.334B - 0.112B^2 - 0.062B^3 - \cdots)Y_t + (1 - 0.334B - 0.112B^2 - 0.062B^3 - \cdots)0.148Y_{t-1}$$
$$- (1 - 0.334B - 0.112B^2 - 0.062B^3 - \cdots)0.117Y_{t-2} = e_t,$$
$$\Leftrightarrow Y_t - 0.186Y_{t-1} - 0.278Y_{t-2} - 0.04Y_{t-3} - \cdots = e_t,$$
$$\Leftrightarrow Y_t = 0.186Y_{t-1} + 0.278Y_{t-2} + 0.04Y_{t-3} - \cdots + e_t,$$

The results of the forecasted wind speed in Soekarno-Hatta airport from period of December 1st, 2018 to December 14th, 2018 using ARFIMA $(2{,}0.334{,}0)$ can be seen in Table 7.

Table 7. Results of forecasted wind speed value.

Date	Wind Speed Forecast (Knot)	Date	Wind Speed Forecast (Knot)
01/12/2018	5.582	08/12/2018	5.585
02/12/2018	5.582	09/12/2018	5.586
03/12/2018	5.583	10/12/2018	5.586
04/12/2018	5.583	11/12/2018	5.587
05/12/2018	5.584	12/12/2018	5.587
06/12/2018	5.584	13/12/2018	5.588
07/12/2018	5.585	14/12/2018	5.588

5. CONCLUSIONS

The semiparametric methods yield different estimate of fractional difference parameter, i.e. $\hat{d}_{gph} =$ 0.334, $\hat{d}_{Sperio} =$ 0.359, $\hat{d}_{Whittle} =$ 0.352, and $\hat{d}_{R/S} =$ 0.238 obtained from the GPH, Sperio, Local Whittle, and R/S methods, respectively, while the estimated fractional difference parameter is $\hat{d}_{EML} = 0.439$ based on parametric method. The best fitted model to forecast the wind speed is ARFIMA$(2, d, 0)$ with GPH semiparametric method with MAPE accuracy value of 17.76. The selected model can be expressed as follows:

$$Y_t = 0.186Y_{t-1} + 0.278Y_{t-2} + 0.04Y_{t-3} - \cdots + e_t.$$

From the above equation, it can be seen that the wind speed at Soekarno-Hatta airport have long-ter m memory. This might be due to the tendency of repeated wind cycles over time. The forecasted val ues in the next 14 days in the beginning of December 2018 show very little increase in wind speed.

REFERENCES

[1] A. R. Damanhuri, A. Priyadi and M. H. Purmono, "Prediksi Kecepatan Angin Jangka Pendek Menggunakan Metode Fuzzy Linear Regression Untuk Mendapatkan Masukan Pada Kontroler Turbin Angin," *JURNAL TEKNIK POMITS,* vol. I, no. 2, pp. 1-6, 2014.

[2] N. Ulinnuha and Y. Farida, "Prediksi Cuaca Kota Surabaya Menggunakan Autoregressive Integrated Moving Average (ARIMA) Box Jenkins dan Kalman Filter," *Jurnal Matematika "MANTIK",* vol. 4, no. 1, pp. 59-67, 2018.

[3] A. P. Desvina and M. Anggriani, "Peramalan Kecepatan Angin Di Kota Pekanbaru Menggunakan Metode Box-Jenkins," *Jurnal Sains dan Matematika Statistika,* vol. I, no. 2, 2015.

[4] W. W. Wei, Time Series Analysis : Univariate and Multivariate Methods. Second Edition, New York: Pearson Addison Wesley, 2006.

[5] C. W. Granger and R. Joyeux, "An introduction to long-memory time series models and fractional differencing," *Journal of time series analysis,* vol. I, pp. 15-29, 1980.

[6] F. Sowell, "Maximum likelihood estimation of stationary univariate fractionally integrated time series models," *Journal of econometrics,* vol. 53, no. 1-3, pp. 165-188, 1992.

[7] J. Geweke and S. Porter-Hudak, "The Estimation and Application of Long Memory Time Series Models," *Journal of Time Series Analysis,* vol. 4, no. 4, pp. 221-238, 1983.

[8] V. A. Reisen and S. Lopes, "Some Simulations and Applications of Forecasting Long-memory Time Series Models," *Journal of Statistical Planning and Inference,* vol. 80, no. 1-2, pp. 269-287, 1999.

[9] H. R. Kunsch, "Statistical aspects of self-similar processes," in *Proceedings of the First World Congress of the Bernoulli Society, 1987*, Zurich, 1987.

[10] P. M. Robinson, "Gaussian semiparametric estimation of long range dependence," *The Annals of statistics,* vol. 23, no. 5, pp. 1630-1661, 1995.

[11] National Climatic Data Center, "NOAA Satellite and Information Service," [Online]. Available: http://www7.ncdc.noaa.gov/CDO/dataproduct. [Accessed 5 November 2018].

[12] H. E. Hurst, "The problem of long-term storage in reservoirs," *Hydrological Sciences Journal,* vol. I, no. 3, pp. 13-27, 1956.

[13] J. A. Doornik and M. Ooms, "A package for estimating, forecasting and simulating ARFIMA models: Arfima package 1.0 for Ox," *Preprint, Erasmus University,* 1999.

[14] J. R. Hosking, "Fractional differencing," *Biometrika,* vol. 68, no. 1, pp. 165-176, 1981.

[15] M. Boutahar and R. Khalfaoui, "Estimation of the long memory parameter in non stationary models: A Simulation Study," 2011.

[16] R. A. H. Mohamed, "Using Arfima Models in Forecasting The Total Value Of Traded Securites On The Arab Republic of Egypt," *International Journal of Research and Reviews in Applied Sciences,* vol. 27, no. 1, pp. 26-34, 2016.

[17] B. Joe and R. Sisir, "US Housing Price Bubbles: A Long Memory Approach," in *46th Annual Conference of the Money ,Macro and Finance Research Group*, Durham University Business School, 2014.

[18] G. Bharwaj and N. Swanson, "An Empirical Investigation of the Usefulness of ARFIMA Models for Predicting Macroeconomic and Financial Time Series," *Journal of Econometrics,* vol. 8, no. 1, pp. 539-578, 2006.

[19] J. D. Cryer and K.-S. Chan, Time Series Analysis With Applications in R Second Edition Springer Science+ Business Media, LLC, 2008.

[20] D. Rosadi, Analisis Ekonometrika & Runtun Waktu Terapan dengan R : Aplikasi untuk Bidang Ekonomi, Bisnis, dan Keuangan, Yogyakarta: ANDI, 2011.