IJDASEA
International Journal of Data Science Engineering, and Analytics

e-ISSN : 2807-1689
p-ISSN : 2798-9208

Visitor Forecasting Wisata Bahari Lamongan (WBL) Using Hybrid Particle Swarm Optimization (PSO) and Seasonal ARIMA

Dinita Rahmalia[1,*]

[1] Universitas Islam Darul Ulum Lamongan; dinitarahmalia@gmail.com
* Correspondence: dinitarahmalia@gmail.com

Citation: Rahmalia, D. Visitor Forecasting Wisata Bahari Lamongan (WBL) using Hybrid Particle Swarm Optimication (PSO) and Seassonal ARIMA. *C***2021**, Volume *1*, Number 2, Page 27-35.

Received: October 21, 2021
Accepted: November 16, 2021
Published: November 25, 2021

Abstract: The revenue of city is determined by some factors, one of them is tourism sector. A problem of tourism sector is forecasting visitors Wisata Bahari Lamongan (WBL). Because data of the number of visitors WBL are fluctuating and seasonal, then it is required Seasonal ARIMA method. In the Seasonal ARIMA method, there are some parameters that should be optimized for producing forecasting with small mean square error (MSE). In this research, Seasonal ARIMA parameters will be optimized by Particle Swarm Optimization (PSO). PSO is optimization algorithm inspired by behavior of birds group in searching food. Based on simulation results, PSO algorithm can optimize Seasonal ARIMA parameter which is optimal and it can produce forecasting result with small MSE.

Keywords: time series; seasonal arima; forecasting; tourism

1. Introduction

The revenue of city is determined by some factors, one of them is tourism sector. A problem of tourism sector is forecasting visitors Wisata Bahari Lamongan (WBL) located in Lamongan, East Java. According to Badan Pusat Statistika (BPS) Lamongan, the number of visitors WBL is very high moreover there are many visitors from outside of city and outside of province [1].

Because data of the number of visitors WBL are fluctuating and seasonal, then it is required forecasting method. From previous method, there are some suitable methods used in forecasting problem such as Exponential Smoothing [2], Kalman Filter [3], Neural Network [4], Fuzzy System [5], Adaptive Neuro Fuzzy Inference System (ANFIS) [6]. ARIMA method is also applied in forecasting on nonstationary data [7]. Seasonal ARIMA is also applied in forecasting of coffee production [8] and forecasting of airline passenger [9].

Before seasonal ARIMA method is applied, first we decide whether data are either nonstationary or stationary. In nonstationary data, differencing process is required so that data become stationary data. After that, constructing Autocorrelation Function (ACF) plot for determining Moving Average (MA) order, constructing Partial Autocorrelation Fuction (PACF) plot for determining Autoregressive (AR) order.

In the Seasonal ARIMA method, there are some parameters that should be optimized for producing forecasting with small mean square error (MSE). Generally, methods used for parameter estimation are methods of moment, maximum likelihood, and least square for identifying model [10]. In this research, Seasonal ARIMA parameters will be optimized by Particle Swarm Optimization (PSO).

PSO is optimization algorithm inspired by behavior of birds group in searching food. This algorithm was discovered by Kennedy dan Eberhart in 1995. Population is called swarm and the individu is called particle. When the particle find the food, other particle will follow them [11]. In

PSO applied in Seasonal ARIMA, there is initialization of particle i.e. Seasonal ARIMA parameters as decision variable.

Based on simulation results, Seasonal ARIMA method can determine where both ACF and PACF plot cut off for identifying the order of ARIMA and after that PSO algorithm can optimize Seasonal ARIMA parameter which is optimal and it can produce forecasting result with small MSE.

2. Related Works

The related works of this research are about time series data, autocorrelation function (ACF), and partial autocorrelation function (PACF), Autoregressive Integrated Moving Average (ARIMA). Then we use Particle Swarm Optimization (PSO) for optimizing the Seasonal ARIMA parameters.

2.1. Stationary and Nonstationary Time Series

A time series is strictly stationary if its properties are not affected by a change in the time origin. A time series that exhibits a trend is a nonstationary time series [10].

The approaching to remove trend is by differencing the data, applying the difference operator to the original time series to obtain a new time series. Differecing can use backward difference operator ∇ or backshift operator B. Equation (1) is the first order differencing.

$$\begin{aligned} x_t &= y_t - y_{t-1} \\ x_t &= \nabla y_t \\ x_t &= (1-B)y_t = y_t - By_t \end{aligned} \tag{1}$$

In general, the d-th order differencing can be expanded in (2) and (3)

$$B^d y_t = y_{t-d} \tag{2}$$

$$\nabla^d y_t = (1-B)^d y_t \tag{3}$$

2.2. Autocovariance and Autocorrelation Function (ACF)

The covariance between y_t and its value at another time period y_{t+k} is called autocovariance at lag k as in (4).

$$\begin{aligned} \gamma_k &= \operatorname{cov}(y_t, y_{t+k}) \\ \gamma_k &= E\left((y_t - \bar{y})(y_{t+k} - \bar{y})\right) \\ \gamma_k &= \frac{1}{T}\sum_{t=1}^{T-k}(y_t - \bar{y})(y_{t+k} - \bar{y}) \end{aligned} \tag{4}$$

with $k = 0,1,2,\ldots,T-1$

The autocorrelation coefficient (ACF) at lag k for a stationary time series is in (5) and (6)

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\operatorname{cov}(y_t, y_{t+k})}{\operatorname{var}(y_t)} = \frac{E\left((y_t - \bar{y})(y_{t+k} - \bar{y})\right)}{E\left((y_t - \bar{y})^2\right)} \tag{5}$$

$$\rho_k = \frac{\sum_{t=1}^{T-k}\left(y_t - \bar{y}\right)\left(y_{t+k} - \bar{y}\right)}{\sum_{t=1}^{T}\left(y_t - \bar{y}\right)^2}, \quad k = 0,1,2,...,T-1 \tag{6}$$

2.3. Partial Autocorrelation Function (PACF)

Consider the equation for ACF process in (7) :

$$\begin{aligned} \rho(1) &= \phi_{1k}\rho(0) + \phi_{2k}\rho(1) + ... + \phi_{kk}\rho(k-1) \\ \rho(2) &= \phi_{1k}\rho(1) + \phi_{2k}\rho(0) + ... + \phi_{kk}\rho(k-2) \\ &\vdots \\ \rho(k) &= \phi_{1k}\rho(k-1) + \phi_{2k}\rho(k-2) + ... + \phi_{kk}\rho(0) \end{aligned} \tag{7}$$

We can write the equation in matrix notation :

$$\begin{bmatrix} 1 & \rho(1) & \rho(2) & \cdots & \rho(k-1) \\ \rho(1) & 1 & \rho(3) & \cdots & \rho(k-2) \\ \rho(2) & \rho(1) & 1 & \cdots & \rho(k-3) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho(k-1) & \rho(k-2) & \rho(k-3) & \cdots & \rho(0) \end{bmatrix} \begin{bmatrix} \phi_{1k} \\ \phi_{2k} \\ \phi_{3k} \\ \vdots \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho(1) \\ \rho(2) \\ \rho(3) \\ \vdots \\ \rho(k) \end{bmatrix} \tag{8}$$

For solving partial autocorrelation coefficient (PACF) at lag k (ϕ_{kk}) for a stationary time series is in (9).

$$\begin{bmatrix} \phi_{1k} \\ \phi_{2k} \\ \phi_{3k} \\ \vdots \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} 1 & \rho(1) & \rho(2) & \cdots & \rho(k-1) \\ \rho(1) & 1 & \rho(3) & \cdots & \rho(k-2) \\ \rho(2) & \rho(1) & 1 & \cdots & \rho(k-3) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho(k-1) & \rho(k-2) & \rho(k-3) & \cdots & \rho(0) \end{bmatrix}^{-1} \begin{bmatrix} \rho(1) \\ \rho(2) \\ \rho(3) \\ \vdots \\ \rho(k) \end{bmatrix} \tag{9}$$

2.4. Autoregressive Integrated Moving Average (ARIMA) Model

The model of moving average process $MA(q)$ is as follows [10] :

$$y_t = \mu + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - .... - \theta_q\varepsilon_{t-q}$$
$$y_t = \mu + (1 - \theta_1 B - \theta_2 B^2 - .... - \theta_q B^q)\varepsilon_t$$
$$y_t = \mu + \Theta(B)\varepsilon_t$$

with ε_t is white noise.

The feature of the ACF is very helpful identifying the MA model and its appropriate order as it cuts off after lag q. Therefore for $MA(q)$, ACF cuts off after lag q.

The model of autoregressive process $AR(p)$ is as follows [10] :

$$y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \varepsilon_t$$
$$y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} - .... - \phi_p y_{t-p} = \delta + \varepsilon_t$$
$$(1 - \phi_1 B - \phi_2 B^2 - .... - \phi_p B^p) y_t = \delta + \varepsilon_t$$
$$\Phi(B) y_t = \delta + \varepsilon_t$$

with ε_t is white noise.

The feature of the PACF is very helpful identifying the AR model and its appropriate order as it cuts off after lag p. Therefore for $AR(p)$, PACF cuts off after lag p.
The model of Autoregressive Integrated Moving Average $ARIMA(p,d,q)$ process is given as [10] :

$$\Phi(B)(1-B)^d y_t = \delta + \Theta(B)\varepsilon_t \tag{10}$$

With ε_t is white noise and $\Phi(B)$ and $\Theta(B)$ can be constructed as follows :

$$\Phi(B) = (1-\phi_1 B - \phi_2 B^2 - .... - \phi_p B^p) \tag{11}$$

$$\Theta(B) = (1-\theta_1 B - \theta_2 B^2 - .... - \theta_q B^q) \tag{12}$$

2.5. Seasonal ARIMA Model
Sometimes time series data exhibit periodic pattern so that it is required Seasonal ARIMA. Equation (13) shows the difference of data between seasonal periodic.

$$S_t - S_{t-s} = (1-B^S)S_t = 0 \tag{13}$$

Process $w_t = (1-B^S)y_t$ can be seen seasonally stationary. When ARMA model is used, the it is obtained :

$$\Phi(B)w_t = (1-B^S)\Theta(B)\varepsilon_t \tag{14}$$

with ε_t is white noise.

So that the equation of Seasonal ARIMA with order $(p,d,q)\times(P,D,Q)$ is :

$$\Phi^*(B^S)\Phi(B)(1-B)^d(1-B)^D y_t = \delta + \Theta^*(B^S)\Theta(B)\varepsilon_t \tag{15}$$

with
$\Phi^*(B^S) = (1-\phi_1^* B^s - \phi_2^* B^{2s} - \ldots - \phi_P^* B^{Ps})$
$\Theta^*(B^S) = (1-\theta_1^* B^s - \theta_2^* B^{2s} - \ldots - \theta_Q^* B^{Qs})$

2.6. Particle Swarm Optimization (PSO)
Particle Swarm Optimization (PSO) was the algorithm discovered by Kennedy and Eberhart in 1995. This algorithm is inspired from the flock of birds in searching the food. Population in this algorithm is called swarm and the individu is called particle. When the particle find the source of food, other particle will follow them [12].
In this research, Seasonal ARIMA parameters $\phi_1, \phi_1^*, \theta_1, \theta_2, \theta_1^*, \delta$ will be optimized by PSO algorithm with velocity update in (16) and particle position in (17).

$$v_i^k(t+1) = wv_i^k(t) + c_1 r_1(p_i^k - x_i^k(t)) + c_2 r_2(g_i - x_i^k(t)) \tag{16}$$

$$x_i^k(t+1) = x_i^k(t) + v_i^k(t+1) \quad (17)$$

Where w is inertia weight between 0.9-1.2, $p_{ij}^{\ k}$ is the local best position of particle k, g_{ij} is the global best position in the swarm. $c_1 = c_2 = 2$ and r_1 and r_2 are uniform random numbers between 0 and 1 [13].

3. Experiment and Analysis

Data used are the number of visitors Wisata Bahari Lamongan (WBL) during year 2015 obtained from BPS Kab. Lamongan. Observation time perday is started from 08.00 until 15.00 so that in each month is derived the average of the number of visitors started from time 08.00, time 09.00, time 10.00, time 11.00, time 12.00, time 13.00, time 14.00, and time 15.00 with graph as in Figure 1.

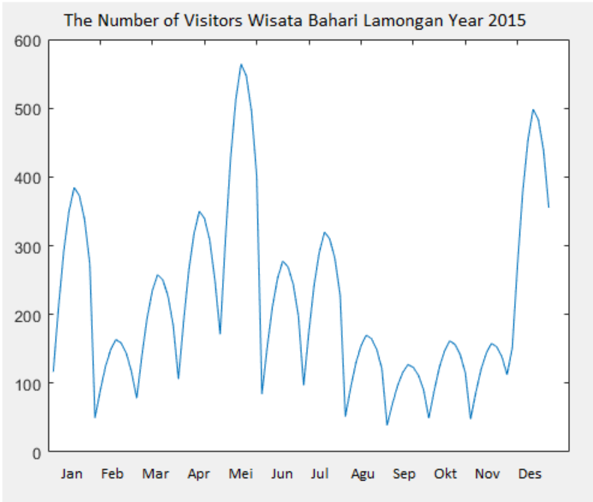

Figure 1. The number of visitors WBL year 2015

Due to every month data exhibit nonstationary, then it is applied first differencing and seasonal in lag 8, so that differencing data are :

$$w_t = (1-B)(1-B^8)y_t$$

with graph as in Figure 2.

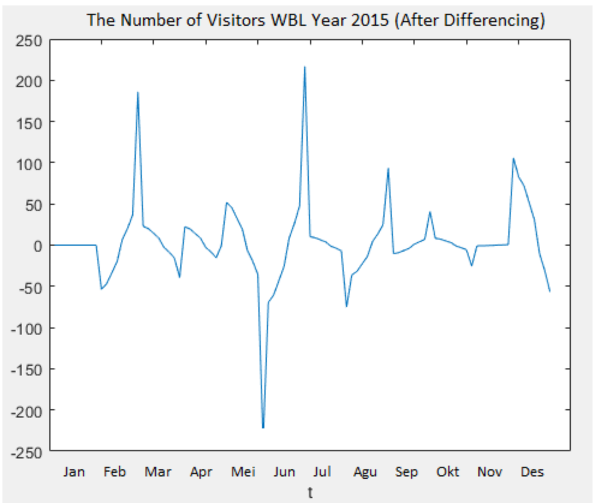

Figure 2. The number of visitors WBL year 2015 after first differencing

Data which have been differencing exhibit stationary process. From differenced data, the ACF and PACF are computed so that obtained graph ACF as in Figure 3 and graph PACF as in Figure 4.

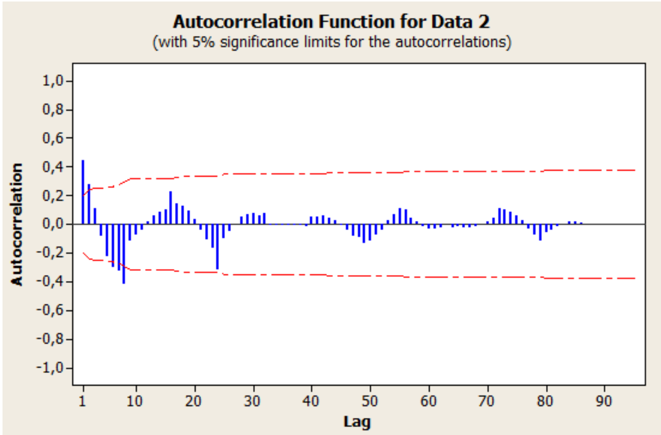

Figure 3. The graph of ACF on differenced data

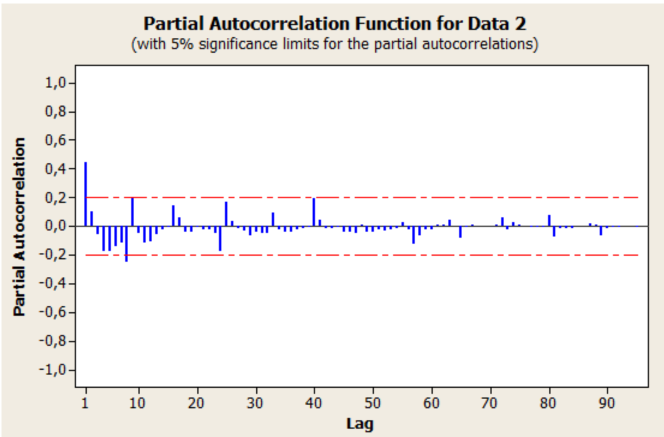

Figure 4. The graph of PACF on differenced data

In Figure 3, the graph of ACF exhibit cut off in lag 2 so that MA(2) model is derived. While in Figure 4, the graph of PACF exhibit cut off in lag 1 so that AR(1) model is derived.
For seasonal data, lag multiplied by 8 will be analyzed, such as lag 8, lag 16, lag 24 and so on. In Figure 3, the graph of ACF exhibit cut off only in lag 8 so that MA(1) model is derived. In Figure 4, the graph of PACF exhibit cut off only in lag 8 so that AR(1) model is derived.
From the graph ACF and graph PACF the model Seasonal ARIMA is :

$$ARIMA(1,1,2)\times(1,1,1)_8$$

With the formula in (16)

$$\Phi^*(B^8)\Phi(B)(1-B)(1-B)y_t=\delta+\Theta^*(B^8)\Theta(B)\varepsilon_t \quad (16)$$

with

$\Phi^*(B^8)=(1-\phi_1^* B^8)$

$$\Theta^*(B^8) = (1-\theta_1^* B^8)$$

$$\Phi(B) = (1-\phi_1 B)$$

$$\Theta(B) = (1-\theta_1 B - \theta_2 B^2)$$

From Seasonal ARIMA parameters obtained $\phi_1, \phi_1^*, \theta_1, \theta_2, \theta_1^*, \delta$, they will be optimized by PSO algorithm.
The result of PSO simulation is shown in Figure 5. In early iteration, from Seasonal ARIMA parameters resulted, MSE (Mean of Square Error) value computed is large enough. Then it is done velocity update process and particle position so that the optimal Seasonal ARIMA parameters are :

$$(\phi_1; \phi_1^*; \theta_1; \theta_2; \theta_1^*; \delta) = (-0,6740\ ; -0,9376;\ -0,0522\ ; -0,0313\ ;\ 0,2389\ ;\ 5.9572)$$

with MSE value is 7200.

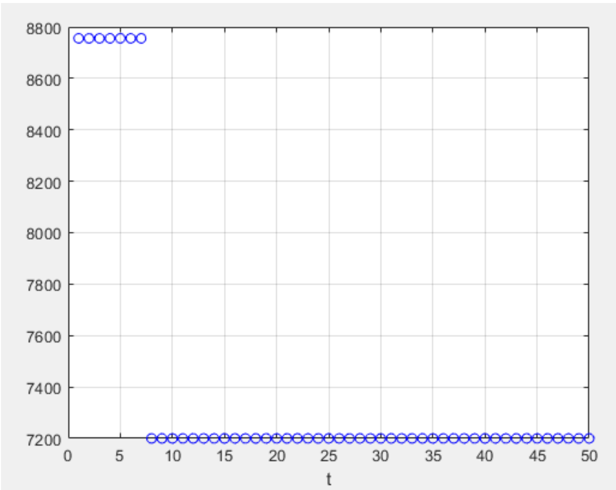

Figure 5. PSO simulations in optimization of Seasonal ARIMA parameters

Figure 6 is the comparison actual value and forecasting result the number of visitors WBL year 2015, every month started from time 08.00, time 09.00, time 10.00, time 11.00, time 12.00, time 13.00, time 14.00, and time 15.00.

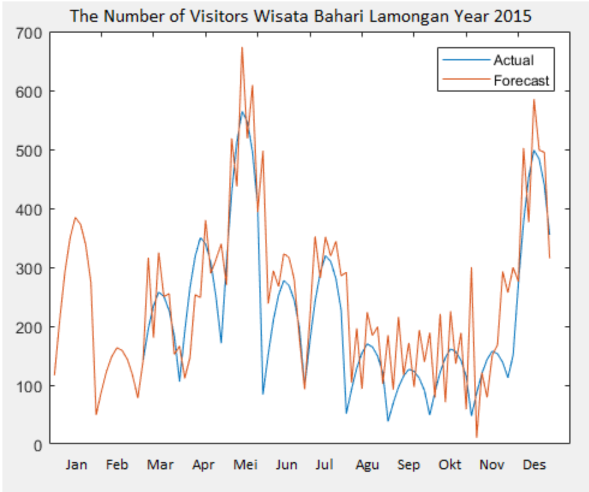

Figure 6. Forecasting result using Seasonal ARIMA

4. Conclusions

In nonstationary data, differencing process is required so that data become stationary data. After that, constructing Autocorrelation Function (ACF) plot for determining Moving Average (MA) order, constructing Partial Autocorrelation Fuction (PACF) plot for determining Autoregressive (AR) order. Because data of the number of visitors WBL are fluctuating and seasonal, then it is required Seasonal ARIMA method. In the Seasonal ARIMA method, there are some parameters that should be optimized for producing forecasting with small mean square error (MSE). In this research, Seasonal ARIMA parameters will be optimized by Particle Swarm Optimization (PSO). Based on simulation results, PSO algorithm can optimize Seasonal ARIMA parameter which is optimal and it can produce forecasting result with small MSE. The development of this research is other optimization techniques of Seasonal ARIMA parameter so that better forecasting can be produced.

References

1. Bhaskoro, B.; Mustain, M.; Suntoyo. Evaluasi dan Optimalisasi Fasilitas Taman Hiburan Wisata Bahari Lamongan di Era New Normal. *Jurnal Teknik ITS* **2020**, *10*
2. Rahmalia, D. Estimation of Exponential Smoothing Parameter on Pesticide Characteristic Forecast Using Ant Colony Optimization. *Eksakta* **2018**, *18*, pp. 56–63
3. Muhith, A.; Herlambang, T.; Rahmalia, D. Estimation of Thrombocyte Concentrate (TC) and Whole Blood (WB) Using Unscented Kalman Filter. *International Journal of Advanced Science and Technology* **2020**, *29*, pp. 25–32
4. Rahmalia, D.; Pradana, M.S. Backpropagation Neural Network pada Data yang Tak Stationer (Studi Kasus: Jumlah Penderita Penyakit Ebola). *Jurnal Riset dan Aplikasi Matematika (JRAM)* **2019**, *3*, pp. 32-42
5. Rahmalia, D. Optimizing the Membership Degree of Fuzzy Inference System (FIS) and Fuzzy Clustering Means (FCM) in Weather Data Using Firefly Algorithm. *Cauchy* **2021**, *6*, pp. 169-180
6. Rahmalia, D.; Rohmatullah, A. Pengaruh Korelasi Data pada Peramalan Kelembaban Udara Menggunakan Adaptive Neuro Fuzzy Inference System. *Applied Technology and Computing Science Journal* **2019**, *2*, pp. 10-24
7. Pradana, M.S.; Rahmalia, D.; Prahastini, E.D.A. Peramalan Nilai Tukar Petani Kabupaten Lamongan dengan ARIMA. *Jurnal Matematika* **2020**, *10*, pp. 91-104

8. Tantika, H.N.; Supriadi, N.; Anggraini, D. Metode Seasonal ARIMA untuk Meramalkan Produksi Kopi dengan Indikator Curah Hujan. *Jurnal Matematika* **2018**, *12*
9. Ukhra, A.U. Pemodelan dan Peramalan Data Deret Waktu dengan Metode Seasonal ARIMA. *Jurnal Matematika UNAND* **2016**, *3*, pp. 59-67
10. Montgomery, D.C.; Jennings, C.L.; Kulahci, M. *Introduction to Time Series Analysis and Forecasting*; John Wiley and Sons: New Jersey, USA, 2015
11. Shi, Y.H.; Eberhart, R.C. A Modified Particle Swarm Optimization. In Proc. Int. Conf. On Evolutionary Computation, 1998
12. Rao, S.S. *Engineering Optimization Theory and Practice*; John Wiley and Sons: New Jersey, USA, 2009
13. Rahmalia, D. Particle Swarm Optimization-Genetic Algorithm (PSOGA) on Linear Transportation Problem, AIP Conference Proceeding, Surabaya, Indonesia, 2017