Analysis of statistical models for forecasting PM\textsubscript{10} in Kototabang region

S Alfiandy\textsuperscript{1,*}, R S Davi\textsuperscript{2,*}

\textsuperscript{1} Global Atmosfer Watch Station (GAW), Palu, 94111, Indonesia
\textsuperscript{2} Global Atmosfer Watch Station (GAW), Sumatera Barat, 26151, Indonesia

*E-mail: solihalfiandy@gmail.com

Abstract. PM\textsubscript{10} is one of the aerosol particles that can endanger human health. This research conducted by forecasting for PM\textsubscript{10} concentration. Forecasting is an activity of estimating or predicting events in the future, therefore it is necessary to do analysis simple statistical model to know good results. In this case, several forecasting models are used for the daily PM\textsubscript{10} concentration in Kototabang, that is Linear, Quadratic, and Exponential Trend Model. As the results of this research, monthly forecasting using Linear Trend Model has the highest correlation value in October (-60) with MAD value 0.254, MSE 0.0651, RMSE 0.255, and MAPE 1848.8. Monthly forecasting using the Quadratic Trend Model has the highest correlation value in February (+0.52) with MAD value 0.013, MSE 0.0002, RMSE 0.015, and MAPE 103.934. Then, for monthly forecasting using the Exponential Trend Model has the highest value in October (+0.61) with MAD value 0.124, MSE 0.0154, RMSE 0.124, and MAPE 893.484. The output of the forecasting model obtained the best model for forecasting the daily PM\textsubscript{10} concentration in Kototabang i.e by using a simple statistical model, linear trend model.

1. Introduction

To solve the problems in the future and of course this can not be ascertained the magnitude, researchers always try to solve in order to find out the magnitude with the approach models that are in accordance with behavior actual data, in this case for forecasting problems. In general, the notion of forecasting is interpretation. But using certain techniques, deepening is therefore not just an interpretation without techniques. Forecasting is an activity forecasts or predictions of events in the future, of course with the help of preparing plans in advance, where these plans are made based on the capacity and capability of demand or production that has been done in the company [3].

The main purpose of forecasting is to predict demand in the future so that an estimate is approaching the actual situation. In forecasting systems, the use of forecasting methods greatly influences the results of forecasting. The division of deepening methods can be divided into several aspects, depending on the point of view[7].

In forecasting, there are two methods of deepening, namely qualitative methods and quantitative methods. Quantitative deepening method is a forecasting method that in calculations uses mathematical calculations. Quantitative methods are grouped into two types of methods of periodic time series(time-series)and correlation – causality causal method[2].

PM\textsubscript{10} is defined as aerosol particles which have a size of up to 10 μm when finding in high concentrations that will endanger human health. Therefore, the measurement of PM\textsubscript{10} concentration is very important[1]. One of the air pollutants is dust that has a size of 0.1 to 100 μm in diameter and is of...
mutual concern, especially dust generated from the management of industrial solids and disposal from motor vehicles. Environmental and health experts believe that PM$_{10}$ and PM$_{2.5}$ can trigger human respiratory tract infections because these particles can settle to the respiratory tract of the bronchi and alveoli.

Based on the Republic of Indonesia Government Regulation No.41 of 1999 concerning the control of air pollution, for the national ambient air quality standard, the PM$_{10}$ concentration is 150 μg / m$^3$ (24 hours). If the concentration exceeds the quality standards set by the government, it can be stated that the concentration of PM$_{10}$ has polluted the air\cite{1}.

The main objective of this research is to analyze a simple statistical forecasting model that is Linear Trend Model, Quadratic Trend Model, and Exponential Trend Model-based on data for Aerosol PM$_{10}$ concentrations daily in 2015 for forecasting in 2016. After forecasting, then the data is verified between forecasting values. 2016 daily data with actual data using the verification method Mean Absolute Deviation (MAD), Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Percent Error (MAPE), and correlation.

2. Data and Methods

2.1. Data
Data used in this study are PM$_{10}$ concentration data for the daily period January 2015 – December 2016. The data was obtained from the Kototabang Global Atmospheric Monitoring Station (GAW) West Sumatra. In the analysis of this statistical model supported by the application of Minitab 16 and Microsoft Excel 2016. The use of the application is a means to facilitate data processing, such as for example the analysis of PM$_{10}$ concentration forecasting models.

2.2. Linier Trend Model
Trend Linear Model is an analytical method intended for forecasting. Linear Trend is data tendency where the change is based on a fixed and constant time \cite{5}:

$$Y_t = a + bt$$  \hspace{1cm} (1)

2.3. Quadratic Trend Model
The quadratic Trend Model is a data tendency method with curved patterned curves. Means a quadratic trend model \cite{8}:

$$Y_t = a + bt + ct^2$$  \hspace{1cm} (2)

2.4. Exponential Trend Model
Exponential Trend Model is an equation expressed in the form of a time variable (X) and expressed as a rank. To find the a, and b values of Y and X data, used the following formula is\cite{4}:

$$y^2 = ab^x$$  \hspace{1cm} (3)

2.5. Mean Absolute Percent Error
Mean Absolute Percent Error (MAPE) is used to calculate the average percentage error (Average Absolute Percent Error) with the equation as follow \cite{9}:

$$MAPE = \frac{1}{N} \sum_{t=1}^{N} \left| \frac{A_t - F_t}{A_t} \right| \times 100$$  \hspace{1cm} (4)
2.6. Mean Square Error
Mean Square Error (MSE) is used to calculate the average rank error (Average of Squared error) with the following equation [10]:

$$MSE = \sum_{t=1}^{n} (A_t - F_t)^2$$

(5)

2.7. Root Mean Square Error
Root Mean Square Error (RMSE) is used to find out how much error is obtained from the results of forecasting by taking into account the value of MSE. Following is the formula[10]:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n} (A_t - F_t)^2}{N}}$$

(6)

2.8. Mean Absolute Deviation
Mean Absolute Deviation (MAD) is used to calculate the average error (Average Absolute Error) with the following equation [6]:

$$MAD = \frac{\sum_{t=1}^{n} |A_t - F_t|}{n}$$

(7)

2.9. Correlation
Correlation is the level of closeness of relationships between variable that have values of -1 to 1. If the correlation coefficient gets closer to 1, it indicates that the relationship between variables is getting tighter. If the coefficient is close to 0, then the variables are considered to have no relationship, and if the coefficient is getting closer to -1, it indicates that the variables have relationships the opposite[9]:

$$r_{XY} = \frac{n \Sigma XY - \Sigma X \Sigma Y}{\sqrt{n \Sigma X^2 - (\Sigma X)^2} \sqrt{n \Sigma Y^2 - (\Sigma Y)^2}}$$

(8)

3. Results and Discussion

3.1. Equation Model
Following is a model equation from Linear Trend Model, Quadratic Trend Model, and S-Quve Trend Model:

| Month   | Equations                   |
|---------|-----------------------------|
| 1       | January: $Y_t = 0.01473 + 0.000188 \times t$ |
| 2       | February: $Y_t = 0.02362 + 0.000268 \times t$ |
| 3       | March: $Y_t = 0.00463 - 0.000530 \times t$ |
| 4       | April: $Y_t = 0.01972 + 0.000228 \times t$ |
| 5       | Mei: $Y_t = 0.01345 + 0.000199 \times t$ |
| 6       | June: $Y_t = 0.01277 + 0.000453 \times t$ |
| 7       | July: $Y_t = 0.01430 + 0.000618 \times t$ |
| 8       | August: $Y_t = 0.00790 + 0.00139 \times t$ |
| 9       | September: $Y_t = 0.0939 - 0.000629 \times t$ |
| 10      | October: $Y_t = 0.1419 + 0.00269 \times t$ |
| 11      | November: $Y_t = 0.2813 - 0.000534 \times t$ |
| 12      | December: $Y_t = 0.00319 - 0.000282 \times t$ |
Table 2. Equations table of quadratic trend model.

| Month | Equations |
|-------|-----------|
| 1     | January $Y_t = 0.014 \times (1.01231 \times t)$ |
| 2     | February $Y_t = 0.023 \times (1.00959 \times t)$ |
| 3     | March $Y_t = 0.0134 \times (1.01162 \times t)$ |
| 4     | April $Y_t = 0.020 \times (0.98344 \times t)$ |
| 5     | Mei $Y_t = 0.013 \times (1.02213 \times t)$ |
| 6     | June $Y_t = 0.015 \times (1.02351 \times t)$ |
| 7     | July $Y_t = 0.015 \times (1.02351 \times t)$ |
| 8     | August $Y_t = 0.013 \times (1.04295 \times t)$ |
| 9     | September $Y_t = 0.088 \times (0.9782 \times t)$ |
| 10    | October $Y_t = 0.164 \times (0.9964 \times t)$ |
| 11    | November $Y_t = 0.024 \times (0.98174 \times t)$ |
| 12    | December $Y_t = 0.003 \times (1.0355 \times t)$ |

Table 3. Equations table of exponential trend model.

| Month | Equations |
|-------|-----------|
| 1     | January $Y_t = 0.01861 - 0.000517 \times t + 0.000022 \times t^2$ |
| 2     | February $Y_t = 0.02343 + 0.000306 \times t - 0.000001 \times t^2$ |
| 3     | March $Y_t = 0.01301 - 0.000996 \times t + 0.000048 \times t^2$ |
| 4     | April $Y_t = 0.01325 + 0.000986 \times t - 0.000039 \times t^2$ |
| 5     | Mei $Y_t = 0.01508 - 0.000098 \times t + 0.000009 \times t^2$ |
| 6     | June $Y_t = 0.02449 - 0.001744 \times t + 0.000071 \times t^2$ |
| 7     | July $Y_t = 0.03077 - 0.002376 \times t + 0.000094 \times t^2$ |
| 8     | August $Y_t = 0.02594 - 0.00189 \times t + 0.000103 \times t^2$ |
| 9     | September $Y_t = 0.1341 - 0.00189 \times t + 0.000243 \times t^2$ |
| 10    | October $Y_t = 0.1303 + 0.00480 \times t - 0.000066 \times t^2$ |
| 11    | November $Y_t = 0.03820 - 0.002422 \times t + 0.000061 \times t^2$ |
| 12    | December $Y_t = 0.01052 - 0.001050 \times t + 0.000042 \times t^2$ |

3.2. Results of estimated daily PM$_{10}$ concentration.
The following is the output or forecast results from the PM$_{10}$ daily in 2016 from the three methods analyzed, the forecast results will be displayed in graphical form.

In tables 1, table 2, and table 3 which is the output of the equation forecasting models. There is an error in Exponential Trend Model on March, in making the forecast come out (ERROR - some data are
non-positive; cannot fit growth model) the data writing is not convincing, meaning that the PM$_{10}$ concentration data in March cannot be forecast using the Exponential Trend Model. The error is suspected because the model trend does not match with the curve pattern that will be modeled on the forecast.

It can be seen in figure 1 and figure 2 which is a comparison chart of the results of forecasting the daily PM$_{10}$ concentration between one model and another model, against the actual value in 2016. Based on the picture, almost all forecast models have the same pattern with the value of the value that is not too far towards the actual value, it's just for the output of the Quadratic Trend Model that has a higher estimated value than the other models.

![Figure 1. Comparison graph between Actual PM$_{10}$ Concentration and Forecasting Results, January - June 2016](image1)

![Figure 2. Comparison graph between Actual PM$_{10}$ Concentration and Forecasting Results, July - December 2016](image2)

It can be seen in table 4 and table 5 which is a verification of the forecast results of PM$_{10}$ concentrations from January to June. From the three forecast models, for January the MAD, MSE, and values were obtained, Lowest RMSE namely the Linear Trend Model with MAD values of 0.10, MSE 0.0001, and RMSE values of 0.011. Then for the lowest MAPE and Correlation values found in the Exponential Trend Model with a value of MAPE 108,769 with a correlation value of -0.16. in February, the MAD, MSE, and RMSE values for the three forecast models have the same values. Then the lowest MAPE value is in the Quadratic and Exponential Trend Model with the value MAPE 103,934. Furthermore, the highest Correlation value occurs in the Linear and Quadratic Trend Model with a Correlation value of 0.52. Verification of March forecast results, from the two models available between Linear and Quadratic Trend Model, the MAD, MSE, Lowest RMSE, and MAPE occur in the Linear Trend Model values with a MAD value of 0.008, MSE 0.0001, RMSE 0.011, MAPE 53.369. Then for the highest correlation value of the two models, namely the model Quadratic with a value of 0.17. In April, the values lowest MAD, MSE, and RMSE were in the Linear and Exponential Trend Model with MAD values of 0.11, MSE 0.0002, and RMSE 0.14. Then for the MAPE value, Lowest namely the Exponential Trend Model with a MAPE value of 50,298 with the value Correlation of 0.56. In May, the value and Exponential
Trend Model with a value of MAD 0.008 and an MSE value of 0.0001. Furthermore, the lowest RMSE and MAPE values are in the Linear Trend Model with RMSE 0.009 and MAPE values 62.673. Then for the lowest Correlation value occurs in the Quadratic Trend Model with the Correlation value -0.42.

| No | Month | Model          | MAD   | MSE   | RMSE  | MAPE      | Korelasi |
|----|-------|----------------|-------|-------|-------|-----------|----------|
| 1  | January| Linier Trend Model | 0.010 | 0.0001| 0.011 | 103.130   | -0.33    |
|    |       | Quadratic Trend Model | 0.030 | 0.0012| 0.034 | 275.523   | -0.17    |
|    |       | Exponential Trend Model | 0.011 | 0.0002| 0.012 | 108.769   | -0.16    |
| 2  | February| Linier Trend Model | 0.014 | 0.0002| 0.015 | 108.822   | 0.52     |
|    |       | Quadratic Trend Model | 0.013 | 0.0002| 0.015 | 103.934   | 0.52     |
|    |       | Exponential Trend Model | 0.013 | 0.0002| 0.015 | 103.934   | 0.51     |
| 3  | March | Linier Trend Model | 0.008 | 0.0001| 0.011 | 53.369    | 0.09     |
|    |       | Quadratic Trend Model | Error | Error | Error | Error     | Error    |
|    |       | Exponential Trend Model | 0.052 | 0.0037| 0.061 | 275.503   | 0.17     |
| 4  | April | Linier Trend Model | 0.011 | 0.0002| 0.014 | 50.607    | 0.53     |
|    |       | Quadratic Trend Model | 0.022 | 0.0007| 0.026 | 112.345   | -0.36    |
|    |       | Exponential Trend Model | 0.011 | 0.0002| 0.014 | 50.298    | 0.56     |
| 5  | Mei   | Linier Trend Model | 0.008 | 0.0001| 0.009 | 62.673    | -0.43    |
|    |       | Quadratic Trend Model | 0.016 | 0.0003| 0.019 | 127.849   | -0.42    |
|    |       | Exponential Trend Model | 0.008 | 0.0001| 0.010 | 65.840    | -0.43    |
| 6  | June  | Linier Trend Model | 0.018 | 0.0004| 0.020 | 166.699   | -0.01    |
|    |       | Quadratic Trend Model | 0.081 | 0.0083| 0.091 | 687.636   | 0.06     |
|    |       | Exponential Trend Model | 0.021 | 0.0005| 0.023 | 192.184   | 0.04     |

In June, the MAD, MSE, RMSE, and Lowest MAPE in the Linear Trend Model values were with MAD values of 0.018, MSE 0.0004, RMSE 0.020, and MAPE values of 166.699. Then for the highest correlation value, there is a Quadratic Trend Model with a correlation value of 0.06. In July, the MAD, MSE, Lowest RMSE, and MAPE occurred at Linear Values 0.0008, RMSE 0.027, and MAPE 180.780. Then for the lowest correlation value is at Quadratic Trend Model with value -0.31 correlation In August, the MAD, Lowest MSE, RMSE, and MAPE occurred in the Linear Trend Model values with MAD values of 0.058, MSE 0.0035, RMSE 0.059, and MAPE 389.221. Then for the lowest Correlation value on the Exponential Trend Model with a Correlation value of -0.02. In September, the lowest MAD, MSE, RMSE, MAPE values, and the highest correlation occurred in the Exponential Trend Model with MAD values of 0.018, MSE 0.0004, RMSE 0.019, MAPE 127.253, and Correlation of 0.09. In October, the lowest MAD, MSE, RMSE, MAPE values, and the highest correlation occurred in the Exponential Trend Model with MAD values of 0.001, MSE 0.0030, RMSE 0.124, MAPE 893.484, and Correlations of 0.61. In November, the lowest MAD value was found in the Exponential Trend Model with a MAD value of 0.001. Then the lowest MSE, RMSE, and MAPE values occur in the Linear Trend Model with MSE values 0.0001, RMSE 0.009, and MAPE 70.332. Furthermore, the highest correlation value occurs in the Linear Trend Model with a Correlation value of 0.45. In December, the lowest MAD, MSE, RMSE, MAPE values were found in the Linear Trend Model.
with MAD 0.007, MSE 0.008, and RMSE 91.003. Then for the highest correlation found in the Exponential Trend Model with a Correlation value of 0.33.

| No | Month | Model               | MAD  | MSE  | RMSE  | MAPE  | KORELASI |
|----|-------|---------------------|------|------|-------|-------|----------|
| 7  | July  | Linier Trend Model  | 0.010| 0.0001| 0.011 | 103.130| -0.33    |
|    |       | Quadratic Trend Model| 0.030| 0.0012| 0.034 | 275.523| -0.17    |
|    |       | Exponential Trend Model| 0.011| 0.0002| 0.012 | 108.769| -0.16    |
| 8  | August| Linier Trend Model  | 0.014| 0.0002| 0.015 | 108.822| 0.52     |
|    |       | Quadratic Trend Model| 0.013| 0.0002| 0.015 | 103.934| 0.51     |
|    |       | Exponential Trend Model| 0.013| 0.0002| 0.015 | 103.934| 0.51     |
| 9  | September| Linier Trend Model  | 0.008| 0.0001| 0.011 | 53.369 | 0.09     |
|    |       | Quadratic Trend Model| 0.052| 0.0037| 0.061 | 275.503| 0.17     |
|    |       | Exponential Trend Model| Error | Error | Error | Error | Error     |
| 10 | October| Linier Trend Model  | 0.011| 0.0002| 0.014 | 50.607 | 0.53     |
|    |       | Quadratic Trend Model| 0.022| 0.0007| 0.026 | 112.345| -0.36    |
|    |       | Exponential Trend Model| 0.011| 0.0002| 0.014 | 50.298 | 0.56     |
| 11 | November| Linier Trend Model  | 0.008| 0.0001| 0.009 | 62.673 | -0.43    |
|    |       | Quadratic Trend Model| 0.016| 0.0003| 0.019 | 127.849| -0.42    |
|    |       | Exponential Trend Model| 0.008| 0.0001| 0.010 | 65.840 | -0.43    |
| 12 | December| Linier Trend Model  | 0.018| 0.0004| 0.020 | 166.699| -0.01    |
|    |       | Quadratic Trend Model| 0.081| 0.0083| 0.091 | 687.636| 0.06     |
|    |       | Exponential Trend Model| 0.021| 0.0005| 0.023 | 192.184| 0.04     |

Figure 3 is the average PM$_{10}$ concentration in hours for 1 year. Based on the pictures above, it can be seen that the PM$_{10}$ concentrations highest are in the afternoon from 11 to 13 local time. Then for the PM$_{10}$ concentration lowest occurs at night at 20 to 05 morning.

According to Solih Alfiandy, the relationship between PM$_{10}$ concentration and air temperature, where the resulting correlation is positive, means that when sunlight starts to shine, the temperature of the air rises, the mass of air that settles will burst due to exposure to sunlight which causes the air temperature to rise. Because the mass of air that settles on the surface has broken out in the sun which causes the temperature of the air to rise, then the dust particles that settle will hover back in the air.

When viewed from its geographical location, the Kototabang Global Atmospheric Monitoring station is on a plateau with the surrounding environment being a forest with low air temperatures and high levels of humidity. So as can be seen in figure 3, the highest concentration is only at 11 to 13 noon where the sun is above the Global Atmospheric Monitoring station Kototabang, and for hours before and after the sun shines on the monitoring station PM$_{10}$ concentration has a value that does not occur fluctuation (up and down).

Based on the picture above it can be seen that the daily average PM$_{10}$ concentration reaches the highest value on Thursday (weekdays) with a value reaching 0.041 μg/m$^3$ while the lowest average concentration value is on Friday with an value average of 0.036 μg/m$^3$. The results show that Thursday has averaged the highest concentration where the day is a working day so there is a lot of activity motorized, offices that cause a PM$_{10}$ concentration higher than on weekends.
4. Conclusion
It is known that the results of the statistical model analysis for forecasting PM$_{10}$ concentrations at Kototabang Global Atmospheric Monitoring Station is using Linear Trend Model based on data for forecasting 2015. Forecasting is conducted monthly per month in 2016, to anticipate the inequality between data in other months. There is no uniformity in the sense of differences in seasons, rainy seasons and dry seasons. Aerosol particles that float in the air during the rainy season are less than in the dry season because during the rainy season aerosol particles that hover will dissolve to the surface with rainwater that falls. Therefore, to get good results, it is necessary to separate between wet months and dry months. One way is to do forecasting month by month, for example; January for January forecasting and so on. From the results of the verification of PM$_{10}$ concentration forecasting models, the best model is Linear Trend Model with the best specifications in January, March, May, June, July, August, November and December. Then for the next best model, the Exponential Trend Model with specifications, namely April, September, and October. Furthermore, the Quadratic Trend Model is best used, namely in February. In selecting the model for forecasting, the writer can conclude that if the chart pattern time series on the curve is linear, then use the Linear Trend Model. If the pattern on the curve is parabolic, then choose the Quadratic and Exponential Trend Model. Based on the three models forecasting, the best model for forecasting the concentration of PM$_{10}$ obtained at Atmospheric Monitoring Station Kototabang Global is the Linear Trend Model.

References
[1] Sulistiyo A et al 2016 Megasains 7 34-39
[2] Baroto T 2002 Perencanaan dan pengendalian produksi Ghalia Indonesia: Jakarta
[3] Ginting R 2007 Sistem produksi Publisher: Graha Ilmu, Yogyakarta
[4] Hyndman R J, Koehler A B, Synder, R D and Grose S 2002 International Journal of Forecasting 18 439-454
[5] Makridakis S, Wheelwright S C and Hyndman R J 2008 Forecasting methods and applications John Wiley & Sons: New York
[6] Makridakis S, Wheelwright S C and McGee V E 1999 Metode dan Aplikasi Peramalan (U.S. Andriyanto dan A. Basith, terj.) Jakarta: Erlangga.
[7] Nasution A H 2015 Perencanaan & pengendalian produksi PT Candimas Metropole First Edition, Jakarta
[8] Sukerti N K 2015 Peramalan Deret Waktu Menggunakan S – Curve dan Quadratic Trend Model Proc. National Conference System & Information: STMIK STIKOM Bali, 9 – 10 October 2015
[9] Alfiandy S 2018 Estimation of PM$_{2.5}$ Daily Concentration In January – June 2018 By Using Principal Component Regression Method in South and Central Jakarta Undergraduate Dissertation Climatology Study Program: The School of Meteorology, Climatology and Geophysics,.