M – 27

Rainfall Forecasting Using Bayesian Nonparametric Regression

Suwardi Annas[1], Rizwan Arisandi[2]
[1] Department of Statistics, State University of Makassar
[2] Department of Statistics, State University of Makassar
suwardi_annas@yahoo.com

***Abstract*—** In the present years, climate change due to global warming, resulting in the change of seasons in Indonesia is high variability and unpredictable. Many methods that can be used to predict rainfall pattern, such as parametric regression and ARIMA. However, the model obtained through parametrics statistical approach only concerned to information of samples, therefore, it is poor to interpret the parameters of the rainfall pattern. This study proposes a bayesian nonparametric regression with Gaussian Regression Process approach for rainfall forecasting in the City of Makassar, Indonesia. Based on the value of Root Mean Square Error Prediction (RMSEP), the best covariance function that can be used to forecast is quadratic exponential.

Keywords: *quadratic exponential, rainfall forecasting, regression gaussian process, RMSEP*

I. INTRODUCTION

In the latitude and longitude, the characteristics of equatorial and monsoonal circulations in Indonesia are very different. These natural conditions resulting the rainfall in Indonesia are very unstable, complex, and have high variability. This climate change, due to global warming, also have resulted irregular change of seasons in various regions in Indonesia including the city of Makassar. This natural phenomenon causes that the currently rainfall is very difficult to predict accurately with traditional forecasting methods. Even in climatology, rainfall in Indonesia has become one of the most difficult factor to be predicted accurately.

Along with the development of modern technology, technology of rainfall forecasting has also developed rapidly, ranging from deterministic approach to the stochastic approach (Sutikno et al, 2010). The deterministic approach conducts through an analysis based on physical laws expressed in mathematical forms by using the classical statistics approaches such as (ARIMA), Fourier analysis, analysis of Kalman Filter and other methods. These forecasting methods are widely used to identify the relationship between rainfall and temperature, air pressure, wind velocity, air humidity, and solar radiation intensity (Pramudia *et al*, 2008).

Besides those approaches, some researchers concerning to climate and weather have also developed a model of rainfall prediction based on nonparametric models and based on neural networks (Estiningtyas and Kharmila 2008, Subarna 2009, Warsito and Sri, 2007). Nevertheless, the parameter estimations of model and their inferences only concern to the information from the samples obtained and ignore the prior information from researchers or field workers. According to Casella and Berger (2002), this classical approach has drawbacks in term of interpretation of the confidence interval of a parameter model. The significance of the confidence interval of a model parameter can not be based on the real level setting up before building the model.

To overcome the problem, a predictive model that can be used is Nonparametric Bayesian approach i.e. Gaussian Process Regression (GPR). The advantage of nonparametric models lies in the flexibility form of the model, in particular there are no assumptions on the parametric form (Eubank, 1999; Takezawa, 2006). Bayesian models can accommodate the researchers' prior information. Prior information is usually quantified into distribution form the parameters or the functions (Box and George, 1973).

Some researchers had used the Gaussian process regression approach to develop a model from the cases faced. Williams and Rassmusen, *et al* (2006) used the Gaussian process regression to obtain a model on the robot arm motions. In addition, Chen, *et al* (2007) used the Gaussian process regression to develop a calibration model on the spectroscopic data. Therefore, this study proposes the use of Gaussian processes regression to forecast the rainfall in the city of Makassar, Indonesia. The results of this study are expected to be able to generate a good estimation value of Gaussian Process Regression model parameter so that it produces an accurate rainfall forecasting.

II. MATERIAL AND METHOD

This study uses data of 120 observations obtained each month for last 10 years, from January 2006 to December 2015. The data are obtained from Indonesian Agency for Meteorology, Climatology, and Geophysics of province of South Sulawesi. The data used consist of one dependent variable that is the amount of rainfall (Y), and four independent variables namely; air humidity (X1), air temperature (X2), air pressure (X3), and wind velocity (X4).

The definitions of each variable used in this forecasting include; (1) rainfall is defined as the amount of water that falls on the flat land surface on certain period measured in height unit (mm) none of evaporation over horizontal surface, runoff and infiltration, (2) air humidity is the amount of water vapor content or water vapor concentration in the air measured in percentage (%), (3) air temperature is the level or degree of heat from the activity of molecules in the atmosphere or the size of kinetic energy average from the movement of molecules in the atmosphere measured in Celsius unit (^{0}C), (4) the air pressure is the pressure appeared by weight of air layer or pressure working to move air mass in each certain area unit measured in millibar (mb), and (5) wind velocity is the speed of air moving horizontally at a height of two meters above the ground measured in knot.

In analyzing the data, researcher conducts several procedure stages: the first, identify distribution patterns of Gaussian of variables used in the study, the second, make Gaussian process regression model with a review and an assignment of covariance function that will be used, the third, validate the result model of stage 2 with the criteria of Root Mean Square Error Prediction (RMSEP), and the fourth, make a prediction of the amount of monthly rainfall in 2016.

III. RESEARCH RESULTS

A. *Graphic Desciption of Each Variable*

Figure 1 is a graph of the dynamics of monthly rainfall from 2006 to 2015. One millimeter of rainfall means that one square meter area on the flat land accommodates as high as one millimeter of water or as much as one liter of water. Rainfall intensities are the amount of rainfall in certain time period.

Big intensity means heavy rain and this condition is very dangerous because the impact can cause flood, landslide, and drawback effect to the plants. Short-term rainfall is stated in per hour intensity called as rainfall intensity (mm/hr). The rainfall intensity average is in ½ hour. The amount of rainfall intensities are different due to rainfall durations or occurrence frequencies. Varieties of time distribution of a rainfall records can give signs of the increasing or decreasing of tendency. For the monthly rainfall occurring based on the graph, there is no repeated frequency in a year, so that it does not make a seasonal distribution. It shows that the diversity of the amount of annual rainfall is quite high. In addition, standard deviation reaches 273.1243 mm. The highest monthly amount of rainfall occurs in January of 2013 reaching 982 mm.

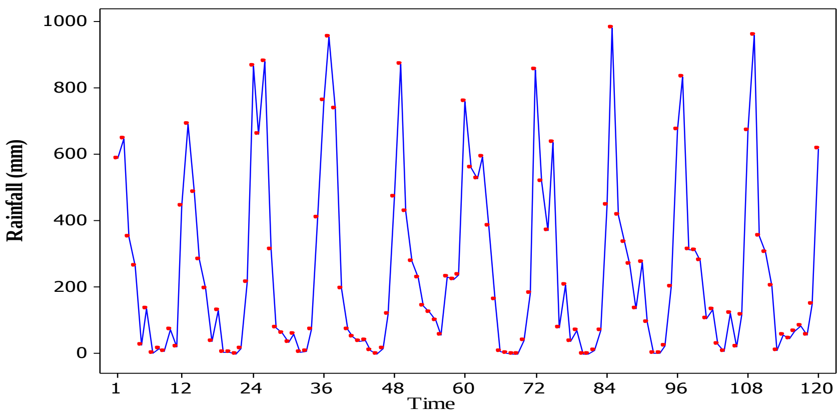

FIGURE 1. GRAPH OF THE MONTHLY RAINFALL IN THE CITY OF MAKASSAR IN 2006-2015

Furthermore, the dynamics of air humidity are presented in Fig. 2., which is the graph of the dynamics of air humidity are from 2006 to 2015. It can be seen that the air humidity each month at the early year tends to be high, while on August and September tend to decrease. The increasing and decreasing of air humidity is relatively low because the city of Makassar is in the equatorial area having tropics climate. Generally, the annual air humidity is in the range 67% to 97%. It states that the tendency of the temperature is above of the temperature average i.e. 79%.

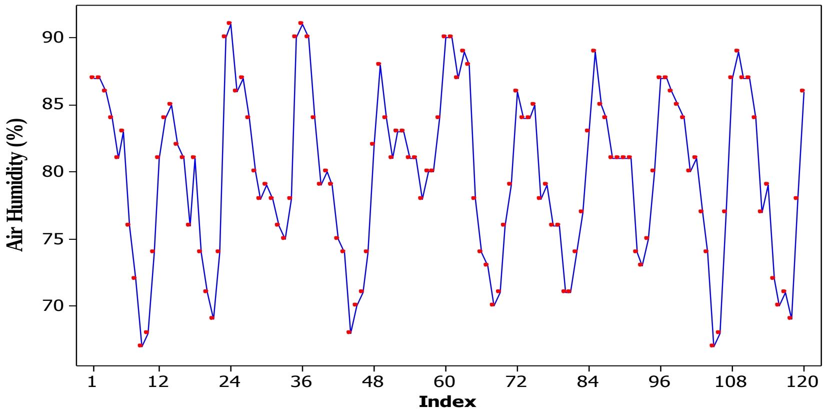

FIGURE 2. GRAPH OF THE AIR HUMIDITY IN THE CITY OF MAKASSAR IN 2006-2015

The dynamics of the air temperature average from 2006 to 2015 is revealed in Fig. 3. It shows that the annual air temperature average is erratic each month, but the annual air temperature average is generally in the range of 26.3 ^{0}C to 29 .4 ^{0}C. It states that the tendency of the temperature is above the average i.e. 27.78 ^{0}C.

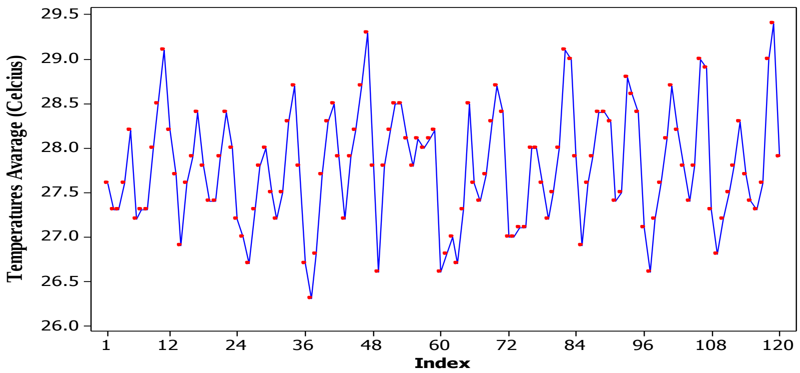

FIGURE 3. GRAPH OF THE AIR TEMPERATURE AVERAGE IN THE CITY OF MAKASSAR IN 2006-2015

The dynamic of air pressure from 2006 to 2015 is shown in Fig. 4. It shows that the dynamics of air pressure is in addictive pattern. The kinds of data has increased every year from the early 2012, but the annual air pressures are generally in the range of 1008 mb to 1013 mb. It reveals that the tendency of air pressure is above the air pressure average reaching 1010.72 mb.

ISBN 978-602-74529-0-9

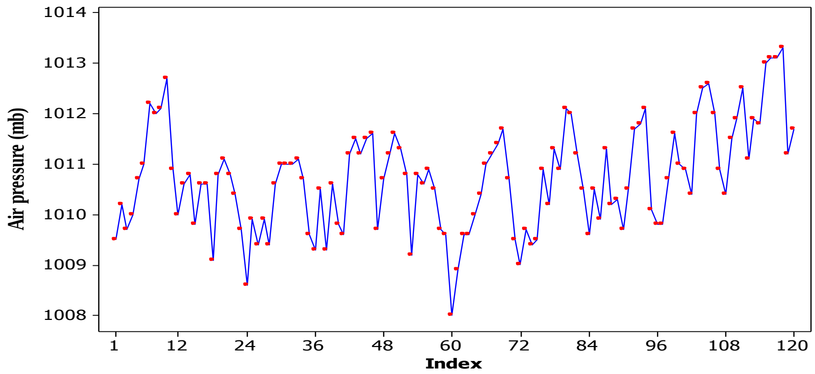

FIGURE 4. GRAPH OF THE AIR PRESSURE IN THE CITY OF MAKASSAR IN 2006-2015

Figure 6 is a graph of the dynamics of the wind velocity from 2006 to 2015. The dynamics of wind velocity tends to be almost the same from 2011 to 2015, unless the early 2014 the wind velocity is increasing. But the annual wind velocity are generally in the range from 2 knots to 7 knots. It shows that the tendency of the wind velocity is above the average air pressure i.e. 4.5 knots.

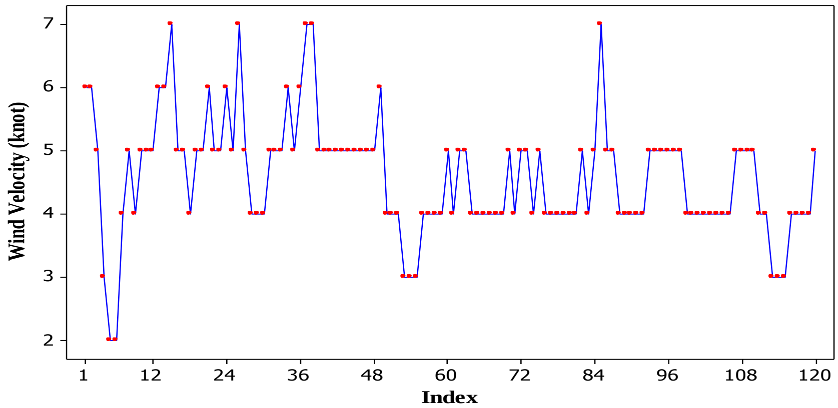

FIGURE 5. GRAPH OF THE WIND VELOCITY IN THE CITY OF MAKASSAR IN 2006-2015

B. Selection of Covariance Function

A crucial aspect of modeling Gaussian process regression is selection of covariance function. This study uses a Quadratic Exponential covariance function. Hyper-parameter values are estimated by using marginal maximum likelihood method. Table 1 shows the estimated value of the hyper-parameters of covariance function.

The covariance functions examined in this study are: (1) the function of Quadratic Exponential Covariance with distance measurements of Automatic Relevance Determination (QE-ARD), (2) the Linear covariance function with the Automatic Relevance Determination parameters (Linear-ARD), (3) the Linear covariance function with single hyper-parameter (Linear-1), and (4) Quadratic Exponential covariance function with the Isotropic distance measurement (QE-Iso).

The covariance function is conducted with concerning to the value resulted on RMSEP. After the covariance functions above are attempted to be used in the Gaussian process regression modeling, the results are shown at Table 1. It can be concluded that the covariance function, which is relevant to smallest value RMSEP, is obtained when the Gaussian process regression model uses quadratic exponential covariance function (QE-ARD), which is 137.0867. Furthermore, the covariance function will be used to predict the amount of rainfall in 2016.

TABEL 1. ESTIMATION VALUE OF HYPER-PARAMETER OF COVARIANCE FUNCTION AND RMSEP

Covariance Function	Hyper-parameter	Parameter Estimation	RMSEP
Quadratic exponential (QE- ARD)	Length scale 1 Length variance 2 Signal variance Error variance	11.285 10.185 8074.285 16034.730	137.087
Linear ARD	Bias parameter Controller 1 Bias parameter Controller 2 Error variance	10.215 10.305 20798.480	142.357
Linear-1	Bias parameter Controller Error variance	10.225 20798.48	142.327
quadratic exponential-Iso	Length scale Signal variance Error variance	8981.035 11.445 16227.670	139.857

C. Rainfall Forecasting in 2016

Before performing a prediction of the amount of monthly rainfall during 2016, the first stage is making predictions of the air humidity, the air temperature average, the air pressure, and the wind velocity during 12 months in 2016. In this case, the method used is the moving average with the period 3. The result of the prediction is shown at Table 2.

The prediction about air pressure is revealed at Table 2, the highest values is on September and October in 2016 with 1011.6 mb and the lowest value is in February 2016 reaching 1009.1 mb, the highest monthly temperature average is 28.7 ^{0}C falling in September 2016 and the lowest temperature average is in October 2016 with 26.6 ^{0}C. Based on the prediction, the highest value of air humidity is in August 2016 with 91% and the lowest value -which reached 71%- is in October 2016, while the highest value of the wind velocity based on the prediction is in May 2016 reaching 7 knots and the lowest value is in January 2016 reaching 2 knots.

TABEL 2. ESTIMATION VALUE OF HYPER-PARAMETER OF COVARIANCE FUNCTION AND RMSEP

Month	Air humidity (%)	Temperature average (^{0}C)	Air pressure (mb)	Wind velocity (knot)
January	83	27.2	1011.0	2
February	81	27.8	1009.1	4
March	90	28.0	1009.7	5
April	86	27.0	1009.9	5
May	87	26.7	1009.4	7
June	78	27.2	1011.0	5
July	75	28.3	1011.1	5
August	91	26.7	1009.3	6
September	79	27.7	1010.6	5
October	71	28.7	1011.6	5
November	88	26.6	1011.2	6
December	83	28.5	1009.2	3

The prediction results of air pressure, temperature average, humidity and wind velocity in 2016 will be the input for the interest amount of rainfall prediction in 2016 by using a Gaussian process regression model. Prediction of rainfall for 12 months using Gaussian process regression model shows at Table 3. Table 3 shows that the highest amount of rainfall prediction on the February reaching 787.2 mm with diversity 881 mm and the lowest in July 2016 reaching 38.3 mm with diversity 6 mm.

TABEL 3. ESTIMATION VALUE OF HYPER-PARAMETER OF COVARIANCE FUNCTION AND RMSEP

Month	Rainfall	Standard deviation
January	626.2	137
February	787.2	881
March	110.4	251
April	455.9	662
May	287.2	130
June	62.0	35
July	38.3	6

August	747.7	764
September	151.8	197
October	39.6	16
November	700.6	873
December	544.5	114

IV. DECISSION

Nowadays, the rainfall forecasting has also developed rapidly beginning from the deterministic approach to the stochastic approach. Gaussian process regression is one of methods that can be used to do that. This method is classified in category of non-parametric Bayesian regression model where the model specification connecting between dependent variable and independent variables is not needed to be set in advance. The data are allowed to "speak to themselves" to form the relevant structure models. Gaussian process regression using stochastic approach assumes that the amount of rainfall is random. Based on the value of Root Mean Square Error Prediction (RMSEP), the best covariance function which can be used to predict is Quadratic Exponential Automatic Relevance Determination parameters (QE-ARD) with RMSEP i.e. 137.0867. The prediction of the highest amount of rainfall is on February reaching 787.2 mm with 881 mm and the lowest amount of rainfall is in July 2016 i.e. 38.3 mm with 6 mm. Compared to other classical methods that are not able to accommodate the initial information and not to impose the modeling, Gaussian process regression approach is relatively good to be used.

REFERENCES

[1] Badan Pusat Statistika Kota Makassar, 2010. Makassar Dalam Angka 2010 (Makassar in Figure 2010). UD Areso, Makassar.

[2] Box G.E.P and George C.T., Bayesian Inference in Statistical Analysis, Addison-Wesley Publishing Company, Inc. Canada, 1973.

[3] Casella G. and Berger R.L., Statistical Inference, Second Ed. Thomson Learning, Duxbury, 2002.

[4] Chen T, Morris J, Martin E. "Gaussian Process Regression for Multivariate Spectroscopic Callibration." Chemometrics and Intelligent Laboratory Systems, 2007, 87: 85-97.

[5] Eubank R.L., Nonparametric Regression and Spline Smoothing, Second Ed. Marcel Deker, Inc. New York, 1999.

[6] Estiningtyas W., Elsa S., dan Kharmila S.H., "Penyusunan Skenario Masa Tanam Berdasarkan Prakiraan Curah Hujan di Sentra Produksi Pangan," Jurnal Meteorologi dan Geofisika, 2008, Vol. 9, No.2: 65-77

[7] Pramudia A., Yonny K., Irsal L., dan Tania J., "Pewilayahan Hujan dan Model Prediksi Curah Hujan untuk Mendukung Ananlisis Ketersediaan dan Kerentanan Pangan di Sentra Produksi Padi," Forum Pascasarjana, 2008, Vol. 31, No. 2: 13 1-142.

[8] Rasmussen C.E., Evaluation of Gaussian Processes and Other Methods for Non-linear Regression (Dissertation), Department of Computer Science, University of Toronto. 1996.

[9] Subarna D., "Aplikasi Jaringan Neural untuk Pemodelan dan Prediksi Curah Hujan," Berita Dirgantara, 2009, Vol.10, No.1: 13-18.

[10] Sutikno, Rokhana D.B, Putri S, dan Istriana, "Prakiraan Cuaca dengan Metode Autoregressive Integrated Moving Average, Neural Network dan Adaptive Splines Treshold Autoregression di Stasiun Juanda Surabaya, " Jurnal Sains Dirgantara, 2010, Vol. 8, No. 1: 43-61.

[11] Takezawa K., Introduction to Nonparametric Regression, John Wiley & Sons, New Jersey, 2006: 1-15.

[12] Warsito B. dan Sri S., "Prediksi Curah Hujan Kota Semarang dengan Feedforward Neural Network Menggunakan Algoritma Quasi Newton BFGS dan Levenberg¬Marquardt," Jurnal Presipitasi, 2007, Vol. 3, No.2: 60-64.