IOP Conference Series: Earth and Environmental Science

An Evaluation Graph of Hourly Rainfall Estimation in Malang

An Evaluation Graph of Hourly Rainfall Estimation in Malang

R M Putra [1], A Kurniawan[2], I A Rangga[2], M Ryan[2], Endarwin[1] and A Luthfi[1]

[1]Agency for Meteorology, Climatology, and Geophysics. Angkasa I Street No.2, Kemayoran, Capital Jakarta, 10620 Indonesia
[2]State College of Meteorology Climatology and Geophysics. Perhubungan I Street No.5, Pondok Betung, South Tangerang, 15221 Indonesia

richardmp.rmp@gmail.com

Abstract. Satellite-based rainfall estimation is evolving rapidly. Most studies use data, which is spatially fine, but poorly regarding time. On the other hand, availability of verification data is also quite rare. This study used Hillman Form B report that was corrected by ME-48 from Malang Climatological Station. 2009-2016 IR1 satellite data were used in hourly temporal resolution (only less than 3% data missing). Four estimation methods were compared: Auto Estimator, CST, mCST, and Quantile Analysis Equation. Data processing was carried out using Python and R statistic as a quality control. The analysis was done by creating a graph that combines False Alarm and Miss Information for each rainfall intensity. Binary transformation was done for enabling information to be plotted. All rainfall estimation methods have a high false alarm (more than 74% at 1 mm) but quite low miss (less than 0.03%). By taking into account its error pattern, satellite data can be used in rainfall observation. The Quantile equation is slightly superior to other methods. This study is relatively inexpensive to be duplicated so it can be used as an evaluation tool for rainfall estimation best practice for Meteorological and Climatological Agency's network.

Keywords : rainfall, satellite, CST, Malang.

1. Introduction

One of the most significant weather elements [1] for all human activities is rainfall. Rainfall is the intensity of rain that occurs at a certain duration of time [2]. Rainfall information is expressed in millimeters (mm). Rainfall of 1 (one) millimeter means that the area of 1 square meter has a water level of 1 millimeter or equal to one litter (1 dm^3) assuming that it does not evaporate, does not seep and does not flow [3]. Bulk information is needed primarily for early warning of natural disasters caused by hydro-meteorological factors such as floods, landslides, and droughts as well as in research related to the hydrological cycle [4]. However, the current problem is the uneven distribution of rain gauge [5] so that many locations do not have this information. One alternative solution is to estimate the rainfall intensity using satellite data [6].

Rainfall estimation techniques that have been studied previously produce relatively good values of rainfall estimation based on satellite imagery. Swarinoto and Husein conducted research on satellite data capabilities to estimate rainfall in the Jayapura area and its surrounding area [7] . Other research [8] uses GOES Infrared for making rainfall estimation using regression fit of radar rain rate and GOES – 8 Cloud Top Temperature. Another technique for estimating rainfall using satellite is by clouds classification

IOP Conf. Series: Earth and Environmental Science **303** (2019) 012031 doi:10.1088/1755-1315/303/1/012031

based on convective and stratiform types [9, 10]. This technique was improved by [11] by using the passive microwave (PMW) to increase performance on rainfall estimation in Indonesia.

This study aims to evaluate the results of rainfall estimation using satellite data for Malang region. In addition, the aim is to identify the best methods for rainfall estimation in Malang and to address problem in Quality Control (QC) before conducting research.

2. Data and Methods

The data used in this research consists of three datasets from January 2009 – December 2017. These data are rainfall observation from Climatological Station in Malang every 3 hours (Form ME.48 Meteorological Archive), rainfall data every 1 hour from Hellman rainfall observation (Form.B Meteorological Archive), and Himawari – 8 Satellite from Kochi University (http://weather.is.kochi-u.ac.jp/archive-e.html) with Portable Gray Map (PGM) data format. Kochi University's satellite data have an hour temporal resolution and have a spatial resolution of about 5.5 km as shown in figure 1.

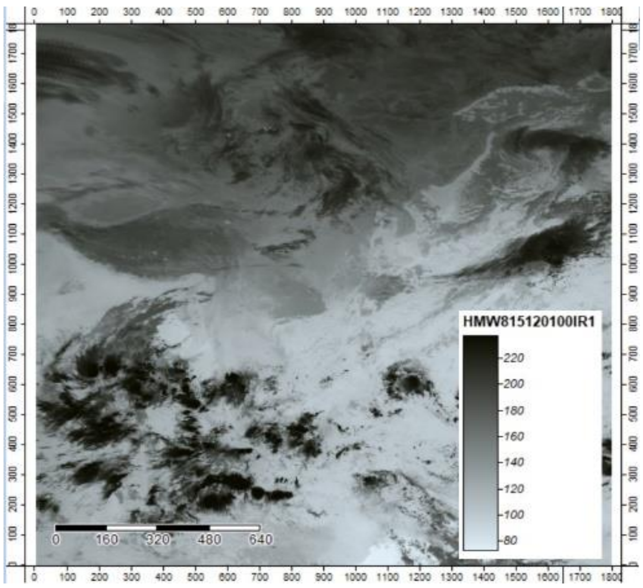

Figure 1. Satellite data from Kochi University in PGM Format

In the PGM format, the Cloud Top Temperature (CCT) cannot be directly obtained from the data. This is because the Portable Gray Map (PGM) format only shows relative CTT based only on the grayscale. The minimum color code is zero, and the maximum is 255. Therefore, in addition to requiring PGM format satellite image data; this study also used CAL data to convert grayscale from PGM into the cloud top temperature (CTT) in Kelvin (K) units.

This research was conducted in the region of Malang Climatological Station located in -7.9^{o} S and 112.6^{o}E. Unfortunately, the pixel position of Malang Climatological Station is at the center of four pixels' satellite from Kochi University. Therefore, in this research, we also tried to find the best pixel for making rainfall estimation in Malang Climatological Station using Himawari – 8 from Kochi University.

Because of that condition of satellite pixel, in this research, we make five schemes for processing the satellite rainfall estimation by using pixel 1, pixel 2, pixel 3, pixel 4 and the average of all pixels around Malang Climatological Station as shown in figure 2. The methods used in this research consist of 4 rainfall estimation methods, i.e. Convective Stratiform Technique (CST), modified Convective Stratiform Technique (mCST), Auto Estimator (AE), and Quantile Analysis formed by regression function of sorted rainfall intensity in Malang for 2009 – 2017 and cloud top temperature (CTT) in that locations using 5 schemes of pixel.

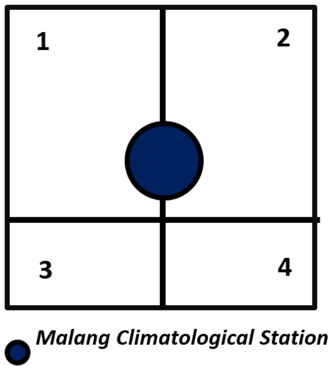

Figure 2. Location of Malang Climatological Station in Pixel's Satellite

2.1. Convective Stratiform Technique (CST) and modified Convective Stratiform Technique (mCST)
The methods used in this research are similar to the procedures proposed by [9, 10]. The technique classifies cloud types into convective or stratiform cloud by using the difference between CTT in the pixel of the study's area with that in the surrounding pixels. [12] Also used brightness temperature difference between the pixel and the eight surrounding neighbour pixels. In this research, we classify convective and stratiform cloud by using eight pixels in two form, referred as "Square" form and "Plus" form as illustrated in figure 3.

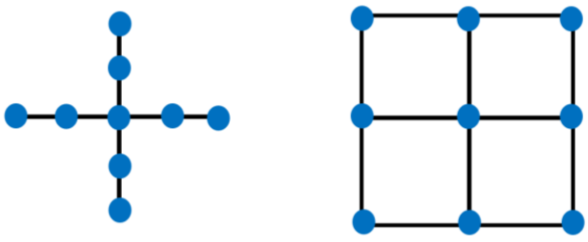

Figure 3. Classification technique for Convective and Stratiform Cloud by "Plus" (a) and "Square" (b) form

The technique of classification whether convective and stratiform cloud is determined by the Slope Parameter (S) using eight pixels around the main location:

$$S = k(T_{i-2,j} + T_{i-1,j} + T_{i+1,j} + T_{i+2,j} + T_{i,j-2,} + T_{i,j-1} + T_{i,j+1} + T_{i,j+2} - 8T_{i,j}) \quad (1)$$

With k = 0.125
Where the values i and j are pixels indices around the center (-7.6 LS and 112.6 BT), while k is a factor that depends on data resolution. Furthermore, the Slope Parameter values that indicate the type of convective cloud must meet the following requirements:

$$S \geq \exp[0.0826\,(T_{min} - 207)] \quad (2)$$

With T_{min} = cloud top temperature (CTT) in Kelvin units

If the slope parameter value does not satisfy the criteria, then the type of is stratiform cloud [11]. The next step to estimate the rainfall is to determine the area of rain. The area of cloud that produce rainfall are determined by the following equation:

$$A = \exp(15.27 - 0.0492 \; x \; T_{min}) \quad (3)$$

The next step is to estimate rainfall from the convective and stratiform cloud by using these equations:

$$Convective\ rainfall\ (mm) = C \left(\frac{A_c}{A}\right) TR_c \quad (4)$$

$$Statiform\ rainfall\ (mm) = S \left(\frac{A_s}{A}\right) TR_s \quad (5)$$

Where c = number of convective cells within a grid, s = number of stratiform cells within a grid, A_c and A_s = rain area from Eq. (3), A= average area covered by each pixel, T = length of period in hour, R_c = convective rain rate in mm h^{-1} (20 mm h^{-1}), R_s = stratiform rain rate in mm h-1 (3.5 mm h^{-1}). The differences between CST and modified CST are the coefficient values of A, R_c, and R_s. [11] has modified this method by the result of best rainfall intensity from the convective cloud (R_c) is 26 mm h^{-1} and stratiform cloud (R_s) is 0.8 mm h^{-1}.

2.2. Auto Estimator

This technique uses regression fit of radar rain rate and GOES – 8 Cloud Top Temperature [8]. Rainfall estimation is calculated by using cloud top temperature (CTT) data from a satellite then converted into rainfall intensity by using the following equation :

$$R = 1.183 \; x \; 10^{11}. \exp(-3.6382 \; x \; 10^{-2} x T^{1.2}) \quad (6)$$

With R = rainfall intensity (mm h^{-1}) and T = cloud top temperature in Kelvin units

2.3. Quantile Analysis Equation

In this research, we did not only compare the rainfall estimation from the equations that were studied previously, but also tried to derive equation by regression [13] of sorted rainfall intensity and sorted cloud top temperature called "Quantile Analysis" method. We have eight schemes for this equation which depend on the pixel that we choose from the satellite data. The table below shows all schemes to estimate the rainfall intensity from regression data:

Table 1. Quantile Analysis Equation

Rainfall = B*exp(CTT*A), CTT=Cloud Top Temperature, A and B = Coefficients

Coeff.	Quantile Upper Limit			
	P1	P2	P3	P4
A	-0.15561	-0.15940	-0.15605	-0.15446
B	3.5494E+14	8.0260E+14	4.0251E+14	2.8841E+14
Coeff.	Quantile Lower Limit			
	P1	P2	P3	P4
A	-0.15792	-0.16023	-0.15813	-0.15736
B	5.1820E+14	8.8531E+14	5.8972E+14	5.0726E+14

The coefficient values of A and B are obtained from the regression equation of sorted rainfall intensity and sorted cloud top temperature. However, the condition sometimes has a problem with this pattern from rainfall data and cloud top temperature, which have the same value of rainfall intensity but the cloud top temperature more than one conditions. In this case, we use two approaches namely upper limit and lower limit. Upper limit means we choose the highest data of cloud top temperature for the same-sorted rainfall intensity and lower limit means we choose the lowest data. Therefore, for this condition, we have eight schemes that consist of four pixels and two conditions of approaches.

2.4. Double Checked Analysis

The rainfall intensity from satellite data is double-checked by both software "R" [14] and "Python"[15]. This step is to ensure that both software agree with each other. The main software for calculating the rainfall intensity from satellite data is "Python" and the result of calculation is compared to result from software "R" for Quality Control. Furthermore, the plotting was also done using QGIS software [16] [17] to ensure that 4 pixels of satellite data are in the right position. In comparison, "R" software has a better configuration script and performance. Therefore, python calculation for rainfall intensity was checked and compared with R calculation. Python have a better processing time than that of R, so we use Python software and checked by R software. Based on this case, it can be concluded that if we use one software, it has to be checked by the other software to ensure that our calculation and script's programming are working well.

2.5. False Alarm and Miss

In this research, verification of rainfall estimation method used 2 x 2 contingency table [18]. In this method, we will calculate the False Alarm and Miss condition for all rainfall estimation's scheme that we used. The previous research from [19, 20, 21], also used FAR for validation.

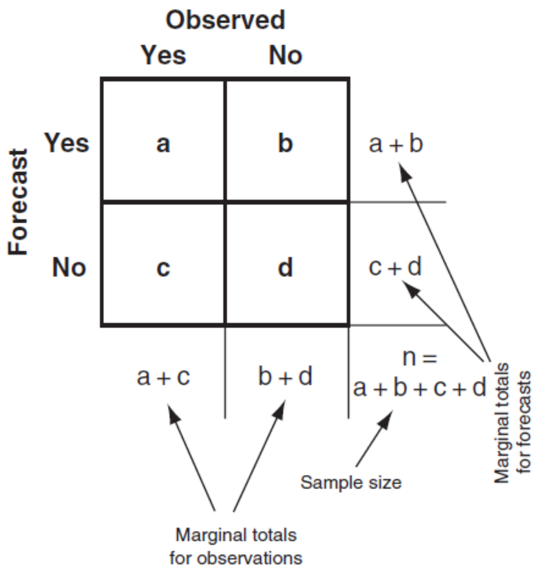

Figure 4. 2 x 2 contingency table (Source: [18])

In term of figure 4, *a, b, c* and *d* are the number of pairs for estimation data and observed data. When the estimation data is same class value with observed data, it is called *hits* (*a*). Similarly, on *b* occasions called *false alarm* (FAR). It is a condition of occurring event in estimation, but did not happen in observed data. There is also *c* that show the number of estimations did not occur, but in observed data is happen. It is called *miss* (MS). The last symbol is *d* means the same condition of estimation and observed show the event does not occur.

In this research, validation using FAR and MS values were done for any rainfall intensity. They are 1 mm, 2 mm, 3 mm, 4 mm, 5 mm, 7 mm, 9 mm, 11 mm, 15 mm, 17 mm, 20 mm, 25 mm, and 30 mm. Hereafter, we plot all of FAR and MS values into one graphic called FAMS picture and analysed it.

3. Result and Discussion

3.1. Data completeness of rainfall observation

The first step before we calculate rainfall intensity using satellite data and compare which one is best method, we want to discuss on completeness of data from 2009 – 2016. The problem that we have to know first is about the different type of report for rainfall intensity before and after 2015. Before 2016, the rainfall intensity of Hellman (Form.B) is reported on the same day as the rainfall occurred. However, starting in 2016, rainfall intensity was reported the day after rainfall occurred. This problem was found by comparing the rainfall intensity every 3 hour from ME.48 and every 1-hour from Hellman observation (Form.B). This figure below is the ratio of corrected data for rainfall intensity caused by that problem:

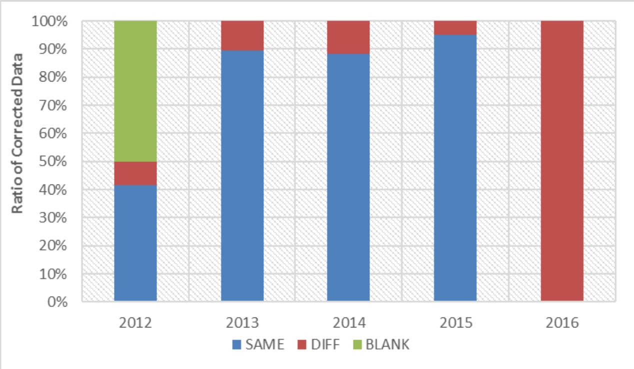

Figure 5. Ratio of corrected data

Based on figure 3, we know that before performing the calculation, we must be performing Quality Control (QC) of our input data. The reason of why data analysis started in 2012 is due to the fact that digitalization of hourly rainfall intensity was only started from this year. Also, in 2012 data is not complete for one year. The green color indicates blank data of rainfall observation. The red color shows different type of report for rainfall observation, every 3 hours and every 1 hour. Also, the blue color shows the same report of rainfall observation every 3-hour and 1 hour based on a different kind of meteorological form. In this discussion, we want to show that rainfall data observation has been verified and has a good quality before we compare them with rainfall estimation method using satellite.

3.2. Data completeness of Satellite data from Kochi University

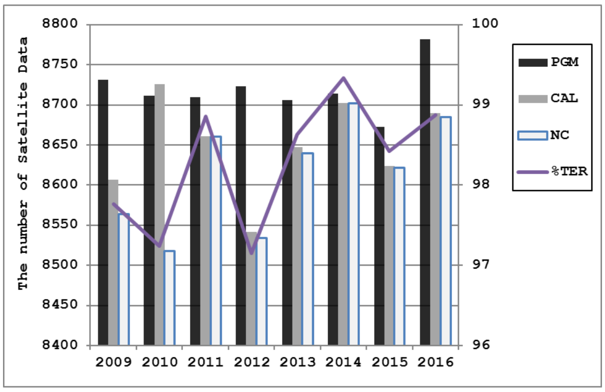

Figure 6. Completeness of satellite data from Kochi University

Satellite data from Kochi University has a different condition with rainfall data observation. Start from 2009 – 2016; the satellite data is existing in PGM format and for Calibration data (CAL). However, since rainfall observation data in digital format was only existing from the middle of 2012, this research is only done when both satellite and observation data exist. Satellite data from Kochi University in PGM format and CAL format was converted into NetCDF (NC) format first for calculating rainfall intensity easily. However, the problem was that we could not convert all of the data to rainfall intensity. So, figure 4 shows the number of data successfully converted to NC format and then rainfall estimation. The highest number of data that could be converted to NC is in 2014; it is almost 100% of data. The lowest number of data that could be changed is 2010, it is about 97.2% of data. However, for all of the data used in this research, successful data conversion are higher than 97% for all of the data.

3.3. Comparison of All Schemes of Rainfall Estimation Using Satellite Data

In figure 7, we plot all of the value for *false alarm* and *miss* for all schemes. Due to its similar result, only best scheme is plotted in black. *False alarm* and *miss* have its own best scheme. Where there is rainfall estimation with high intensity, there will be high *false alarm*. The *false alarm* is about only 0.7 in separating rain and no rain condition. It means that statistically, for every ten times rainfall estimation say there will be a rain, only three times of rains will be true. The best method of rainfall estimation for all schemes based on *false alarm* value is Quantile Analysis in pixel 3 using end time of observation.

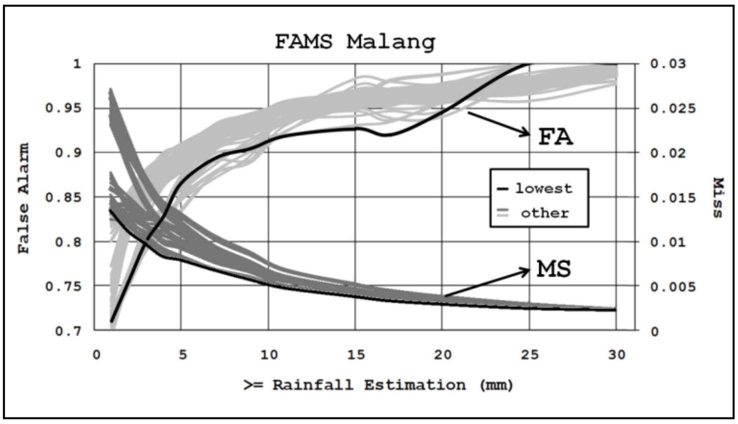

Figure 7. All experiment result of rainfall estimation methods

Miss condition is also shown in figure 7 with axis on the right side. For all methods, miss value is less than 0.03 % and it means that for rainfall estimation, all schemes have a good performance so that if there is condition of rain that declared by someone lie, rainfall estimation is good enough to negate it. Based on the result, the best method for rainfall estimation in Malang based on *miss* value is CST calculated by "square" method for separating cloud type. We can also know that the best rainfall estimation was obtained using pixel that located southwest of Malang coordinate (pixel 3).

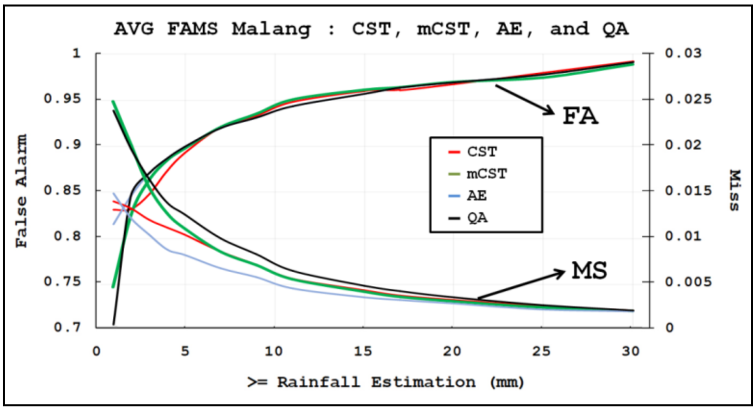

Figure 8. Average result of estimation methods in all method schemes

Unlike figure 7, figure 8 shows average result for all schemes and methods that we used for making rainfall estimation. The red one is CST method, green is mCST method, black is QA method, and blue is AE method. Based on the result, the best average methods for rainfall estimation in all classes is

Quantile Analysis (QA) methods. It has shown from the minimum value of *false alarm* in this method, although *miss* condition is higher than that of others. Miss condition is quite small and not significant compared with *false alarm*.

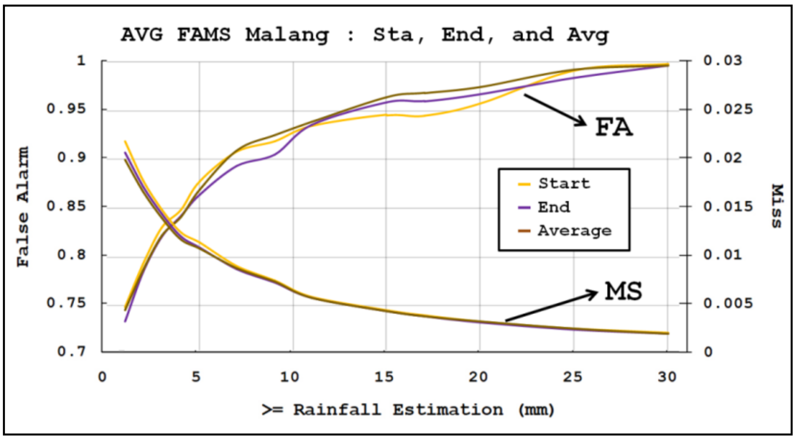

Figure 9. Average result of estimation methods in all time schemes

Figure 9 shows time for making estimation of rainfall. We divide into 3 kind of schemes. First scheme is called "start time". Its means that rainfall estimation is done at first time of satellite data. For example, the satellite data at 08.00 UTC is for making rainfall estimation during 08.00 – 09.00 UTC. Second scheme is called "end time". Rainfall estimation is done at the end of time for satellite data. For example, the satellite data at 08.00 UTC is for making rainfall estimation from 07.00 – 08.00 UTC. The last scheme is called "average time". It means we use both "start time" and "end time" for making rainfall estimation for one hour. In figure 9, the color of orange is the start time. The color of violet is the end time, and brown is an average time. Based on the result, it is shown that the effect of different time is not significant but using start time point should be avoided due to relatively higher FA and MS.

4. Conclusion

Based on the research, we learned that many factors must be considered before proceeding. The most important thing before research is the Quality Control (QC) of the data. In this research, the rainfall observation data from 2009 – 2016 is not existing all in digital format. So, we only use start from 2009 – 2016. Besides, there is difference in reporting for rainfall intensity, so we have to double-check the rainfall data using two types of report (ME.48 and Form B).

The other lesson is performing the calculation by two types of software; they are Python and R. In this research, we must double check calculation system using another software for checking the result of data calculation. In this research, the main software for calculating the rainfall intensity from satellite data is "Python" and the result of the calculation is compared to result from software "R" for Quality Control.

Based on the data analysis, we can understand about the pattern of *false alarm* and *miss* condition in rainfall estimation using satellite by different rainfall intensity and different schemes. *A false alarm* is high when rainfall intensity is high and decreases gradually with decreasing rainfall estimation. For a miss, condition has an opposite pattern with a *false alarm*. Miss value is high when rainfall intensity is slight and decreases gradually until rainfall intensity become moderate – high. The best schemes of rainfall estimation using satellite data are Quantile Analysis using southwest pixel of Malang location and by the end of time observation satellite. The scheme gives the lowest *false alarm* in all intensities, although miss condition is higher than those of others *miss* condition is quite small and not significant compared with the *false alarm*. Quantile analysis is better due to its empirical-based coefficient. It avoids error from generalization process. In the end, this study encourages the idea that local equation is better than the general one so local study for it is urgent to be done.

Acknowledgments
The authors wish to thank Kochi University Weather Home for providing the infrared weather satellite data. We also like to thank to Malang Climatological Station for providing the rainfall intensity in some reported documents.

References
[1] Caraka R E, Supari and Tahmid M 2018 Copula-Based Model for Rainfall and El- Niño in Banyuwangi Indonesia *J. Phys. Conf. Ser*
[2] Caraka R E, Bakar S, Tahmid M, Yasin H and Kurniawan I D 2018 Neurocomputing Fundamental Climate Analysis *Telkomnika*
[3] BMKG 2010 KEP.009 Tahun 2010 Tentang Prosedur Standart Operasional Pelaksanaan Peringatan Dini, Pelaporan dan Diseminasi Informasi Cuaca Ekstrim *Jakarta: BMKG*
[4] Suwarsono P, Kusumaning A D S and Kartasamita M 2009 Penentuan Hubungan Antara Suhu Kecerahan dangan MTSAT dengan Curah Hujan Data QMORPH *J. Penginderaan Jauh* **6** 32–42
[5] Caraka R E and Yasin H 2018 *Spatial Data Panel 1st ed* Wade Group
[6] Phipps M and Rowe S 2010 Seeing satellite data *Public Underst. Sci.* **19** 311–321
[7] Swarinoto Y and Husain 2012 Estimasi Curah Hujan Harian Dengan Metode Auto Estimator (Kasus Jayapura Dan Sekitarnya) *J. Meteorol. dan Geofis*
[8] Vicente G A, Scofield R A and Menzel W P 1998 The Operational GOES Infrared Rainfall Estimation Technique *Bull. Am. Meteorol. Soc.* **79** 1883–1893
[9] Adler R F and Negri A J 1988 A Satellite Infrared Technique to Estimate Tropical Convective and Stratiform Rainfall *Journal of Applied Meteorology* **27** 30–51
[10] Goldenberg S B, Houze J R A and Churchill D D 1990 Convective and stratiform components of winter monsoon cloud cluster determined from geosynchronous *J. Meteorol. Soc. Japan* **68**37 – 63
[11] Endarwin, Tjasyono B, Gunawan D and Siswanto 2014 *Modifikasi Convective Statiform Technique dengan Modifikasi Kombinasi Data Satelit Gelombang Mikro Pasif dan Inframerah Untuk Estimasi Curah Hujan di Indonesia,* ITB
[12] Anagnostou E N and Kummerow C 1997 Stratiform and convective classification of rainfall using SSM/I 85-GHz brightness temperature observations *J. Atmos. Ocean. Technol.* **14** 570 – 575
[13] Jamilatuzzahro, Herliansyah R and Caraka R E 2018 *Aplikasi generalized linear model pada R, 1st ed*. Yogyakarta: INNOSAIN
[14] Sutiksno D U, Gio P U, Caraka R E and Ahmar A S 2018 Brief Overview of STATCAL Statistical Application Program *J. Phys. Conf. Ser.,* **1028**
[15] Payne J 2010 *Beginning Python: Using Python 2.6 and Python 3.1.* Wrox Press Ltd.
[16] Caraka R E and Yasin H 2017 *Geographically Weighted Regression (GWR) Sebuah Pendekatan Regresi Geografis, 1st ed*. MOBIUS GRAHA ILMU
[17] QGIS Development Team 2015 *QGIS Geographic Information System* Open Source Geospatial Foundation Project.
[18] Wilks D S 2006 *Statistical Methods in the Atmospheric Sciences* Elsevier Inc
[19] Dinku T, Chidzambwa S, Ceccato P, Connor S J and Ropelewski C F 2008 Validation of high-resolution satellite rainfall products over complex terrain Validation of high-resolution satellite rainfall products over complex terrain *International Journal of Remote Sensing* **29** 4097 – 4110
[20] Kenabatho P K, Parida B P and Moalafhi D B 2017 Evaluation of satellite and simulated rainfall products for hydrological applications in the Notwane catchment, Botswana, *Phys. Chem. Earth* **100** 19–30
[21] Hossain F and Huffman G J 2008 Investigating Error Metrics for Satellite Rainfall Data at Hydrologically Relevant Scales *Journal of Hydrometeorology* **9** 563–575