VALIDATION OF TROPICAL RAINFALL MEASURING MISSION (TRMM) DATA IN THE UPPER KAPUAS RIVER BASIN

Stefanus Barlian Soeryamassoeka[1], Robertus Wahyudi Triweko[2]* and Doddi Yudianto[3]

[1]Civil Engineering Department, Tanjungpura University, Pontianak, Indonesia,

[2,3]Civil Engineering Department, Parahyangan Catholic University, Bandung, Indonesia.

Abstract — Rainfall is a difficult parameter to measure, due to large spatial and temporal variations. Lack of data availability, data incompletely, less spreading of station, less observer, and manual data entry are other problems for rainfall predicting. To encourage these problems rainfall satellite can be used, because it has high temporal and spatial resolution, widely coverage, near real-time and fast accessibility. This research was conducted in the upper Kapuas River Basin, West Kalimantan, to determine how TRMM satellite-derived rainfall compares with ground-measured values and the possibility of using it to complement ground-measured rainfall. The statistical analyses and correction factor development for TRMM data are conducted to validate and correct the TRMM data on eleven sub basin in Kapuas River basin. Validation showed high correlation between TRMM and gauge data. Validation shows a high correlation and lowest RMSE between TRMM and gauge data in the sub basin adjacent to the gauge station (r= 0.76-0.8, RMSE 0,84-0,92). The results of the analysis also show that after correction, the corrected TRMM data errors were reduced for the eleven rainfall.

1.0 INTRODUCTION

Rainfall is one of the important elements in the hydrological cycle., so many region in the tropics depend on rainfall for their various needs [1]. This is because the accuracy of any hydrologic study depends heavily on the availability of good-quality precipitation estimates. However, rainfall is a difficult parameter to measure, due to large spatial and temporal variations. Lack of data availability, data incompletely, less spreading of station, less observer, and manual data entry are other problems for rainfall predicting. Because of this in several regions of the world, the availability of rainfall data is very limited as well as Kapuas River Basin. Tropical Rainfall Measuring Mission (TRMM), which was developed by NASA, became an alternative solution to the limitation of rainfall data. Although the TRMM data provide quick access, near real-time data, spatially and temporarily distributed, the satellite data need to be verified and evaluated with the ground station data. There are several reasons the data needs to be evaluated, one of them because of the satellite is using the infrared and microwave radiation to measure the rainfall and sometimes the waves are interrupted by something in the atmosphere [2]. A few studies have developed methods to correct the daily satellite rainfall by using ground station data in Indonesia, such as research on validation and correction of Trmm satellite data on three rainfall Patterns in indonesia [3], evaluation and bias correction of satellite rainfall data for drought monitoring in Indonesia [4]. However, these studies have not specifically discussed the validation of TRMM rainfall in the upper Kapuas River Basin, hherefore, this study aims to present the results of the correlation and validation of rainfall data in the upper Kapuas River Basin.

2.0 METHODOLOGY

This study is concerned with validation of a monthly TRMM data in uppper Kapuas River Basin. The data used in this study is the Tropical Rainfall Measuring Mission (TRMM) data and rainfall observation data from 5 ground station (KPH-06 Ng Kantuk, KPH-08 Silat, KPH-09 Jongkong, KPH-13 Tepuai and BMKG Pangsuma) by period 1998-2016. Rainfall data obtained from 5 rainfall ground stations in upper

Kapuas River Basin and TRMM measurements are compared first. This comparison is made to check if the ground station and TRMM rainfall data reliable enough to be used in the analysis. A correlation coefficient is used to assess these data. As the measurement conducted in the same area for both TRMM and ground stations, both data must have a reasonable correlation value and homogeinity since it shows similar characteristics. Rainfall data that passes the correlation assessment are going to be used in TRMM correction determination. TRMM data correction is done on a monthly basis. As on a monthly basis, rainfall event randomness is reduced, and certain lag caused by the TRMM measurement method and ground station measurement method can be ignored (lag less than a day). Ground stations rainfall data are assumed to be the correct rainfall measurement since it showing the record of an actual event. In the correction process, objective functions are used to show TRMM error to ground station data. The correction will be done using a linear regression model with one unknown variable to be determined. Using the objective function, the best correction value, which results in the least error, is obtained. These correction coefficients, for monthly correction, will be tested to be used in Upper Kapuas River Basin with its respective TRMM and ground station data.

In validating the TRMM rainfall data, first the observation stations and the TRMM grid were used, plotted on the map. Then all of the daily rainfall data from ten rainfall stations are summarized to monthly rainfall. Before these data are used in the analysis, a data reliability test is needed to see whether the data from the five ground stations are reliable enough to be used as data to validate the TRMM rainfall data, by means of which, monthly rainfall data from all ground stations are compared with monthly TRMM data. This is to see the relationship between ground station rainfall data and TRMM data using the correlation coefficient. If the correlation value is higher than 0.6, that mean ground station data is matched with TRMM data and can be used for further analysis. If the correlation coefficient is less than 0.6, this indicates there is something wrong with the ground station data collection and needs to be checked. Thus, the data is eliminated for this study. If all the ground station data available in a grid did not satisfy the correlation requirement, that indicates the TRMM data cannot be used and will be eliminated.

3.0 LITERATURE REVIEW

Tropical Rainfall Measuring Mission (TRMM), which conducted by NASA, is a mission to measure precipitation using satellite. TRMM data used for this research is the 3B42 V7 TRMM product. TRMM Multi-Satellite Precipitation Analysis (TMPA) 3B42 is the daily accumulated precipitation product. It provides daily many selection TRMM products (3-hourly, daily, and monthly) from 1998 to the present day. TRMM data extracted from open access NASA's Giovanni website. TRMM satellite launched in 1997 and provides gridded precipitation data with a spatial resolution of 0.25^{o} x 0.25^{o} or equal to 28 km x 28 km and temporal resolution from daily up to 3-hour rain. Although the satellite data provide quick access, near real-time data, spatially and temporarily distributed, the satellite data need to be verified and evaluated with ground station data. As mentioned before, TRMM data need to be evaluated because the satellite is using the infrared and microwave radiation to measure the rainfall, which sometimes the waves are interrupted by something in the atmosphere. The objective function used in this study area Correlation coefficient and Root Mean Square Error (RMSE). These objective functions will be used in assessing the performance of ground station data and TRMM in hydrologic modeling.

3.1. Correlation

Correlation is a statistical technique that can show whether and how strongly pairs of variables are related. The main result of a correlation is called the correlation coefficient (or "r"). It ranges from -1.0 to +1.0. The closer r is to +1 or -1, the more closely the two variables are related.

If r is close to 0, it means there is no relationship between the variables. If r is positive, it means that as one variable gets larger the other gets larger. If r is negative it means that as one gets larger, the other gets smaller (often called an "inverse" correlation). The correlation coefficient is calculated as:

$$r = \frac{n\left(\sum XY\right) - \left(\sum X\right)\left(\sum Y\right)}{\sqrt{\{n\left(\sum X^2\right) - \left(\sum X\right)^2\} - \{n\left(\sum Y^2\right) - \left(\sum Y\right)^2\}}} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots(1)$$

r : correlation coefficient (the closer r is to +1 or -1)

X : rainfall data of ground station (mm)

Y : TRMM data (mm)

n : number of pairs of scores.

A satisfactory correlation coefficient is dependent on the purpose of which is to be used, and on the nature of raw data. For evaluating the correlation coefficient, Asuero et al provides a rule-of-thumb scale as shown in the following table [5].

Table 1. Strength of Correlation

Size of r	Interpretation
0.90 to 1.00	Very high correlation
0.70 to 0.89	High correlation
0.50 to 0.69	moderate correlation
0.30 to 0.49	low correlation
0.00 to 0.29	little if any correlation

3.2. Root Mean Square Error (RMSE)

Root Mean Square Error (RMSE) is the standard deviation of the residuals (prediction errors). Residuals are a measure of how far from the regression line data points are; RMSE is a measure of how spread out these residuals are [6]. The Root Mean Square Error (RMSE) has been used as a standard statistical metric to measure model performance in meteorology, air quality, and climate research studies [7]. RMSE is expressed as [3];

$$RMSE = \sqrt{\left[\frac{1}{N}\sum_{i=1}^{N}(X - Y)^2\right]}\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots(2.)$$

X : rainfall data of ground station (mm)

Y : TRMM data (mm)

n : number of pairs of scores.

4.0 RESULT AND DISCUSSION

4.1. Digitizing data

Data digitization was used to plot the ground station position and the TRMM grid based on its latitude and longitude degrees. The digitization results are as shown below
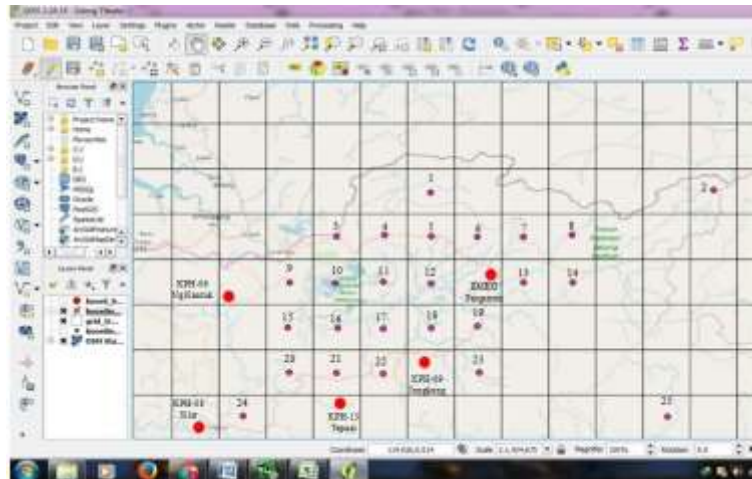
Figure 1. Digitization results position of ground station and TRMM grid

From the results of digitizing the position of the ground station and the TRMM grid, it can be seen that each ground station and TRMM that will be used in the analysis for this study is located on each different grid. So that the next analysis of each ground station and TRMM will be compared to select a ground station that will be used to validate the TRMM data.

4.2. Data Screening

In conducting screening data to select a ground station to validate TRMM rainfall data, the five rainfall stations will be calculated for their correlation with 25 TRMM data grids. Each rainfall station will see its correlation coefficient against each TRMM data. Therefore, the correlation coefficient value of 0.6 is used as a measure of whether the rainfall station represents the rainfall event at that location or not.

The results of the analysis show that four of the five observation rainfall stations in the upstream Kapuas River Basin have a correlation coefficient value of less than 0.6, so that only 1 observation station can be used to correct the 25 TRMM grid data, namely the BMKG Pangsuma station with the selected regression model is a linear regression model.

Table 2. Correlation of the observation station and the TRMM grid

TRMM Grid	KPH-06 Ng Kantuk			KPH-08 Silat			KPH-09 Jongkong			KPH-13 Tepuai			BMKG Pangsuma		
	Model			Model			Model			Model			Model		
	Linear	Exponential	Logarithmic	Linear	Exponential	Logarithmic	Linear	Exponential	Logarithmic	Linear	Exponential	Logarithmic	Linear	Exponential	Logarithmic
1	0,38	0,36	0,36	0,29	0,28	0,27	0,33	0,21	0,22	0,31	0,19	0,20	0,63	0,51	0,52
2	0,19	0,22	0,27	0,15	0,15	0,15	0,15	0,14	0,14	0,21	0,08	0,07	0,61	0,49	0,50
3	0,44	0,42	0,41	0,33	0,31	0,32	0,38	0,26	0,29	0,33	0,21	0,24	0,65	0,52	0,51
4	0,44	0,43	0,41	0,32	0,30	0,31	0,36	0,24	0,28	0,36	0,24	0,28	0,63	0,51	0,54
5	0,38	0,36	0,36	0,29	0,27	0,28	0,33	0,21	0,22	0,31	0,19	0,20	0,63	0,51	0,55
6	0,35	0,35	0,36	0,30	0,28	0,29	0,32	0,26	0,21	0,31	0,25	0,20	0,67	0,55	0,56
7	0,25	0,22	0,21	0,21	0,19	0,19	0,25	0,13	0,14	0,23	0,11	0,12	0,60	0,54	0,49
8	0,23	0,23	0,23	0,16	0,14	0,14	0,22	0,21	0,21	0,20	0,18	0,19	0,60	0,48	0,49
9	0,54	0,42	0,43	0,33	0,31	0,32	0,38	0,26	0,29	0,33	0,31	0,31	0,60	0,47	0,46
10	0,58	0,45	0,44	0,48	0,46	0,47	0,49	0,37	0,41	0,42	0,40	0,41	0,61	0,49	0,52
11	0,44	0,32	0,35	0,33	0,31	0,32	0,38	0,26	0,27	0,33	0,31	0,31	0,65	0,53	0,57
12	0,36	0,24	0,28	0,27	0,25	0,26	0,32	0,26	0,21	0,29	0,27	0,28	0,66	0,54	0,55
13	0,34	0,22	0,23	0,24	0,12	0,13	0,33	0,31	0,32	0,31	0,29	0,30	0,67	0,59	0,56
14	0,34	0,28	0,23	0,24	0,11	0,10	0,33	0,31	0,32	0,31	0,29	0,30	0,67	0,55	0,56
15	0,37	0,25	0,26	0,27	0,15	0,18	0,34	0,32	0,32	0,29	0,27	0,28	0,64	0,51	0,50
16	0,37	0,24	0,23	0,27	0,15	0,19	0,34	0,32	0,32	0,29	0,27	0,28	0,64	0,52	0,55
17	0,37	0,25	0,28	0,29	0,17	0,18	0,32	0,30	0,30	0,31	0,29	0,30	0,65	0,53	0,57
18	0,37	0,25	0,29	0,29	0,23	0,18	0,32	0,30	0,30	0,31	0,29	0,30	0,65	0,53	0,54
19	0,34	0,22	0,23	0,24	0,12	0,13	0,33	0,31	0,32	0,31	0,29	0,29	0,67	0,58	0,56
20	0,37	0,31	0,26	0,27	0,14	0,13	0,34	0,32	0,32	0,29	0,27	0,28	0,64	0,52	0,53
21	0,37	0,36	0,32	0,33	0,21	0,24	0,34	0,32	0,33	0,34	0,32	0,33	0,64	0,51	0,50
22	0,36	0,35	0,33	0,27	0,15	0,19	0,32	0,30	0,31	0,29	0,27	0,28	0,66	0,54	0,57
23	0,35	0,34	0,38	0,28	0,16	0,17	0,33	0,31	0,32	0,30	0,28	0,29	0,67	0,55	0,59
24	0,38	0,36	0,35	0,33	0,27	0,22	0,35	0,33	0,34	0,37	0,35	0,36	0,65	0,53	0,54
25	0,34	0,34	0,34	0,27	0,25	0,25	0,31	0,29	0,30	0,33	0,31	0,32	0,66	0,60	0,55

4.3. TRMM Correction Factor

The correction factor for the monthly TRMM satellite data in this study is determined using the least squares equation, with the values a and b as correction factors, the minimum boundary requirements. $\bar{V} \rightarrow \varepsilon \approx 0$ or $\frac{\partial \varepsilon}{\partial a} = 0$, $\frac{\partial \varepsilon}{\partial b} = 0$ or JKG $= \min \sum_{i=1}^{n} [Y_i - (a+bx_i)]^2$, as the X variable is BMKG Pangsuma rainfall data, Y variable is the TRMM rainfall data. By entering these provisions, iteration is then carried out and the value that meets the correction factor a is 0.4455 and b is 165.067, so that the correction equation for TRMM rainfall data at the study location is Y = 0.4455 X + 165.067. By using these correction and correction factors, we look for the amount for which TRMM rainfall is corrected, in order to obtain information, RMSE, MAE and relatively bias. The result of the analysis is as follows;

Table 3. Validation of TRMM data

TRMM Grid	Description	Correlation	RMSE	MAE	Relative Bias	TRMM Grid	Description	Correlation	RMSE	MAE	Relative Bias
1	Before validation	0,63	129,88	23,80	7,38	14	Before validation	0,67	121,25	34,85	11,50
	After validation	0,73	98,05	7,23	2,06		After validation	0,75	79,29	26,69	4,17
2	Before validation	0,61	142,17	37,97	11,77	15	Before validation	0,68	82,76	27,91	4,69
	After validation	0,61	94,26	19,93	6,84		After validation	0,64	127,87	24,06	7,46
3	Before validation	0,65	124,89	18,18	5,64	16	Before validation	0,65	120,83	23,63	3,27
	After validation	0,74	65,52	12,56	3,83		After validation	0,64	127,87	24,06	7,46
4	Before validation	0,63	127,09	37,48	2,32	17	Before validation	0,65	130,83	23,63	3,27
	After validation	0,74	90,79	9,86	2,04		After validation	0,65	124,45	15,97	4,95
5	Before validation	0,63	129,88	23,80	7,38	18	Before validation	0,66	116,87	15,74	3,40
	After validation	0,73	98,05	5,23	2,06		After validation	0,65	124,45	15,97	4,95
6	Before validation	0,67	123,07	76,70	5,32	19	Before validation	0,66	116,87	15,74	3,40
	After validation	0,75	95,52	17,38	3,15		After validation	0,67	121,25	54,85	6,50
7	Before validation	0,60	135,32	21,05	8,32	20	Before validation	0,68	122,76	14,91	4,69
	After validation	0,69	74,06	17,44	2,28		After validation	0,64	127,87	24,06	7,46
8	Before validation	0,60	103,12	42,46	6,96	21	Before validation	0,65	130,83	23,63	3,27
	After validation	0,69	72,43	22,01	2,16		After validation	0,64	126,66	20,38	6,12
9	Before validation	0,60	174,89	58,18	30,64	22	Before validation	0,65	87,70	10,55	3,27
	After validation	0,62	80,52	12,56	16,83		After validation	0,66	123,24	5,66	1,75
10	Before validation	0,61	143,75	70,71	19,76	23	Before validation	0,67	124,71	15,69	3,95
	After validation	0,62	118,02	35,23	6,67		After validation	0,66	122,76	7,36	2,28
11	Before validation	0,65	124,89	18,18	5,64	24	Before validation	0,67	124,40	17,35	5,50
	After validation	0,74	65,52	12,56	3,83		After validation	0,65	123,76	3,63	1,13
12	Before validation	0,66	123,24	5,65	1,75	25	Before validation	0,66	125,14	13,72	4,30
	After validation	0,77	82,02	0,11	0,03						
13	Before validation	0,67	121,30	34,98	11,55						

The results of the analysis that have been carried out show that the BMKG rainfall data has a strong relationship to all TRMM grids, this is shown from the results of the correlation analysis before the data is corrected and after the data is corrected which is generally ≥ 0.6, as well as the RMSE value which is generally less from 130 mm. From the analysis, it can also be seen that there is an improvement in the correlation value, RMSE, MAE and the relative bias when the TRMM data is corrected. The results of the analysis, it can be said that the TRMM rainfall data can be used as analysis data for the upstream part of the Kapuas basin. Furthermore, the rainfall data at the BMKG Pangsuma and the average rainfall from 25 TRMM grids are plotted in a graph to see the rainfall pattern in the study location. To see the existing rainfall patterns, an example of a 1-year pattern is shown, namely 2003, a 4-year pattern, namely 2003-2006 and the overall pattern, 1998-2016.
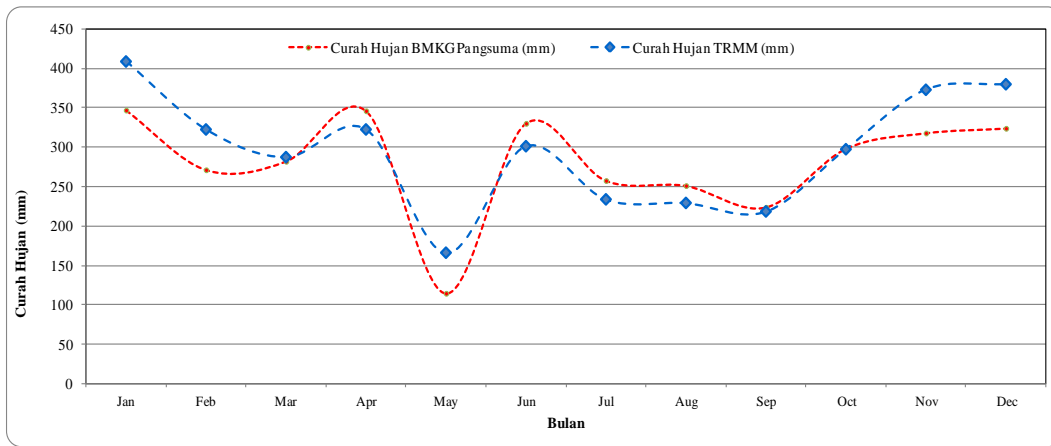
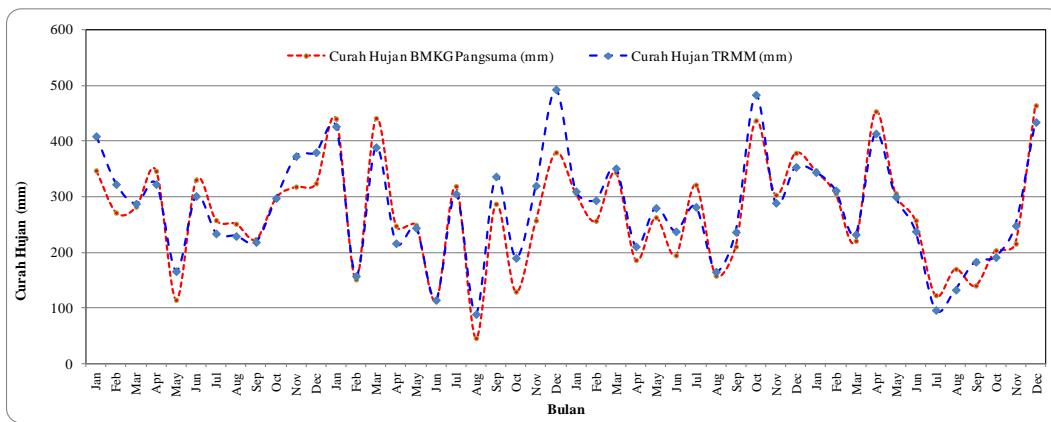Figure 2. Rainfall pattern in the upstream Kapuas River Basin in 2003

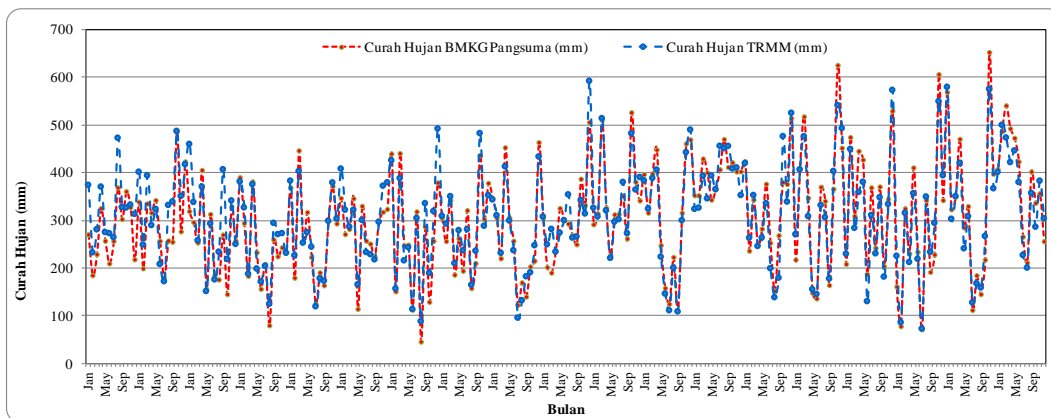Figure 3. Rainfall pattern in the upstream Kapuas River Basin in 2003-2006

Figure 4. Rainfall pattern in the upstream Kapuas River Basin in 1998-2016

From Figure (2-4) it is known that the rainfall patterns of BMKG Pangsuma and TRMM tend to be the same; both have two peak rain peaks, which occur in months other than August. The existence of two bimodial rain peaks in a year means that the rain pattern in the upstream part of the Kapuas River is included in the equatorial rainfall pattern [8], [9], [10]. (Tjasjono, 2006; Aldrian & Susanto, 2003; Aldrian, 2003).

5.0 CONCLUSION

From a series of rainfall analysis in this study, it is known that of the five ground stations in the upstream part of the Kapuas River Basin used in this study, the rainfall recorded at the BMKG Pangsuma Station has a good relationship with the data from the TRMM grid used, with a general correlation> 0.6 and RMSE <130 mm, so it can be said that the resolution of TRMM rainfall with the rainfall recorded at BMKG Pangsuma did not cause a significant bias towards the broad Kapuas River Basin coverage, thus the TRMM algorithm was used for analysis at the study site. can be used to calculate rainfall in accordance with the actual rainfall conditions based on records from the Pangsuma BMKG observation station. Thus, the results of corrections made to representative TRMM rainfall are used in hydrological analysis in the upstream part of the Kapuas River Basin. From the analysis, it can be seen that the rain pattern in the upstream part of the Kapuas River Basin each year has two peaks of maximum rain and it rains throughout the year. The peak of rain does not necessarily occur in the same month, but never outside of October, November, December, January, February and March, thus it can be said that the Kapuas d River Basin in the upstream part enters the equatorial rain pattern, namely the area which has a bimodial monthly rain distribution with two peaks of the maximum rainy season and rain occurs almost throughout the year.

6.0 ACKNOWLEDGEMENT

REFERENCES

[1]   Wu, Z.; Chenyi, Y.; Qingyun, Z; Yueqian, C. (2018). Evaluation of the Validation of TRMM data over the Region of Qilianshan Mountain in Northwest China. Proc. SPIE 10776, Remote Sensing of the Atmosphere, Clouds, and Precipitation VII, 107760T (24 October 2018); doi: 10.1117/12.2324651.

[2]   Wijaya, O.T., Wanny, K.A., Bambang, A.R., Doddi, Y., Anri, N.A.R., and Reva, A.M., 2018. Application of TRMM for Runoff Prediction in Ungauged Basin, Case Study: Nanjung Catchment. 21st IAHR-APD 2018 Yogyakarta.

[3]   Mamenun, H., Pawitan, dan A., Sophaheluwakan. (2014). Validasi Dan Koreksi Data Satelit TRMM Pada Tiga Pola Hujan Di Indonesia. Jurnal Meteorologi dan Geofisika Vol. 15 no. 1 Tahun 2014 : 13-23.

[4]   Vernimmen, R., R., E., A., Hooijer, Mamenun, E,. Aldrian, dan A. I. J. M. van Dijk. (2012). Evaluation and bias correction of satellite rainfall data for drought monitoring in Indonesia. Hydrol. Earth Syst. Sci., 16, 133–146. Published by Copernicus Publications on behalf of the European Geosciences Union.

[5]   Asuero, A. G., A. Sayago, and A. G. González. 2006. The Correlation Coefficient: An Overview. Critical Reviews in Analytical Chemistry 36 (1): 41–59. https://doi.org/10.1080/10408340500526766.

[6]   Hyndman, Rob J.; Koehler, Anne B. (2006). "Another look at measures of forecast accuracy". International Journal of Forecasting. 22 (4): 679–688. CiteSeerX 10.1.1.154.9771. doi:10.1016/j.ijforecast.2006.03.001.

[7]   Chai, T., and R. R. Draxler. 2014. "Root Mean Square Error (RMSE) or Mean Absolute Error (MAE)? -Arguments against Avoiding RMSE in the Literature." Geoscientific Model Development 7 (3): 1247–50. https://doi.org/10.5194/gmd-7-1247-2014.

[8]   Tjasjono, B. (2004). Klimatologi Umum. Penerbit ITB Bandung.

[9]   Aldrian, E, and R.D., Susanto. (2003). Identification of Three Dominant Rainfall Regions Within Indonesia and Their Relationship to Sea Surface Temperature, Int. J. Climatol, Vol. 23, No. 12, page: 1435-1452.

[10] Aldrian., E. (2003). Simulations of Indonesian Rainfall with a Hierarchy of Climate Models.