Available online at www.sciencedirect.com
SciVerse ScienceDirect
Procedia Environmental Sciences 10 (2011) 2265 – 2271
Procedia Environmental Sciences
2011 3rd International Conference on Environmental Science and Information Application Technology (ESIAT 2011)

Accuracy Assessment of Global Satellite Mapping of Precipitation (GSMaP) Product over Poyang Lake Basin, China

Qiaoni Fu [1,2,a], Renzong Ruan [1], Yuanbo Liu [2]

[1]*School of Earth Sciences and Engineering, Hohai University, Nanjing, P.R. China*

[2]*Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing, P.R. China,*

[a]*Email: WANGKuifu@yeah.net*

Abstract

The purpose of this study is to evaluate the accuracy of GSMaP by using the gauge-based precipitation measurements across Poyang Lake Basin at daily, monthly and annual scales. The results show that GSMaP products generally underestimate precipitation amount. The monthly correlation coefficient is 0.85, which shows a significant linear relationship between product estimations and rain-gauged observations while the daily correlation coefficient is less than 0.50 on average. The performance of precipitation estimation based on satellite data is poorer in mountainous areas than that in flatlands. The results also show that relative errors decrease in wet months and increase gradually in dry months, while the trends of mean absolute errors (MAE) and root mean square errors (RMSE) are just opposite. In wet periods, the omission is higher but the commission is lower. However, in dry periods the situation is often opposite. The analyses also show that omission, commission and underestimation of precipitation caused the differences between validation data and remotely sensed data to some degree. The events of strong precipitation have not been detected; or even detected, the amount of the precipitation has been insufficiently estimated are the main reasons why there is a difference between remotely sensed data and validation data. Moreover the underestimation and overestimation of precipitation amount are also the major reasons.

Keywords: Precipitation, Accuracy assessment, Poyang Lake Basin, GSMaP

Introduction

Launched in 1997, the TRMM satellite plays a vital role in monitoring global and regional precipitation. Due to the circumstance that the global rain rate with uniform accuracy cannot be observed by any methods other than satellite remote sensing [1], retrieval of the global distribution of precipitation with enough accuracy and high spatial-temporal resolution has been one of major scientific frontiers. The Global

doi:10.1016/j.proenv.2011.09.354

Satellite Mapping of Precipitation (GSMaP) started in 2002 is one of the projects that use observations of multi-sensor as an input to retrieve rainfall rates [2]. Although it has been applied widely in hydrology and meteorology, it has not been verified in mainland of China. In this study, we attempted to evaluate the accuracy of the datasets in Poyang Lake basin.

Study Area and Data Sources

Study Area.

Our study area is Poyang Lake Basin which is well-known for being listed as one of the first batch of The Ramsar Convention Wetlands of International Importance. It is located between 24.65°-30.05°N in latitude, 113.75°-118.45°E in longitude. The basin has an area of 162, 200 km^2 [3] .It is situated in the southern beaches of the middle and lower Yangtze River, which is a part of the subtropical area in China [4]. There are diverse landform types and a variety of wetland plants in Poyang Lake watershed. This renowned bird sanctuary has a humid subtropical climate and is strongly influenced by the East Asian monsoon. It is connected with five main rivers, namely Xiushui, Xinjiang, Ganjiang, Fuhe and Raohe [3].

Data Sources.

GSMaP provides six products of various spatial and temporal resolution using combined data from the microwave radiometer (MWR) datasets observed from LEO satellites and the infrared radiometer (IR) datasets obtained from GEO satellites [5]. The highest spatial and temporal resolution of the product are 0.1° and an hour respectively [2]. According to the different underlying surfaces and rain/no-rain classification, different algorithms are chosen to estimate surface rainfalls [6]. The product GSMaP_MVK+ 4.8.4 use both datasets from microwave radiometers and infrared sensor to retrieve rain rate, and its version has not been updated during the analysis period, so it has been selected as test data in the process. Because of its high quality, the small-scale variability and non-normal statistical characteristics of precipitation could be represented.

Precipitation is not continuous in time and space; therefore, rain-gauge data were collected at 45 meteorological stations throughout Poyang Lake Basin. They were used to assess the accuracy of GSMaP precipitation products from 2003 to 2006. All the calculations were based on daily accumulation.

Methodology

We selected nine indices to evaluate the accuracy of the GSMaP datasets. In addition to correlation coefficient, relative error, MAE and RMSE, four scenarios were used to express the performance of estimates: *a* represents correctly estimated precipitation events, *b* denotes the events under which precipitation was estimated but did not occur in the validation data, *c* represents events when precipitation was not estimated but did occur in the validation data, and d stands for correctly estimated no-precipitation events [7]. With four scenarios, Bias Score (BS), Probability of Detection (POD), False Alarm Ratio (FAR), Threaten Score (TS) and Equitable Threaten Score (ETS) were use for further evaluation of accuracy of test data [8]. The above mentioned five indices and their description are as Table 1.

Table 1 Description of BS, POD, FAR, TS and ETS

Index	Expression	Implication
POD	$POD = \frac{a}{a+c}$	the ratio of correct precipitation estimates to the number of precipitation events observed
FAR	$FAR = \frac{b}{a+b}$	the fraction of precipitation estimates that turn out to be wrong
BS	$BS = \frac{a+b}{a+c}$	the ratio of the number of “yes” estimates to the number of “yes” observation
TS	$TS = \frac{a}{a+b+c}$	the number of correct “yes” estimates divided by the total number of occasions for which precipitation event is observed or estimated
ETS	$ETS = \frac{a - a_{ref}}{a - a_{ref} + b + c}$ $a_{ref} = \frac{(a+b)(a+c)}{n}$	a modification of the TS that takes into account the number of hits that could be expected, due purely to random chance

To clarify the mistakes of precipitation detection and to quantify their characteristics, we took the precipitation in 2006 as an example in the end. The statistical data show that the precipitation mainly occurs in April, May and June while it rarely occurs in October, December and January in Poyang Lake Basin. In this case, the year 2006 is divided into three periods: the wet period (April, May and June), the dry periods (October, December and January) and the other period (February, March, July, August, September and November), although it couldn’t be absolutely suitable for every year.

Results and Discussions

Fig.1 illustrates that the GSMaP_MVK+ product had a significant linear relationship with a correlation coefficient of 0.85 at monthly scale [5], although the observed values were generally larger than the GSMaP_MVK+ over Poyang Lake Basin during 2003-2006. On the other hand, at daily scale, the correlation coefficient is relatively low for each year, and the value is less than 0.50 on average for all the years. In this calculation, the cases in which both the remotely sensed dataset and the validation dataset were 0 mm/month or 0 mm/day were excluded. Further investigation indicates that the performance of precipitation by using remotely sensed data in each sub-basin (including lake area) is nearly the same at either daily or monthly scale. But remotely sensed precipitation estimation is poorer in mountainous areas than that in other areas. At Jinggang Mountain station, the correlation coefficient at monthly scale falls to 0.77 and a worse linear relationship appeared in comparison with the average value. At daily scale, the value even falls downward to 0.26 in 2006.

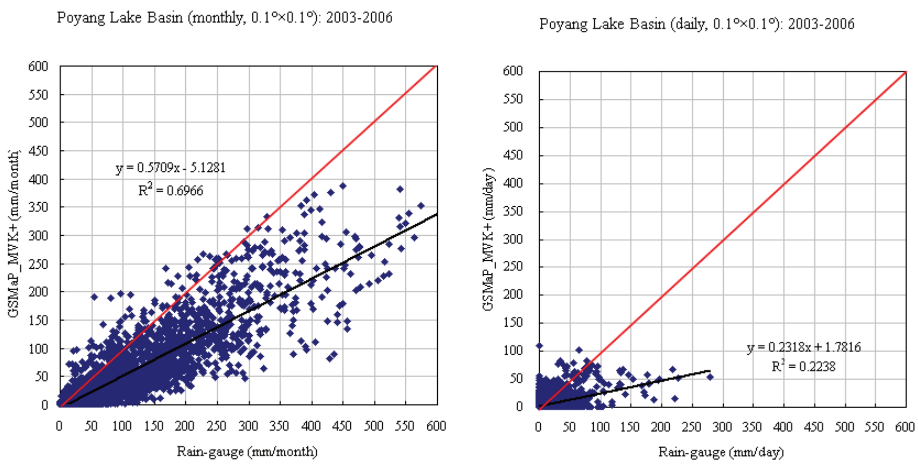

Fig.1(a) Fig.1(b)

Fig.1 Rain-gauge based monthly measures against the GSMaP_MVK+ datasets in Poyang Lake Basin of China for monthly scale during January 2003 -December 2006 (a) and for daily scale during January 2003 -December 2003(b)

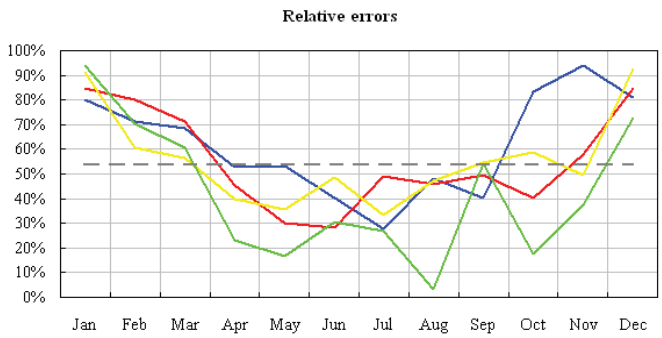

Fig.2 Percentage of relative errors in Poyang Lake Basin for 2003-2006. the values of 2003, 2004, 2005 and 2006 in blue, red, yellow and green respectively the average value for 2003-2006 in dash

Fig.2 demonstrates relative errors by using the rain-gauge datasets as reference data in the study area. The curved line decreases in wet months and it increases gradually in the dry months. Relative errors for average precipitation rate is 37% in wet periods and is 77% in dry periods. The average value of relative errors from 2003 to 2006 is 54%. On the contrary, MAE rises in wet months and decreases in the dry months. The absolute error of precipitation in wet periods is 0.50 mm/d and 0.17 mm/d in dry periods. The former is nearly three times higher than the latter. The trends of change of RMSE are similar as MAE. By further incorporated with Fig.3, the findings are drawn as follows: 1) The average value of POD ranges from 0.70 to 0.85 in wet months and ranges from nearly 0.20 to 0.45 in dry months. Take 2006 as an example, the maximum of POD appears in July and the minimum in November; they are 0.84 and 0.19 respectively. 2) The average of FAR ranges from 0.20 to 0.35 all around the year. In wet periods, the omission is higher but the commission is lower. However, in dry periods the situation is often opposite. 3) ETS is always less than TS, and the average difference is 0.32 in wet periods and is 0.10 in dry periods. The difference between TS and ETS is caused by the frequency of occurrence of random events.

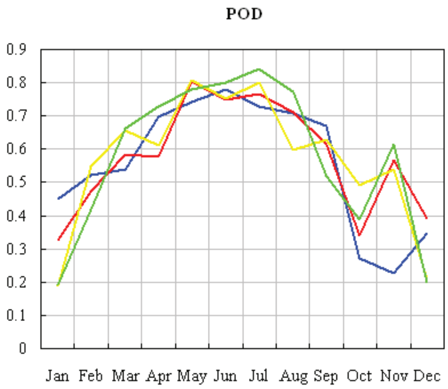

Fig.3(a)

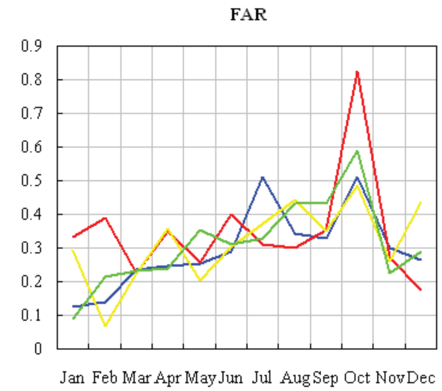

Fig.3(b)

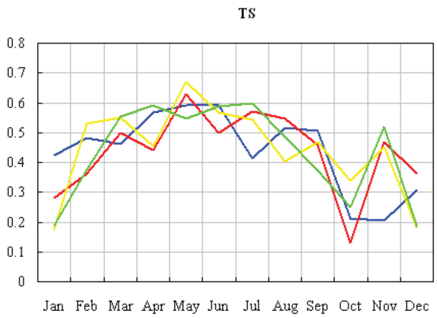

Fig.3(c)

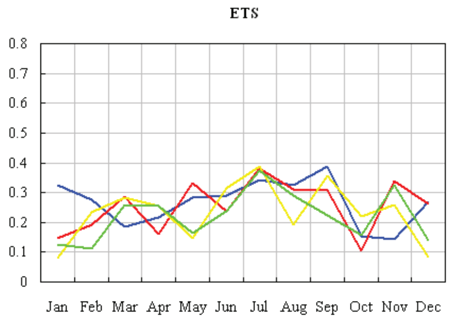

Fig.3(d)

Fig.3 Time series of rain-gauged validation for January through December during 2003-2006 using daily estimates in terms of (a) POD, (b) FAR, (c)TS and (d)ETS.
Blue, red, yellow and green lines indicate values of 2003, 2004, 2005 and 2006, respectively.

Additionally, according to the case study of 2006, it is found that the occurrence in which the misinterpretation happens and the precipitation differences caused by the misinterpretation of precipitation detection were significantly similar in each sub-basin. Here Fuhe sub-basin is taken as specific example for further explanation. Table 2 shows the frequencies and the percentage of the four scenarios above mentioned in different periods. Table 3 shows the difference of precipitation between the remotely sensed data and validation data caused respectively in the scenarios *a*, *b* and *c*. A combination of Table 2 and Table 3 demonstrates that the three scenarios have contributed to the differences of precipitation to some degree. In wet periods, the differences of precipitation is mainly caused by scenario *a*, not mainly by scenario *b* and *c*. In dry periods, precipitation does not occur frequently there. The differences of precipitation is mainly caused by scenario *c*, not by scenario *a* and *b*. Comparing the remotely sensed data with validation data for each day, it is found that much of the underestimation is caused by omission or insufficiently estimated of precipitation amount of strong events (⩾20 mm/d) [6].

Table 2 The frequences and proportions of the four scenarios in every period of 2006

Period \ Scenario		*a*	*b*	*c*	*d*	Sample total
Wet	Frequence	222	77	43	113	455
	Proportion	49%	17%	9%	25%	100%
Dry	Frequence	23	11	86	345	455
	Proportion	5%	2%	18%	74%	100%
Others	Frequence	251	127	138	389	905
	Proportion	28%	14%	15%	43%	100%

Table 3 The differences of precipitation amount coused by three scenarios in every period

Period	Wet				Dry				Others			
Station	Diff	*a*	*b*	*c*	Diff	*a*	*b*	*c*	Diff	*a*	*b*	*c*
58619	153.3	203.3	-81.4	31.4	58.3	-16.5	-5.7	80.5	390.7	280.2	-26.8	137.3
58712	148.2	182.2	-73.8	39.8	87.3	-16.3	-2.5	106.1	320.6	259.1	-54.4	115.9
58715	394.6	373.5	-135.5	156.6	74.4	0.5	-2.0	75.9	267.8	189.8	-93.0	171.0
58719	344.5	262.0	-59.7	142.2	66.8	30.8	-0.5	36.5	202.8	174.2	-110.1	138.6
58813	177.9	218.8	-99.4	58.5	37.8	-1.5	-0.6	39.9	347.0	355.2	-97.1	89.0

*Units of data in this table are mm.

For instance, on April 26, the validation data observed by No.58719 is 54.5 mm while the remotely sensed data is 0 mm. Similarly, the validation data observed by No.58715 is 42.0 mm whereas the remotely sensed data is 0 mm. The validation data observed by No.58712 is 30.0 mm while the remotely sensed data is 0.1mm. Another example, on January 19, the validation data observed by No.58712 is 54.9 mm while the remotely sensed data there is 0 mm. The validation data observed by No.58715 is 31.7 mm, but the remotely sensed data there is 0 mm. The validation data observed by No.58719 is 21.5 mm, but the remotely sensed data there is 0.2 mm. The events of strong precipitation have not been detected; or even detected, the amount of the precipitation has been insufficiently estimated. These are the main reasons why there is a difference between remotely sensed data and validation data. Moreover, the underestimation and overestimation of precipitation amount are also the major reasons.

There were several reasons accounting for the overall underestimation of GSMaP. The main reasons include misinterpretation of precipitation and underestimation of precipitation amount. Fundamentally, it may be caused by topographic factors, mechanism of precipitation, and the defects in GSMaP algorithm. The analysis reveals that the reasons may be as follows: 1) GSMaP estimates are not sensitive to light- and very heavy-rainfall. During the dry periods with less precipitation, lots of interval showers are usually too tiny to be identified. It is easy to misinterpret the precipitation and subsequently affect the accuracy of the results. Snow and ice events may be also cause some inaccuracies of the results. 2) In the rain/no-rain classification, some phase gaps may result in errors when using data observed from unevenly distributed meteorological stations. 3) Errors and non-homogeneity in the rain profile and limited temporal sampling also contributes to the uncertainty in GSMaP products. The strong events of precipitation can not be detected or estimated accurately are the major reasons.

Conclusions

The accuracy of GSMaP were evaluated using the gauge-based precipitation measurements across Poyang Lake Basin at daily, monthly and annual scales. Compared with rain-gauged observations, GSMaP product generally underestimate precipitation amount of Poyang Lake Basin, but it performed better in wet periods than that in dry periods. The monthly and daily correlation coefficients were 0.85 and less than 0.50 on average, respectively. The performance of each sub-basin is not outstandingly better or worse, but satellite estimates were poorer in mountainous areas than that in flatlands.

The relative errors decreases in wet months and it increases in the dry months while the trends of MAE and RMSE are just opposite. Relative errors for average rain rates is 37% in wet periods and is 77% in dry periods. The absolute error of precipitation in wet periods is 0.50 mm/d, which is three times larger than that in dry periods. The trend of RMSE is similar as MAE.

In wet periods, the omission is higher but the commission is lower, and the average POD is between 0.70 and 0.85. However, in dry periods the situation is often opposite, and the average POD is between nearly 0.20 and 0.45. The value range of FAR is not wide, but the curve's amplitude is significantly higher in October. Moreover, occurrence of events influenced overall threaten scores of the remotely sensed product.

Especially, the analyses of 2006 showed that omission, commission and underestimation of precipitation lead to the differences between validation data and remotely sensed data to some degree. The events of strong precipitation have not been detected; or even detected, the amount of the precipitation has been insufficiently estimated are the main reasons. The underestimation and overestimation of precipitation amount are also the major reason.

Consequently, there are many advantages of satellite-retrieved precipitation products that they can get the large-scaled amount, small-scale variability and highly non-normal statistical characteristics of precipitation [7]. However, how to apply the product efficiently in relative fields is still a scientific challenge. Various datasets from different sources, such as the ground-based radar and drop-size-meter sources may be combined to correct the mentioned errors accurately.

Acknowledgements

This work is jointly supported by a Key Program of The Knowledge Innovation Program of CAS (KZCX2–YW–337) and Jiangsu Natural Science Fundation (BK2008360), Jiangsu Postdoctoral Fundation, Hohai University Talent Fundation, and Fundamental Reserch Funds (2009B12714，2009B11714).

References

[1] K. Okamoto, N. Takahashi, K. Iwanami, S. Shige and T. Kubota: High Precision and High Resolution Global Precipitation Map from Satellite Data, ISAP 2007 Proceedings (2007), p. 506-509

[2] K. Okamoto, T. Iguchi, N. Takahashi, K. Iwanami and T. Ushio: The Global Satellite Mapping of Precipitation (GSMaP) project, 25th IGARSS Proceedings (2005), p.3414-3416

[3] S. Q. Fan: Spatial and Temporal Distribution of Rainfall and Hydrological Responses in Poyanghu Watershed (2007)

[4] Information on http://baike.baidu.com/view/

[5] T. Kubota, S. Shige, H. Hashizume, K. Aonashi, N. Takahashi, S. Seto, M. Hirose, Y. N. Takayabu, T. Ushio, K. Nakagawa, K. Iwanami, M. Kachi and K. Okamoto: Global Precipitation Map using Satellite-borne Microwave Radiometers by the GSMaP Project: Production and Validation, IEEE Trans Geosci. Remote Sens. Vol. 45 (2007), p. 2259-2275

[6] D. T. Yu: Evaluation of GSMaP Precipitation Estimates over the Contiguous United States, Journal of Hydrometeor Vol. 11 (2010), p.566-574

[7] T. Kubota, T. Ushio, S. Shige, S. Kida, M. Kachi, and K. Okamoto: Verification of High-resolution Satellite-based Rainfall Estimates around Japan using a Gauge-calibrated Ground-radar Dataset, Journal of the Meteorological Society of Japan Vol. 87A (2009), p.203-222

[8] D. S. Wilks: Statistical methods in the atmospkeric sciences, International geophysics series, Elsevier, Amsterdam, 2nd edn. (2006)