

# The Assessment of the Accuracy of Meteorological Satellite Rainfall Measurement Products in the Heihe River Basin

**Ye wei**

Tianjin Research Institute for Water Transport Engineering, M.O.T., Tianjin 300456, China

Email: tks\_yw@126.com

**Abstract:** With the advent of satellite rainfall products, it makes up the shortage of information in the area with no regional rainfall information. It plays an important role in understanding the spatial and temporal rainfall distribution. By verifying the accuracy of satellite rainfall data, it can provide reference for data providers. Based on the reason of errors of rainfall data, it can further improve the accuracy of the meteorological satellite rainfall data. The accuracy of meteorological satellite rainfall data in day scale is verified with the monitoring data of the stations in the Heihe River Basin by Contingency Table Approach (CTA), Fuzzy Comprehensive Evaluation method and Correlation Coefficient method. Based on the results by the methods above, the classified graded method is proposed, providing a new method for the assessment about the accuracy of the satellite rainfall products.

## 1. Introduction

At present, meteorological satellite rainfall measurement products are mainly used in the field of meteorology. They are widely used in many fields such as precipitation measurement, precipitation prediction, storm research, data assimilation, and so on. As far as the application of meteorological satellite rain measurement products in hydrological field is concerned, there are few studies at home and abroad. In 2006, Md. Nazrul Islam using the rainfall data measured in stations verified the TRMM precipitation products in 3B42 rain test precision, Bangladesh's results show that the TRMM precipitation data in the monsoon period before the measurement value is small, the measured value after the monsoon period is too large, the rain measurements significantly smaller [1]. As far as China is concerned, Yang Yunchuan et al. Have checked the accuracy of TRMM and 3B42V6 in the Sichuan basin and its surrounding areas. It is found that the accuracy of precipitation increases with time scales, [2]. Hu Qingfang assessed the accuracy of 3B42V6 TRMM in Ganjiang River Basin, and the error analysis and explanation of [3].

## 2. Research Area and Data

### 2.1. Research Area Overview

Finite element models of hot-rolled RHS/SHS stub columns were developed by using the non-linear finite element program ABAQUS [5], in which two steps including linear perturbation and non-linear analyses were performed in order to obtain the ultimate carrying capacity and failure modes of RHS/SHS stub columns. Material properties and cross-section dimensions measured from Joanna's test [6] were included in the finite element model. The typical nomenclature is defined in Figure 1.



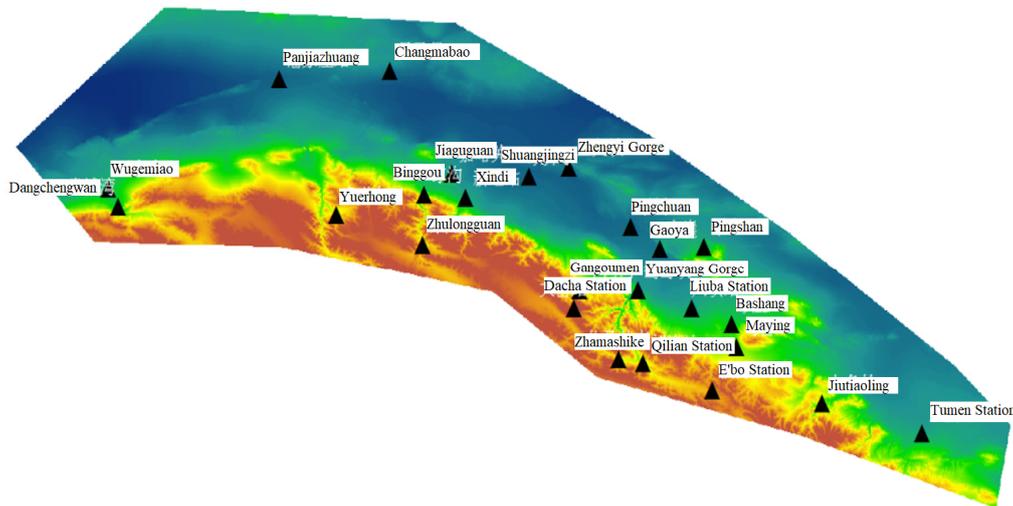


Figure 1 Distribution map of hydrologic stations in the Heihe river Basin

2.2 Research Data

(1) A total of 1980-2003 daily precipitation data of 25 hydrological stations (station distribution is shown in Figure 1) in the watershed were collected, Satellite rainfall data is a data set of surface meteorological factors in China, and it is a set of near surface meteorological and environmental factors reanalysis data set developed by the Institute of Tibetan Plateau of Chinese Academy of sciences. The data set is based on the existing international Princeton reanalysis data, GLDAS data, GEWEX-SRB radiation data and satellite precipitation data, which are fused with the conventional meteorological observation data of China Meteorological Administration. Currently, the data that has been completed and published is a subset of the B-01 of version 1.0, Its time coverage is 1981.01-2008.12 (28 years), the time resolution is 3 hours, the geographical coverage is 70°E-140°E, 15°N-55°N, the horizontal spatial resolution of China's terrestrial region is 0.1 degrees, Contains the data of near ground air temperature, near ground pressure, near ground air specific humidity, near surface full wind speed, surface downward shortwave radiation, surface downward long wave radiation, and ground precipitation rates, a total of 7 elements (variables).

(2) Daily rainfall data of meteorological satellites at corresponding hydrological stations.

3. Precision inspection and evaluation method

3.1 Contingency table approach

Contingency Table Approach (CTA) is a method of marking that the hydrological station observation is taken as the true value to evaluate the forecast value [4]. The contingency table, as shown in table 1, represents a 2 \* 2 matrix. Here, each element of the matrix represents a time measurement and satellite data whether or not reached or exceeds a threshold value. For example, if the observation station at one point reaches or exceeds a threshold, but the satellite data does not reach the threshold, then the number of "Nc" increases by 1.

Table 1 contingency table layout

Hydrological station: Yes Satellite: Yes	$N_A$	Hydrological station: No Satellite: Yes	$N_B$
Hydrological station: Yes Satellite: No	$N_C$	Hydrological station: No Satellite: No	$N_D$

According to the above-mentioned contingency table method, it can be calculated Frequency Bias Index, Probability of Detection, and False Alarm Rate. The deviation Frequency Bias Index is defined as:

$$FBI = \frac{F}{O} = \frac{N_A + N_B}{N_A + N_C} \quad (1)$$

Formula : F is the number of satellites estimated to be equal to or exceeded a given threshold, O is the number of stations measured or equal to or above a given threshold, The deviation score reflects the system on the high side (FBI>1) or the lower system (FBI<1) of the satellite estimated value relative to the rainfall station, its range of change is from O to infinity, and the optimum value is 1.

The accuracy rate POD indicates the correct probability of precipitation estimation, the range of variation is 0 ~ 1, and the optimal value is 1. The false positive rate FAR indicates the probability of precipitation estimation error, the range of variation is 0~1, and the optimal value is 0. The FAR and POD scores are defined as follows:

$$FAR = \frac{N_B}{N_A + N_B} \quad (2)$$

$$FAR = \frac{N_A}{N_A + N_C} \quad (3)$$

Because the above ratings based on the contingency table can only reflect the deviation of the meteorological satellite values above a certain threshold, the meteorological satellite can not be used to determine the error size of the precipitation. however, the error size of meteorological satellite precipitation measurement can not be determined. Therefore, it is necessary to calculate the Mean Relative Error to evaluate the error size of meteorological satellite for precipitation estimation. For a given threshold, MRE is defined as:

$$MRE = \frac{1}{n} \sum_{i=1}^n \left| \frac{T_i - S_i}{S_i} \right| \quad (4)$$

Formula :  $T_i$  is the meteorological satellite rainfall measurement value,  $S_i$  is rainfall measurement value at hydrological station, and  $n$  is the number of rain measured.

CTA method is the most commonly used method of precipitation accuracy evaluation at home and abroad[5].

### 3.2 Fuzzy comprehensive scoring method

Fuzzy comprehensive scoring method is a kind of precipitation forecast scoring method form < Circular on the Interim Provisions Concerning the prediction of surface rainfall in the seven major rivers of China >, China Meteorological Administration, Document number 96, 2002[6]. The method is based on the principle of fuzzy comprehensive evaluation, considering the degree of public feeling, the relationship between fuzzy score and forecast Grade K, prediction error  $E_a$  and grade difference  $\max(k)$  is established, and the accuracy of surface rainfall can be objectively evaluated. The fuzzy scoring formula of level j rainfall forecast is defined as:

$$P = 60 + 40 \left[ 1 - \frac{|R_f - R_0|}{\max(k)} \right] \quad (5)$$

Formula :  $\max(k)$  is the rainfall differential of Grade K rainfall, according to the national meteorological department classification standards, daily rainfall of 0.1 to 10mm is light rain, 10 to 24.9mm is moderate rain, 25 to 49.9mm is heavy rain, and 50 to 99.9mm is rainstorm, for light rain, moderate rain, heavy rain and rainstorm,  $\max(k)$  are 9.9mm, 14.9mm, 24.9mm, 49.9mm.  $R_0$  is Measured rainfall;  $R_f$  is forecasts rainfall.

The first item, "60" is the correct basis for a rain forecast, the second item is the weighted score of grade forecast. the fuzzy score is calculated according to the size of the error, that indicating the extent of the forecast, and the higher the score, the closer the reality is, the higher the public satisfaction.

Rule: when the forecast absolute error is  $< \max(k)$ , it is considered that the rainfall is predicted. When there is no rain or light rain, the value of  $\max(k)$  is small, and the change of the absolute error of prediction has a great influence on the calculation of fuzzy score, and the error range can not be too

large, when rainfall reaches above rainstorm level, max (k) value is larger, the change of the absolute error of prediction has relatively little influence on the calculation of fuzzy score, and the range of error can be enlarged appropriately. Rule: when the absolute error of prediction is less than that of K, the rainfall is predicted; and according to the range of the score P, different evaluation levels are given, The specific evaluation level is shown in table 2.

Table 2 Range Accuracy Rank of different scores

Range of score	Rank
$80 \leq P \leq 100$	Excellent
$70 \leq P < 80$	Good
$60 \leq P < 70$	Medium
$P < 60$	Poor

### 3.3 Correlation coefficient method

Correlation coefficient the statistical index that indicates the degree of correlation between the two elements. For the two elements x and y, if their sample values are  $X_i$  and  $Y_i$  ( $i=1, 2, \dots, n$ ), the correlation coefficients between them:

$$R = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (6)$$

$$\bar{x} = \frac{1}{n} \sum_{i=0}^n x_i, \bar{y} = \frac{1}{n} \sum_{i=0}^n y_i \quad (7)$$

$$BIAS = \frac{\sum_{i=0}^n (x_i - y_i)}{\sum_{i=0}^n y_i} \quad (8)$$

Formula : n is the total number of data records, X is the precipitation provided by meteorological satellite precipitation data, Y is the observed precipitation at ground hydrological stations over the corresponding time, BIAS is relative error.

In practice, firstly, the coordinates of the grid units in the site are calculated by the latitude and longitude of the hydrological station, and then the precipitation in the grid unit is calculated as the value of the precipitation in the grid unit through MATLAB programming language.

R reflects the consistency between meteorological satellite precipitation data and site measured precipitation values. The range of values is [0, 1], and the closer to 1, the better the data consistency. BIAS reflects the deviation of the meteorological satellite precipitation data from the measured precipitation value of the site. The closer the BIAS is to 0, the more accurate the data[3].

## 4. Precision inspection and evaluation results

### 4.1 CTA approach evaluation result

The method of contingency table was used to quantitatively evaluate the accuracy of meteorological satellite rain measuring products. Table 3 is the scoring table of the meteorological satellite grading method. the FBI deviation score ( in figure 2) shows that the meteorological satellite measurements of 0.5mm light rain were better, FBI is close to 1. For light rain less than 0.5mm, the measurement result is too large, FBI is greater than 1. However The FBI value decreases gradually after the rainfall greater than 0.5mm, and more than 0.5mm of light rain, moderate rain, heavy rain and even rainstorm measurements are small, FBI were less than 1. The accuracy rate of POD shows (in figure 2) that the higher the accuracy of estimation(POD) about the meteorological satellite's estimate of the smaller rainfall. The false positive rate FAR shows (in figure 2) that the false positive rate of meteorological satellites decreases gradually with the increase of rainfall, after 4mm, it is basically 0, which means that if the weather satellite data is more than 4mm, the rainfall forecast is basically accurate, while the false positive rate of drizzle is higher.

The average MRE score of relative error absolute value shows ( in figure 2) that the MRE score was smaller for the meteorological satellite calculates the light rain below 0.5mm, for larger than 0.5mm, the MRE score is about 0.8, which means that the relative error of the rainfall greater than 0.5mm is maintained at about 0.8.

Table 3 scores of CTA approach

Threshold	FBI	FAR	POD	MRE
0.1	1.72	0.45	0.94	0.15
0.5	1	0.26	0.74	0.66
1	0.8	0.25	0.6	0.71
1.5	0.71	0.17	0.49	0.78
2	0.48	0.07	0.45	0.82
2.5	0.32	0.06	0.29	0.82
3	0.11	0.05	0.09	0.82
3.5	0.08	0.02	0.05	0.82
4	0.02	0	0.02	0.82

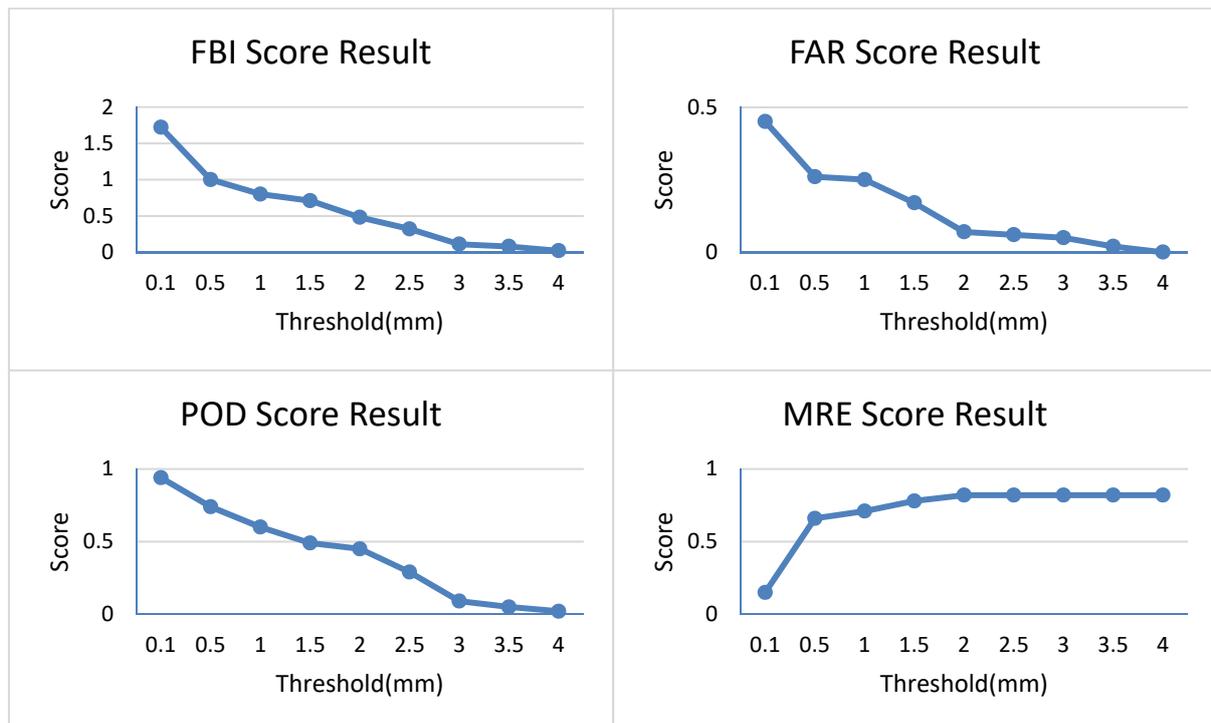


Figure 2 trend lines of FBI, FAR, POD and MRE score

#### 4.2 Fuzzy comprehensive scoring method result

Taking hydrological site data as the true value, the statistics were carried out for all 1980~1982 years of all screenings of precipitation at Babao River Ebo Station, and the average value of fuzzy score is P. Table 4 is the table of fuzzy comprehensive scoring results. it can be seen from table 4 that the accuracy of satellite rainfall estimation is Light rain > Moderate rain > Heavy Rain > Rainstorm. The evaluation levels for light rain, moderate rain, heavy rain and rainstorm are respectively excellent, medium, poor and medium. The evaluation result is the highest in the light rain evaluation, and the lowest score is in the heavy rain evaluation.

Table 4 fuzzy comprehensive evaluation Calculation Table

Daily rainfall	Precipitation type	Precipitation screenings	max (k)	Fuzzy scoring P	Evaluation level
0.4-9.9mm	Light rain	194	9.9	82.96	excellent
10-24.9mm	Moderate rain	59	14.9	65.97	medium
25-49.9mm	Heavy rain	13	24.9	43.77	poor
50-99.9mm	Rainstorm	4	49.9	61.18	medium

#### 4.3 Correlation coefficient method

The measured data for all 1980~1982 years of all screenings of precipitation at Babao River Ebo Station as abscissa, the corresponding meteorological satellite data is used as ordinate to do the correlation map, As shown in figure 3. It was found that the correlation coefficient R was only 0.447, and the correlation between the measured rainfall and the daily precipitation data of the meteorological satellite rain measuring product was poor.

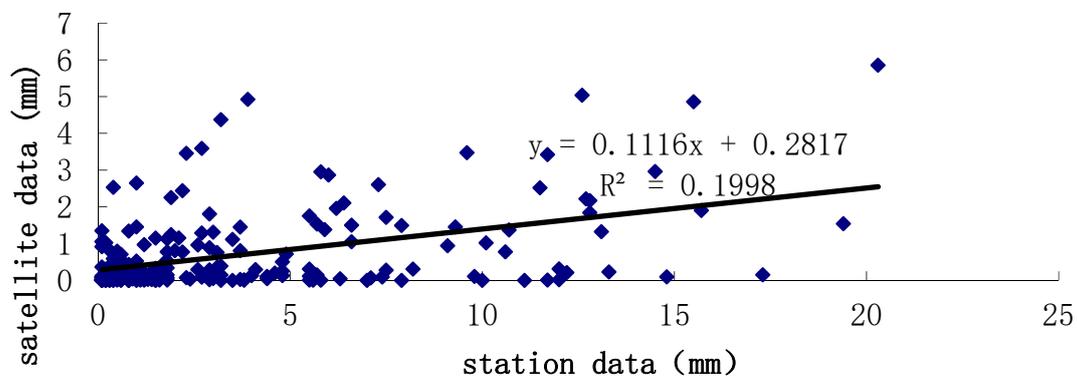


Figure 3 daily precipitation correlation diagram

#### 5. Conclusion

This paper uses the hydrological stations in Heihe basin to measure precipitation, by means of contingency table method, fuzzy comprehensive scoring method and correlation coefficient method, the precision of meteorological satellite precipitation data was tested and evaluated on the daily scale. It is found that the average correlation coefficient between daily precipitation data and corresponding measured precipitation is  $R = 0.447$ , and the correlation is poor. On the basis of the above methods, the grading method of satellite rain measuring products is put forward, which provides a new approach for the inspection and evaluation of the accuracy of the satellite rain measuring products. The rain measuring accuracy of meteorological satellite was checked by the grading method of meteorological satellite rain measuring product. The results show that the meteorological satellite has better measurement of light rain, and the measurement value of moderate rain, heavy rain and rainstorm is too small; The accuracy of rainfall estimation by meteorological satellite is Light rain > Moderate rain > Heavy Rain > rainstorm.

#### References

- [1] Jiang S H, REN L L, Hong Y, et al. Improvement of multi-satellite precipitation products for ensemble streamflow simulation in a middle latitude basin in South China [J]. Water Resources Management, 2014, 28(8): 2259-2278.
- [2] Xu Zhifang, Dang Renqing, Ge Wenzhong. Application of the satellite data to numerical simulation of heavy rainfall [J]. Advances In Water Science, 2005, 16(3): 42-45.
- [3] Hu Qing-fang, Yang Da-wen, Wang Yin-tang, et al. Characteristics and sources of errors in daily TRMM precipitation product over Ganjiang River basin in China [J]. Advances in Water

- Science, 2013, 24(6): 794-800.
- [4] Sasaki Y. Some Basic Formalisms in Numerical Variational Analysis [J]. Monthly Weather Review, 1970, 98(12): 875-883
- [5] Chen Guiying, Zhao Zhenguo. Assessment Methods Of Short Range Climate Prediction And Their Operational Application. [J]. Quarterly Journal Of Applied Meteorology, 1998, 9(2): 178-185.
- [6] Hao Zhenli, Lv Liangjun. Fuzzy Evaluation Of 6h Rainfall Forecasting Method [J]. Yellow River, 2007, 29(5): 23-23.