Article

Delineation of Rain Areas with TRMM Microwave Observations Based on PNN

Shiguang Xu [1,†], Chaoyang Wu [1,*], Alemu Gonsamo [2,†] and Yan Shen [3,†]

[1] The State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100101, China; E-Mail: haibinggis@126.com
[2] Department of Geography and Program in Planning, University of Toronto, Toronto, ON M5S 3G3, Canada; E-Mail: gonsamoa@geog.utoronto.ca
[3] National Meteorological Information Center (NMIC), China Meteorological Administration (CMA), Beijing100081, China; E-Mail: sheny@cma.gov.cn

[†] These authors contributed equally to this work.

[*] Author to whom correspondence should be addressed; E-Mail: hefery@163.com;
Tel.: +86-8725-2343.

Abstract: False alarm and misdetected precipitation are prominent drawbacks of high-resolution satellite precipitation datasets, and they usually lead to serious uncertainty in hydrological and meteorological applications. In order to provide accurate rain area delineation for retrieving high-resolution precipitation datasets using satellite microwave observations, a probabilistic neural network (PNN)-based rain area delineation method was developed with rain gauge observations over the Yangtze River Basin and three parameters, including polarization corrected temperature at 85 GHz, difference of brightness temperature at vertically polarized 37 and 19 GHz channels (termed as TB_{37V} and TB_{19V}, respectively) and the sum of TB_{37V} and TB_{19V} derived from the observations of the Tropical Rainfall Measuring Mission (TRMM) Microwave Imager (TMI). The PNN method was validated with independent samples, and the performance of this method was compared with dynamic cluster K-means method, TRMM Microwave Imager (TMI) Level 2 Hydrometeor Profile Product and the threshold method used in the Scatter Index (SI), a widely used microwave-based precipitation retrieval algorithm. Independent validation indicated that the PNN method can provide more reasonable rain areas than the other three methods.

Furthermore, the precipitation volumes estimated by the SI algorithm were significantly improved by substituting the PNN method for the threshold method in the traditional SI algorithm. This study suggests that PNN is a promising way to obtain reasonable rain areas with satellite observations, and the development of an accurate rain area delineation method deserves more attention for improving the accuracy of satellite precipitation datasets.

1. Introduction

Precipitation has great significance in the study of ecology, hydrology and meteorology [1,2], but it is still a great challenge to acquire the spatial and temporal distribution of precipitation over many developing countries and mountainous regions, due to the sparse rain gauge network. With the development of meteorological satellite networks, the information provided by space-borne sensors has become the main focus for estimating precipitation over un-gauged regions [3,4]. In the past decades, various algorithms have been developed to retrieve precipitation from infrared and microwave satellite observations [5–10]. However, considerable studies revealed that current satellite precipitation datasets contain enormous biases and random errors that could lead to serious uncertainty in many scientific applications [4,11,12].

Tian *et al.* [13] separated the bias of high-resolution satellite precipitation datasets into three independent components: misdetected bias (rain areas that were incorrectly determined as no rain areas by satellite precipitation datasets), false alarm bias (no rain areas that were incorrectly determined as rain areas by satellite precipitation datasets) and hit bias (the rain areas correctly determined by satellite precipitation datasets, but for which the precipitation volumes were inaccurately estimated). Furthermore, they indicated that the false alarm precipitation was a leading error source in many high-resolution satellite precipitation datasets. Gebregiorgis *et al.* [14] revealed that more than 50% of the rainfall biases were caused by misdetected precipitation during winter and that 28% were caused by false alarm precipitation. Gebregiorgis and Hossain [15] pointed out that the uncertainty in the Climate Prediction Center morphing (CMORPH) and Tropical Rainfall Measuring Mission 3B42 was governed by misdetected precipitation and hit bias. From these studies, it can be inferred that a great part of uncertainty in satellite estimates is governed by misclassified rain/no rain areas. Consequently, developing a more accurate rain area delineation method is of great significance for improving the performance of high-resolution satellite precipitation datasets.

Benefitting from the better penetrability of microwave radiation, space-borne microwave sensors can provide information on the horizontal and vertical distribution of hydrometeors inside the cloud system [16]. Therefore, retrieving precipitation information from microwave observations has attracted considerable attention for both global and regional applications [17,18]. Since the 1980s, several methods have been developed to delineate rain areas on the basis of multi-channel passive microwave observations. Grody [19] proposed the polarization-corrected temperature (PCT) method using the polarization difference between the land surface and hydrometeors in order to identify the rain pixels in passive microwave TBs

(brightness temperatures) images. Grody [20] and Ferraro [21] developed the Scatter Index at 85 and 37 GHz to measure the scattering intensity of clouds and eventually retrieved global rain areas with a threshold. Yao *et al.* [22] introduced the dynamic cluster K-means method to delineate rain areas over Tibetan Plateau using multi-channel microwave brightness temperatures. Moreover, several threshold methods based on brightness temperatures at passive microwave (PMW) channels have been proposed to separate rain areas from land surface [23–25]. However, due to the spatially and temporally varying relationship between the occurrence of precipitation and the brightness temperatures at PMW channels, these methods usually cannot identify the rain/no rain pixels accurately.

Artificial neural networks (ANNs) are a non-parametric intelligent tool that is useful for non-linear fitting, pattern recognition and process control [26]. Since the 1990s, several types of ANN have been developed to retrieve rain information from satellite observations. For example, Hsu *et al.* [27] proposed an ANN-based precipitation retrieval algorithm, called precipitation estimation from remotely-sensed information using artificial neural networks (PERSIANN), to estimate rain rate with visible and infrared data; Mahesh, Prakash, Sathiyamoorthy and Gairola [17] estimated the rain rate over land and oceans using ANN with the SI and PCT derived from microwave observations as the input and the radar-derived precipitation data as the output. Previous studies showed the remarkable capability of ANN to illustrate the complicated relationship between the satellite observations and precipitation and indicated that it is a promising approach to accurately delineate rain areas.

In this study, a Bayes-based ANN, termed a probabilistic neural network (PNN), was introduced to delineate rain areas using the brightness temperature derived from Tropical Rainfall Measuring Mission (TRMM) Microwave Imager (TMI). Furthermore, the PNN result was used to improve the Scatter Index (SI) microwave-based precipitation retrieval algorithm, in order to investigate the influence of this new rain area delineation method on the high-resolution satellite precipitation retrieval algorithm.

2. Data and Method

2.1. Study Area

The Yangtze River Basin is located in southern China with an area of about 1.8×10^6 km^2. This Basin contains the Yangtze River (the third longest river in the world), numerous tributaries (such as the LanCang River) and several large lakes (such as Poyang Lake and Dongting Lake). The main source of water in this basin is the precipitation from the Asian monsoon. Therefore, the precipitation over the Yangtze River Basin not only has great significance to ecology and climate over East Asia, but also affects the livelihood of a million people along this river [28]. The study area of this study ranges from 27.5°N to 32°N and 102.6°E to 120.5°E (Figure 1). The temporal distribution of precipitation over this area shows apparently seasonal characteristics, with most of the precipitation happening between June and October [29]. Therefore, the summer season (June, July and August) was selected in this study.

2.2. Data

2.2.1. Rain Gauge Observations

The actual precipitation used in this study is the hourly precipitation measured by 3045 rain gauges over the Yangtze River Basin. These rain gauges measured the precipitation with an intensity above 0.1 mm/h at their locations. The measured precipitation datasets were processed and administrated by the China Meteorological Administration (CMA). The spatial distribution of these rain gauges is presented as red points in Figure 1. The dense rain gauge network over the study area could effectively capture the rainfall event over this area and provide sufficient samples to establish and validate rain area delineation methods.

Figure 1. The topography of the study area and the spatial distribution of the rain gauge network in this study.
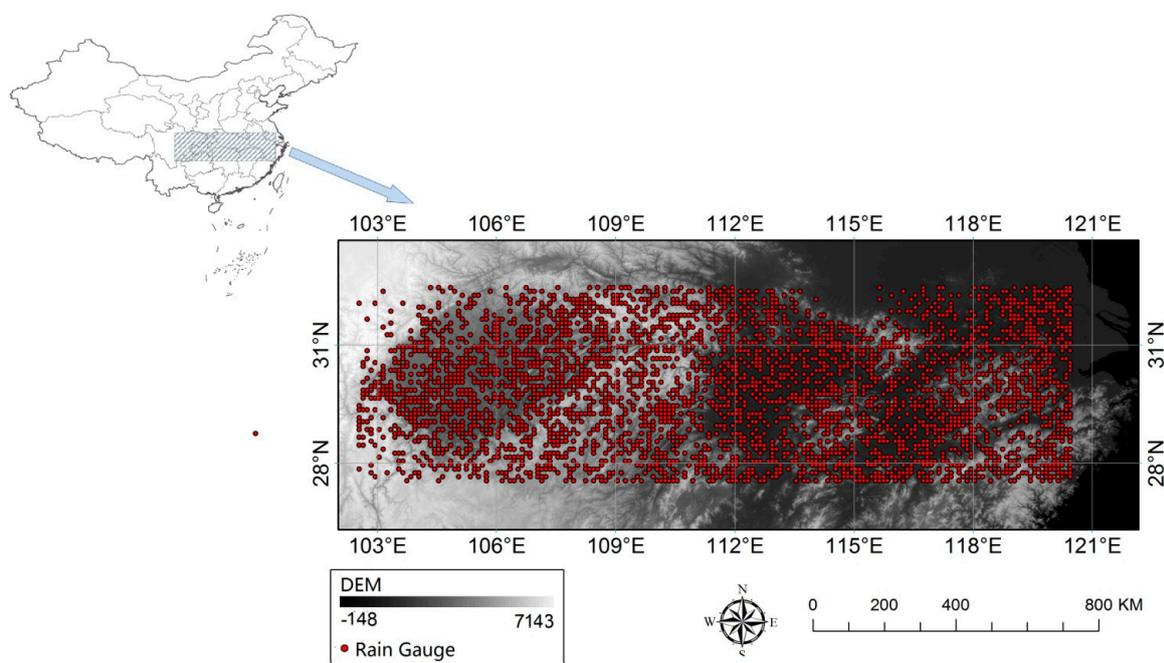
2.2.2. Microwave Brightness Temperature Datasets

In this study, the brightness temperature datasets measured by the Tropical Rainfall Measuring Mission (TRMM) Microwave Imager (TMI) were introduced. TRMM is a joint-mission conducted by America and Japan to monitor tropical rainfall and to estimate its associated latent heating between 38°S and 38°N. This satellite carries three rainfall measuring sensors, including the Visible and Infrared Scanner (VIRS), Precipitation Radar (PR) and TMI. The TMI is designed to observe the microwave radiation of Earth at 10.7, 19.4, 21.3, 37.0 and 85.5 GHz. Except for 21.3 GHz, all frequencies are double-polarized (e.g., one vertically-polarized channel and one horizontally-polarized channel), making it possible to acquire the polarization difference between the land surface and precipitation. The brightness temperatures (termed TBs, hereafter) of these nine channels are provided by the TRMM Goddard Earth Science Data and Information Services Center (GESDISC), with the name TRMM 1B11.

In this dataset, the TBs at nine channels, the time of each scan and the geo-locations of all pixels (e.g., location of the center of the instantaneous field of view (IFOV) at the altitude of the Earth ellipsoid) are stored in the Hierarchical Data Format (HDF). The horizontal footprint spatial resolution of TRMM 1B11 at 85.5 GHz is about 5 km and about 10 km at other four frequencies. Compared with another space-borne microwave sensor, the Special Sensor Microwave/Imager (SSM/I), the TMI could provide higher spatial resolution microwave observations, making it more suitable to monitor the precipitation at the meso- and micro-scale.

2.2.3. TRMM Microwave Imager (TMI) Level 2 Hydrometeor Profile Product

In this study, the latest version (version 7) of the TRMM Microwave Imager (TMI) Level 2 Hydrometeor Profile Product (termed TRMM 2A12, hereafter) was introduced as a comparison for the new rain area delineation method proposed by this study. TRMM 2A12 contains the vertical profiles of hydrometeors, as well as precipitation derived from the TRMM 1B11 brightness temperature using the Goddard profiling algorithm (termed GPROF), which estimates the precipitation information at various vertical heights by blending the TMI observations with dynamical cloud models using the Bayesian method. The vertical spatial resolutions of TRMM 2A12 are 0.5 km (from the surface to 4 km), 1 km (from 4 to 6 km), 2 km (from 6 to 10 km) and 4 km (from 10 to 18 km). The horizontal footprint spatial resolution is about 5 km, which is in accordance with the spatial resolution of TRMM 1B11.

2.3. Methods

2.3.1. Scatter Index

The Scatter Index (SI) is a widely used microwave-based precipitation retrieval algorithm proposed by Grody [20]. Theoretically, the high-frequency microwave radiation emitted by the land surface could be seriously scattered by the rainfall particles (such as big raindrops and ice crystals) in the clouds, so the TBs at high frequency over rain areas are usually lower than that over no rain areas. In contrast, the low-frequency microwave radiation tends to be less scattered by rainfall particles, leading to similar TBs at low frequency over rain and no rain areas. Based on this theory, Grody [20] assumed that the TBs at high frequency over the no rain areas could be simulated by the TBs at low frequency, and the difference between the actual and simulated TBs at high frequency could consequently present the scattering intensity of the rainfall particles in the cloud, which is correlated with the spatial distribution and the physical attributes of the rainfall particles in the clouds. Therefore, the SI is a reasonable indicator of the occurrence and intensity of precipitation. According to the research of Grody [20] and Ferraro [21], SI could be derived from vertically-polarized brightness temperatures at 21.3, 19.0 and 85.5 GHz (termed TB_{21V}, TB_{19V} and TB_{85V}) using a two-step algorithm. First, the brightness temperature at 85 GHz under the no scattering condition at the location k is defined as $E(k)$ and calculated by Equation (1).

$$E(k) = A*TB_{21V}(k)^2 + B*TB_{21V}(k) + C*TB_{19V}(k) + D \qquad (1)$$

The coefficients *A*, *B*, *C* and *D* in this equation are regressed using the brightness temperature of the pixels under the no rain condition. Second, the SI is calculated using Equation (2).

$$SI(k) = TB_{85V}(k) - E(k) \tag{2}$$

After the *SI* is calculated, the rain areas were delineated with a threshold, and then, the precipitation of the test dataset was calculated using delineated rain areas according to Equation (3),

$$P(y) = m * SI(y)^n \tag{3}$$

P(*y*) is the estimated precipitation at the raining location *y*; the coefficients *m* and *n* are acquired by regressing the *SI* and the precipitation of samples under raining conditions in the training datasets.

2.3.2. Dynamic Cluster K-Means Method

To delineate rain areas over the Tibetan Plateau with a sparse rain gauge network, Yao, Li, Zhu, Zhao and Chen [22] attempted to delineate rain/no rain areas with the TBs measured by TMI using an unsupervised classification method, the dynamic cluster K-means (termed as Kmean, hereafter). In this study, the performance of this method was assessed as a comparison of the new rain area delineation method proposed by this study. In the Kmean method, several categories were first defined, and the centers were calculated by averaging the values of similar samples. Then, the distance between a sample and a category center was calculated. For each category, the square sum of the distance between this category and the neighboring samples is called the classification error. Subsequently, the Kmean method attempts to make the classification error as small as possible by assigning each sample to the nearest category. In this step, the distances between a sample *i* and all category centers were first calculated, and then, the sample *i* was assigned to the category with the smallest distance. At the same time, the category center is re-calculated. This process iterated until the smallest classification error was obtained. The advantage of this method is that ground observations are not necessary. According to Yao, Li, Zhu, Zhao and Chen [22], the TBs at 19.4, 21.3, 37.0 and 85.5 GHz were used as the inputs of the Kmean method to delineate rain/no rain areas. In this study, the Kmean method was conducted using the "kmeans" toolkit provide by the MATLAB 2009a software.

### 2.3.3. Brief Description of Probabilistic Neural Network

The probabilistic neural network (PNN) proposed by Specht [30] is a network based on the Bayesian network and kernel Fisher discriminant analysis [31]. It consists of three layers: the input layer, the radial basis layer and the competitive layer. The radial basis layer uses radial basis function to measure the distance between the input vectors and row weight vectors in a weight matrix. Then, the competitive Layer works out the maximum value of these probabilities, and produces a 1 for that class and a 0 for the other classes. Compared with other artificial neural networks, such as back propagation, the PNN can provide a Bayes-optimal classification at considerably faster training speeds [32]. The structure of the PNN is briefly described in Figure 2. *X*, *Y* and *H* in this figure represent input variables, neurons in the output layer and neurons in the hidden layer, respectively, while the variables *k*, *m* and *n* are the number of input vectors, output neurons and hidden neurons, respectively.

2.3.4. Rain Area Delineation Method Based on PNN

With the ground observations of the rain gauge network over the Yangtze River Basin, the PNN-based rain area delineation method was established according to the following steps:

(1) The multi-channel TBs stored in TRMM 1B11 datasets were extracted and pre-processed. First, the multi-channel TBs at the nine channels (including the vertical and horizontal polarization at 10.7, 19.4, 37.0 and 85.5 GHz, and the vertical polarization at 21.3 GHz), as well as the geo-locations of each pixel were extracted using the "hdftool" of the MATLAB 2009a software, generating a dataset composed of the latitude, longitude and the TBs at the nine channels of each pixel. Then, this dataset was resampled to 0.1° × 0.1° spatial resolution using the nearest neighboring method provide by ArcGIS software.

Figure 2. The structure of the probabilistic neural network.
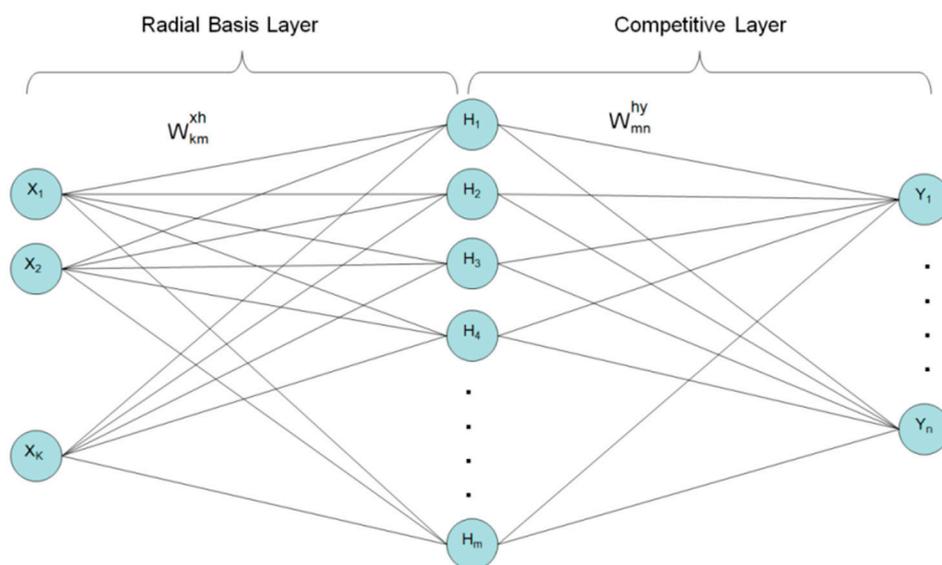
(2) Three parameters, which have the potential to distinguish the rainfall areas from underlying surface, were derived from the resampled TBs. The first parameter is the polarization-corrected temperature (PCT) at 85 GHz (termed PCT_{85}). PCT was proposed by Grody [19] to delineate the rain areas based on the scattering effect of raindrops on the high-frequency microwave radiation, as well as the difference of the microwave polarization between the land surface and rain areas. High values of PCT usually indicate intense raindrops in the clouds and a high probability of raining. Kidd [33] investigated the relationship between the occurrence of rain and the PCT at different frequencies, and the results indicated that the PCT at 85 GHz was sensitive to light precipitation and more suitable for the delineation of the rain areas than other channels. PCT_{85} can be calculated using Equation (4):

$$PCT_{85} = TB_{85V} + \mu * (TB_{85V} - TB_{85H}) \tag{4}$$

In this equation, TB_{85H} is the horizontally polarized brightness temperature at 85 GHz and μ is a constant value (0.818), according to Spencer *et al.* [34].

The second one is the difference of TBs at high frequency and low frequency (termed TD). Ferraro, Smith, Berg and Huffman [25] pointed out that the TBs of land surface (except snow, desert) at high frequency is higher or similar in comparison with those at low frequency; therefore, the TD tends to be

positive over no rain areas. However, because of the serious scattering effect of the raindrops on high-frequency microwave radiation, the TBs at high frequency showed a substantial decrease, leading to negative TD over rain areas. Therefore, the TD is a promising indicator of rain areas. In this study, the vertically polarized TBs at 37 GHz (high frequency) and 19 GHz (high frequency) were used to calculate TD according to the research of Yao, Li, Zhu, Zhao and Chen [22].

The third variable is the sum of TBs at 37 and 19 GHz (termed TS). Because the emissivity of water is much lower than land surface, the water body usually shows low TBs, which is similar to rain areas. As addressed by Yao, Li, Zhu, Zhao and Chen [22], the sum of TBs at 37 and 19 GHz (TS) was an effective indicator to identify the water body, which usually showed a much lower value of TS than other land surface and rain areas.

(3) The three parameters calculated in Step (2) were collocated with the ground observations measured by the rain gauges. For the pixel *i* (with a spatial resolution of 0.1° × 0.1°), the rain gauges located inside pixel *i* were first selected, and then, the mean value of the selected rain gauge observations was calculated as the true precipitation for this pixel. Pixels containing at least one rain gauge were selected to establish and validate the PNN-based rain area delineation method.

(4) Based on the ground observation, a rain flag was generated for each pixel to determine whether this pixel was under raining conditions. The pixels with ground observations above 0 mm/h were deemed under raining conditions, and the rain flag was set to 2, while the pixels with ground observations equal to 0 mm/h were deemed as no raining, and the rain flag of these pixel was set to 1. In order to independently validate the new rain area delineation method, 30% of the pixels were randomly selected as the training datasets, while the rest of the pixels were used to assess the performance of the established rain area delineation method.

(5) The PNN was trained with the three parameters (PCT, TD and TS) of the training datasets as inputs and the corresponding rain flag as the output. In this study, the PNN model was designed and trained by the "newpnn" toolkit provided by MATLAB 2009a. This toolkit could establish a PNN classification model by taking a matrix as input vectors and a matrix with classification information as target class vectors. In this toolkit, a parameter was included to control the spread of radial basis functions in PNN. If this parameter is near zero, the network will be treated as a nearest neighbor classifier. As this parameter becomes larger, the designed network will take into account several nearby input vectors and, thus, takes more time to train the PNN. In this study, the default value of this parameter provided by the MATLAB 2009a software (0.1) was used to establish the PNN model. This step will establish a PNN model that could determine whether a sample was under raining conditions.

(6) The PNN-based rain area delineation method was validated using the test datasets generated in Step (4). In this step, the PNN generates a rain flag for each pixel (1 for the no rain pixels and 2 for the rain pixels). Subsequently, the rain flags determined by PNN were compared with the rain flags determined by the ground observations to assess the performance of the PNN method.

2.3.5. Validation

In this study, six categorical statistics, including the probability of detection (POD), false alarm ratio (FAR), critical success index (CSI), equitable threat score (ETS), Hanssen and Kuiper discriminant (HK) and Heidke skill score (HSS), were introduced to evaluate the performance of the rain area delineation

methods [35]. The POD and FAR were used to measure the misdetected and false alarm precipitation of the rain area delineated by satellite observations, respectively. Higher values of POD and lower FAR indicate less misclassified rain/no rain areas. CSI, ETS, HK and HSS were used to evaluate the strength of the correlation between the rain areas measured by rain gauges and those delineated by satellite observations. More accurate rain area delineation will produce higher values of CSI, ETS, HK and HSS. These categorical statistics are calculated by Equations (5)–(10):

$$POD = \frac{h}{m+h} \tag{5}$$

$$FAR = \frac{f}{f+h} \tag{6}$$

$$CSI = \frac{h}{m+h+f} \tag{7}$$

$$ETS = \frac{h - ((h+m)*(h+f)/(z+f+m+h))}{m+h+f - ((h+m)*(h+f)/(z+f+m+h))} \tag{8}$$

$$HK = \frac{h}{m+h} - \frac{f}{f+z} \tag{9}$$

$$HSS = \frac{2*(z*h - f*m)}{(z+f)*(f+h) + (m+h)*(z+m)} \tag{10}$$

In these equations, h and z represent the number of the correctly-identified rain pixels and no rain pixels, respectively; f is the number of false alarm pixels that are under no raining conditions, as detected by rain gauges, but still misclassified as rain by satellite observations; m is the number of misdetected pixels, which are rain pixels detected by rain gauges, but misclassified as no rain by satellite observations. The meanings of these variables are presented in Table 1.

Table 1. The contingency table of rain areas detected by rain gauges and delineated by rain area delineation methods.

		Rain/No Rain Area Delineation by the Models	
		Rain	**No Rain**
Ground observations	Rain	h	m
	No rain	f	z

3. Results

3.1. The Results of PNN and Other Rain Area Delineation Methods

In this study, 16 rain events over the Yangtze River Basin through June, July and August 2009, were observed by the rain gauge network. The detailed information of these rain events is presented in Table 2. In this section, the results of the PNN-based rain area delineation method (termed PNN) were validated by calculating six categorical statistics described in Section 2.2.3.

The rain event of 26 July 2009, at 9 am was taken as an example to present the results of the Kmean, TRMM 2A12, SI_threshold and PNN rain area delineation methods. Figure 3 shows the rain areas

detected by rain gauge network and delineated by the Kmean, TRMM 2A12, SI_threshold and PNN, respectively. Although the rain areas shared a similar spatial pattern, there were numerous differences among them. For example, the Kmean, TRMM 2A12 and SI_threshold seriously under-estimated the rain areas in the rectangle shown in Figure 3. In contrast, the PNN can detect the rain areas in this region more accurately compared with the other three methods.

Table 2. The information of the 16 rain events over the Yangtze River Basin through June, July and August 2009.

Rain Event	Time	Number of Rain Pixels	Number of No_Rain Pixels
1	June 25 at 2 am	419	2757
2	June 27 at 8 pm	549	2607
3	June 28 at 7 pm	879	2831
4	June 29 at 0 am	1048	2286
5	July 2 at 5 pm	242	3470
6	July 6 at 3 pm	128	3518
7	July 23 at 7 pm	453	2807
8	July 26 at 4 am	482	2215
9	July 26 at 9 am	926	2662
10	July 27 at 8 am	483	2525
11	July 29 at 3 am	330	3268
12	July 31 at 3 am	483	2470
13	August 4 at 1 am	446	2646
14	August 5 at 0 am	317	2935
15	August 5 at 5 am	295	3349
16	August 9 at 9 pm	621	2297

Figure 4 presents the spatial distribution of the misdetected (the red circles) and the false alarm (the green circles) precipitation of the rain area delineated by the Kmean, TRMM 2A12, SI_threshold and PNN, respectively. The false alarm precipitation of the Kmean and TRMM 2A12 was 201 and 225, while the misdetected precipitation of them was 295 and 231, respectively. The SI_threshold produced 489 misclassified precipitations, including 239 misdetected precipitations and 250 false alarm precipitations. Compared with the three rain area delineation methods mentioned above, the PNN produced the smallest amount of false alarm precipitation (129) and misdetected precipitation (180). Besides, the CSI, ETS, HK and HSS of PNN (0.63, 0.53, 0.68 and 0.69) were obviously higher than the Kmean (0.42, 0.32, 0.45 and 0.49), TRMM 2A12 (0.48, 0.35, 0.52 and 0.52) and SI_threshold (0.46, 0.33, 0.51 and 0.49). Therefore, it can be concluded that the PNN could provide more reasonable rain area delineation compared with the other three methods.

3.2. Validation and Comparison of the PNN Method

First, the performance of the PNN method was compared with the threshold method used in the conventional SI method (SI_threshold). Unlike the constant threshold (SI > 10) proposed by Grody [20], the threshold of each case was selected by maximizing the value of HSS according to the research of Haile *et al.* [36]. In this study, the thresholds of SI ranged from 4.7 to 6.4 with a mean value of 5.1.

Figure 3. The rain areas observed by (**a**) rain gauge and delineated by (**b**) dynamic cluster K-means (Kmean); (**c**) TRMM Microwave Imager (TMI) Level 2 Hydrometeor Profile Product (TRMM 2A12); (**d**) threshold method of Scatter Index (SI_threshold); and (**e**) probabilistic neural network (PNN) methods. (On 26 July 2009, at 9 am).
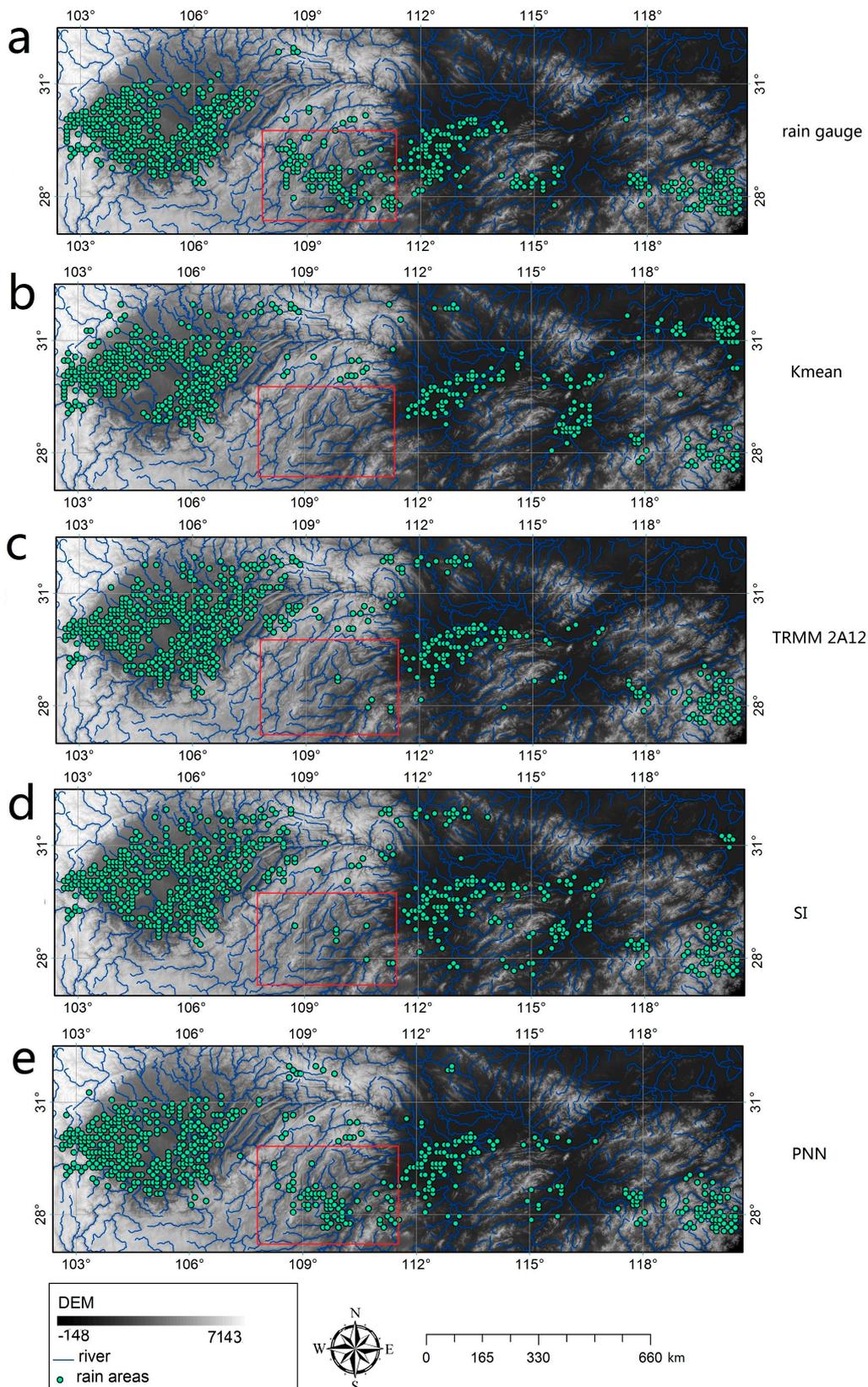
Figure 4. The misdetected and false alarm precipitation of the rain area delineated by (**a**) dynamic cluster K-means (Kmean); (**b**) TRMM Microwave Imager (TMI) Level 2 Hydrometeor Profile Product (TRMM 2A12); (**c**) threshold method of Scatter Index (SI_threshold); and (**d**) probabilistic neural network (PNN) methods (On 26 July 2009, at 9 am).
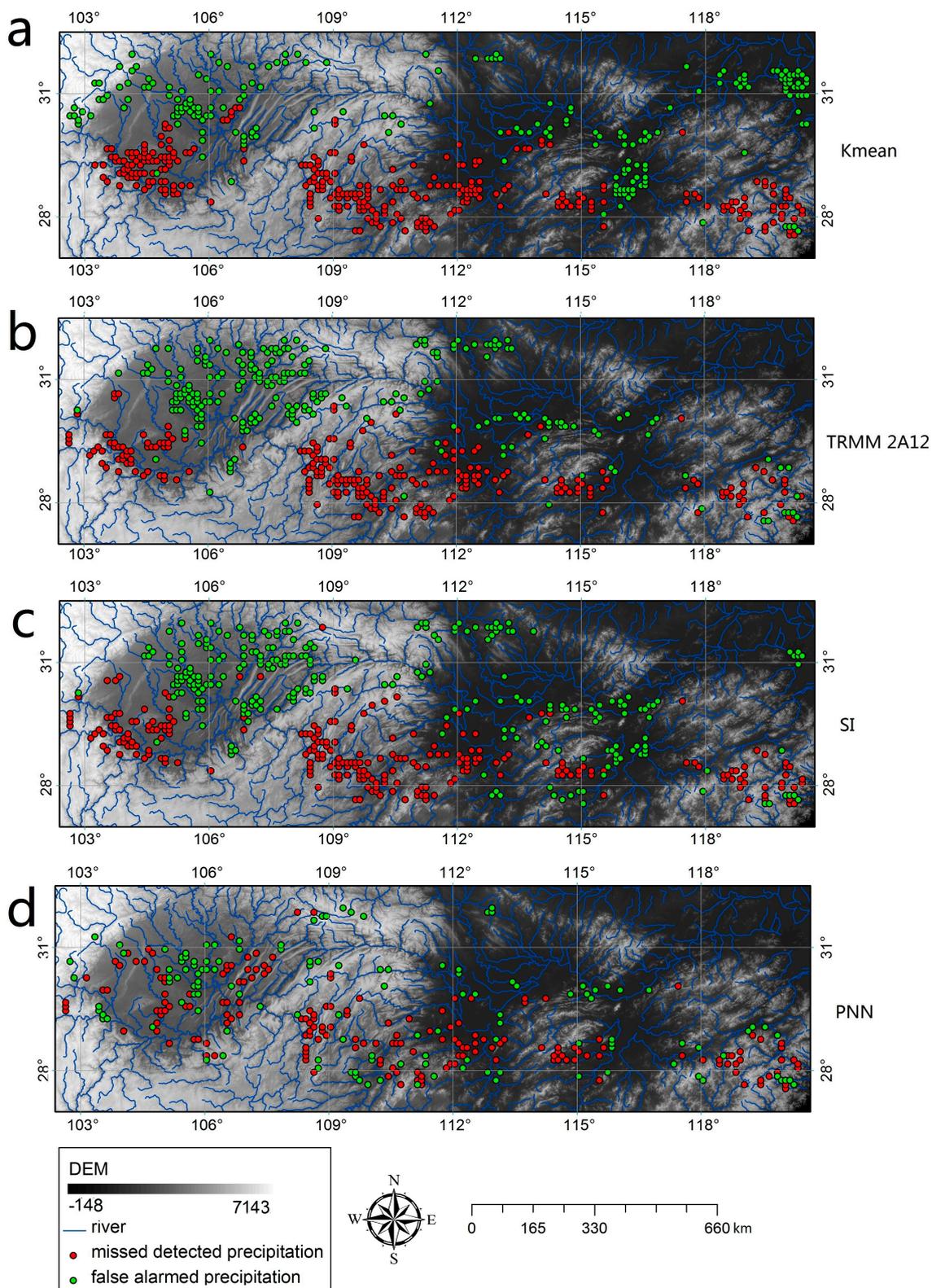
Figure 5. The results of (**a**) false alarm ratio (FAR), (**b**) probability of detection (POD), (**c**) equitable threat score (ETS), (**d**) critical success index (CSI); (**e**) Hanssen and Kuiper discriminant (HK) and (**f**) Heidke skill score (HSS) of the rain area delineated by the threshold method of Scatter Index (SI_threshold), probabilistic neural network (PNN), TRMM Microwave Imager (TMI) Level 2 Hydrometeor Profile Product (TRMM 2A12) and dynamic cluster K-means (Kmean) methods.
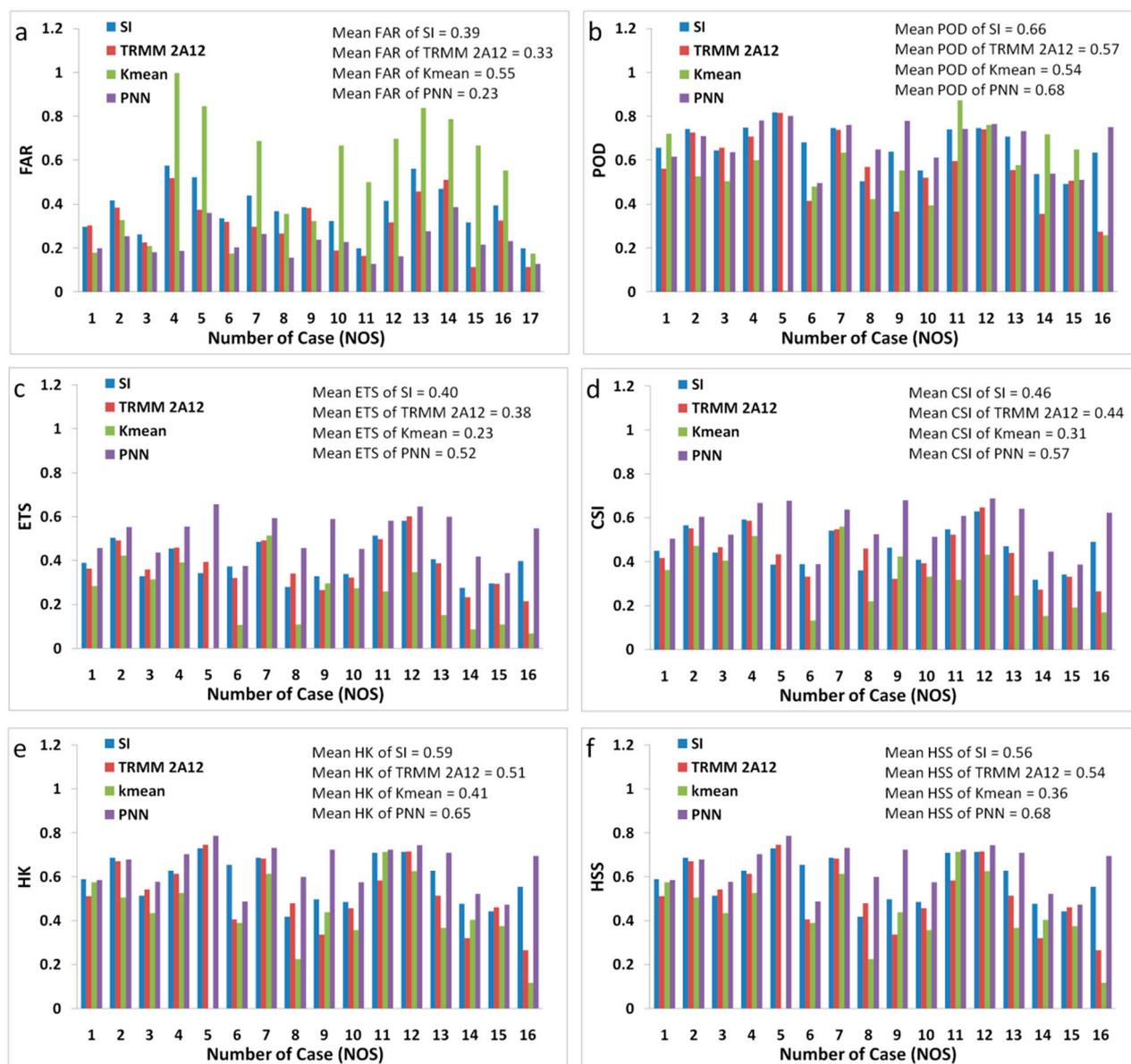

As shown in Figure 5, the FAR of SI_threshold ranged from 0.21–0.58 with a mean value of 0.39, whereas the FAR of PNN decreased to 0.12–0.38 with a mean value of 0.23 (Figure 5a), indicating that the PNN contained less false alarm precipitation than SI_threshold. The POD of PNN does not show an obvious increase compared with SI_threshold (Figure 5b), indicating that the amount of misdetected precipitation in PNN is similar or less than SI_threshold in most cases. The mean values of ETS (Figure 5c), CSI (Figure 5d), HK (Figure 5e) and HSS (Figure 5f) of SI_threshold were 0.40, 0.46, 0.59 and 0.56, respectively, whereas the mean values of these categorical statistics of PNN increased to 0.52, 0.57, 0.65 and 0.68. It can be inferred from the results that the PNN method could delineate the rain

areas more accurately compared with the threshold method used in the traditional SI method.

To further validate the performance of the PNN rain delineation method, the rain area delineated by TRMM 2A12 and the results of Kmean were compared. In this study, the surface rainfall stored in TRMM 2A12 was extracted by the "hdftool" of the MATLAB 2009a software and resampled to 0.1° × 0.1° using the nearest neighbor method provided by the ArcGIS software. Then, the rain pixels were selected by the surface rain above 0 mm/h. The six categorical statistics of these two methods are also presented in Figure 5. As indicated by Figure 5a, b, the FAR of TRMM 2A12 (mean value = 0.33) was higher than that of PNN results (mean value = 0.23), whereas the POD of TRMM 2A12 (mean value = 0.57) was lower than PNN (mean value = 0.68). This indicates that the TRMM 2A12 can obtain higher false alarm and misdetected precipitation rates than PNN. The Kmean method presented a higher FAR and a lower POD compared with PNN, indicating that the results of the Kmean contain more false alarm and misdetected precipitation than PNN. Besides, Figure 5c–f shows that the CSI, ETS, HK and HSS of the TRMM 2A12 and Kmean were all smaller than PNN and even smaller than the SI_threshold. Therefore, the PNN method showed an obvious better performance in the delineation of rain areas than the TRMM 2A12 and Kmean methods.

Figure 6. The false alarm ratio (FAR), probability of detection (POD), equitable threat score (ETS), critical success index (CSI), Hanssen and Kuiper discriminant (HK) and Heidke Skill Score (HSS) of the threshold method of the Scatter Index (SI_threshold), probabilistic neural network (PNN), TRMM Microwave Imager (TMI) Level 2 Hydrometeor Profile Product (TRMM 2A12) and dynamic cluster K-means (Kmean) using the summarized test datasets of the 16 rain events.
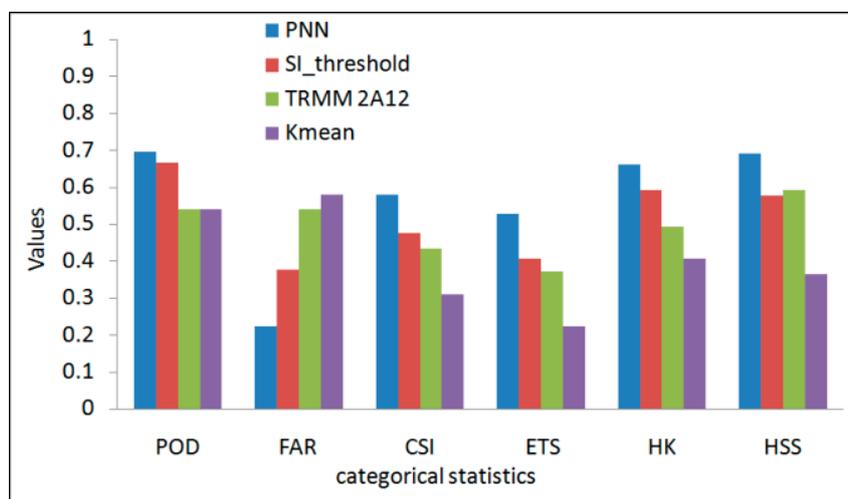
The performances of the four rain area delineation methods were also assessed by calculating the POD, FAR, CSI, ETS, HK and HSS of summarized test datasets (e.g., putting the test datasets of the 16 rain events together). Figure 6 presents the results of the six categorical statistics. It can be seen that the rain area delineated by the PNN method presented the lowest FAR and highest POD, indicating that this method generated the least false alarm and misdetected precipitation. The CSI, ETS, HK and HSS of the PNN were also higher than the other three methods.

Figure 7. (**a**) Mean absolute error (MAE), (**b**) root mean squared error (RMSE) and (**c**) coefficients of determination (R^2) of the precipitation estimated by the Scatter Index using the rain areas delineated by probabilistic neural network (PNN) and delineated by the threshold method.
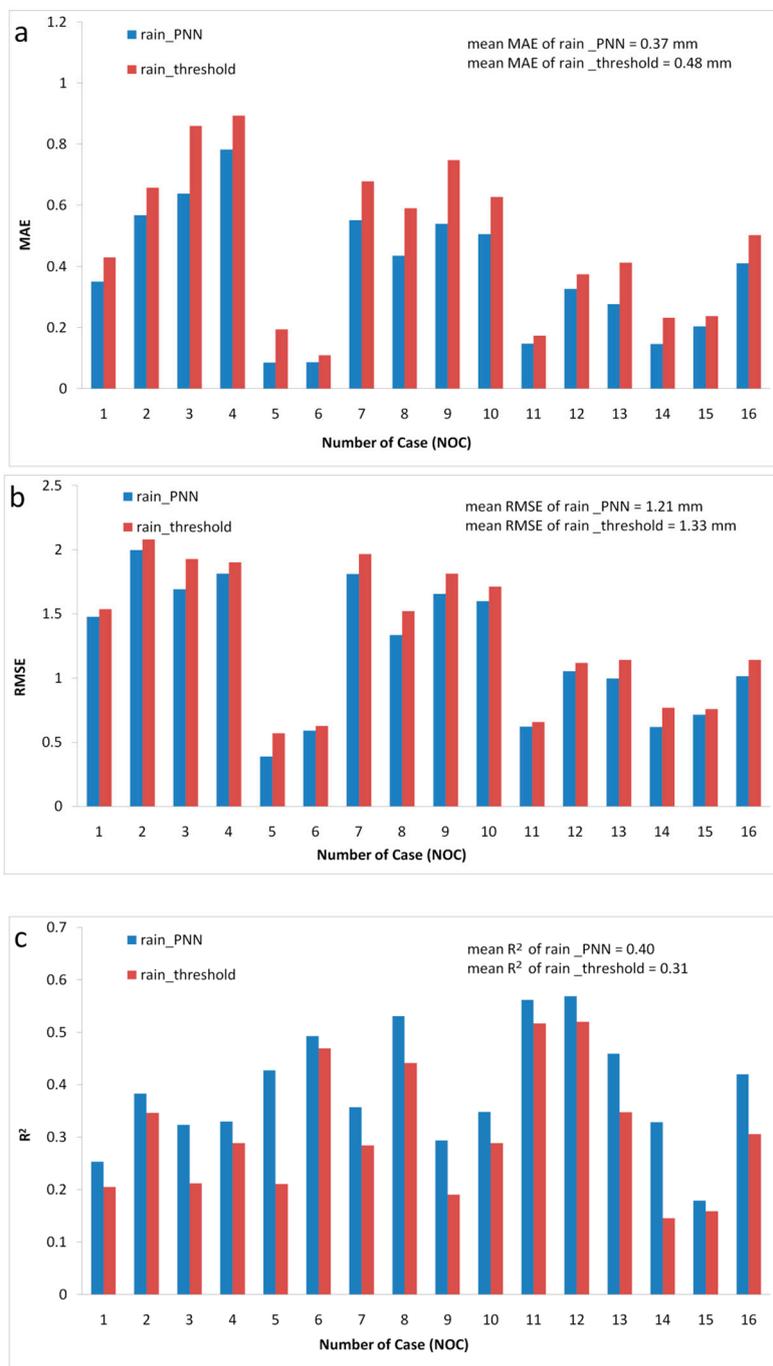
3.3. Improvement of SI with PNN Method

It can be inferred from Section 3.2 that the PNN method can provide more accurate rain area delineation compared with the threshold method used in the traditional SI precipitation retrieval algorithm. Therefore, we investigated whether the PNN method could be used to improve the performance of SI by reducing the misdetected and false alarm precipitation.

In our study, the precipitation of the test dataset in each of the 16 cases was estimated according to Equation (3) using the rain area delineated by PNN and SI_threshold, respectively. The estimated precipitation datasets were termed as rain_PNN and rain_threshold. Figure 7a–c presents the mean absolute error (MAE), root mean squared error (RMSE) and coefficients of determination (R2) of rain_PNN and rain_threshold. As presented by Figure 7, the ranges of the MAE, RMSE and R2 of the Rain_threshold were 0.11–0.84 mm (mean value is 0.45 mm), 1.33–2.21 mm (mean value is 1.33 mm) and 0.11–0.50 mm (mean value is 0.29 mm), respectively. In contrast, the MAE and RMSE of rain_PNN reduced to 0.07–0.8 mm (mean value is 0.39 mm) and 0.42–2.11 mm (mean value is 1.23 mm), respectively, whereas the R2 of rain_PNN increased to 0.16–0.58 mm (mean value is 0.37 mm). The results indicated that the performance of SI can be improved by using the PNN method to delineate rain areas instead of the threshold method.

4. Discussion

It can be inferred from Figure 5 that the threshold method used in the traditional SI method could not provide reasonable rain area delineation over the study area. According to the research of Grody [20] and Ferraro [21], SI is a measurement of the microwave scattering intensity at the high frequencies of the clouds, which is highly related to the distribution of the hydrometeors in clouds. In theory, the high values of SI usually mean a large size and density of water-drops and ice crystals in the clouds, which lead to a high probability of rainfall. However, in fact, the occurrence of precipitation on land not only depends on the physical properties of hydrometeors in the clouds, but also depends on many other factors, such as complicated underlying surface and other meteorological conditions [22]. The disturbance of these factors complicates the relationship between the occurrence of precipitation and the values of SI. Therefore, the rain areas cannot be delineated accurately using the threshold method used in the traditional SI method.

In contrast, the PNN rain delineation method does not rely on a given threshold, but depends on a Bayes-optimal statistical network. During the feedback learning process of PNN, the microwave characteristics of the no rain pixels with high scattering effect and the rain pixels with low scattering effect in the training dataset could be extracted, which should be helpful to identify the raining/no rain pixels with similar microwave characteristics of other areas. This assumption can be demonstrated by the results shown in Figure 5, which suggests that the PNN method can not only substantially reduce the amount of false alarm precipitation and misdetected precipitation in comparison with the SI method, but also show a much better performance than other widely used classification methods (TRMM 2A12 and Kmean). In general, the PNN model is a promising method to provide accurate rain area delineation based on satellite observations.

The accuracy of the rain and no rain areas in TRMM 2A12 over the Yangtze River Basin was also validated. As well-acknowledged and widely-used satellite precipitation datasets, the rainfall information in TRMM 2A12 was derived from TMI observations using a physical retrieval algorithm termed the Goddard profiling algorithm (GPROF) models. This model first built a dataset containing the potential hydrometeor profiles and their computed TBs at microwave channels. Then, the rainfall information was obtained over a location by matching the TBs of this location with the corresponding potential hydrometeor profiles using the Bayesian approach. However, Michaelides [18] pointed out a

potential drawback of the GPROF algorithm, which was that this algorithm was optimized for particular regions and datasets, leading to regional and seasonal errors contained in the TRMM 2A12. As shown in Figure 5, the TRMM 2A12 could only detect 57% rain pixels of the 16 rain events, while misclassifying 33% of no rain pixels as under raining conditions, and the ETS, HK, CSI and HSS of TRMM 2A12 were lower than the SI_threshold method and the PNN method. These results indicate that the TRMM 2A12 could not provide reasonable rain area delineation over the Yangtze River Basin. The misclassified precipitation will be inherited and exacerbated when this dataset is used to drive hydrological models and merged with other precipitation datasets [14,15]. Unfortunately, the false alarm and misdetected precipitation in TRMM 2A12 has rarely been noted and discussed so far. The PNN-based rain area delineation method proposed by this study provides a promising way to improve the performance of TRMM 2A12 over the regions with some rain gauge observations.

Figure 7 indicates that using the PNN method, the performance of the SI method can be improved in most cases. This reveals the importance of rain area delineation in the retrieval of high-resolution satellite precipitation. As addressed by Ferraro, Smith, Berg and Huffman [25], the misdetected and false alarm precipitation is the first-order errors of the satellite precipitation retrieval algorithms, which would be inherited and exacerbated in the retrieval of the precipitation volume. However, most studies only were concerned with improving the physical relationship between the satellite observations and surface rainfall volume or merging the satellite precipitation datasets with ground observations, whereas relatively less attention has been paid to investigating how to reduce the misdetected and false alarm precipitation in high-resolution satellite precipitation datasets. In our opinion, the issue about the rain area delineation with satellite observations deserves more attention in the future.

5. Conclusions

Misdetected and false alarm precipitation generated in the delineation of rain areas is the first-order error in the satellite precipitation retrieval algorithm, which can be inherited and exacerbated in the following process. In this study, we proposed a new rain area delineation method based on PNN using rain flags derived from rain gauges and multi-channel microwave brightness temperature observed by TMI. Independent validation indicated that the false alarm precipitation of PNN is 30% and 55% less than that generated by the TRMM 2A12 satellite precipitation dataset and Kmean method, while the misdetected precipitation of PNN is 19% and 26% less than them. Compared with the threshold method used in traditional SI algorithm, the PNN method can reduce 41% of the false alarm precipitation, and their misdetected precipitation is very similar. Furthermore, with this new rain area delineation method proposed by this study, the precipitation volumes retrieved by the SI algorithm could be improved with a 19% and 10% decrease of MAE and RMSE and a 29% increase of R^2. The results revealed the efficiency of the PNN in the delineation of the rain areas with microwave observations and the significance of reducing the misclassified rain/no rain areas in the retrieval of precipitation based on satellite observations. This research pointed out a new idea to reduce the errors in high-resolution microwave-based precipitation datasets, which has great significance for monitoring the precipitation over ungauged areas. Compared with the threshold method used in the SI algorithm, the PNN-based rain area delineation method could not significantly reduce the misdetected precipitation, which limited the performance of this new method. In the future, we will attempt to add more information (such as

topography, soil moisture and infrared brightness temperature) into the PNN to provide more accurate rain area delineation.

Acknowledgments

Author Contributions

The idea of this research is proposed by the first author, Shiguang Xu. The experiments of this research are conducted by Shiguang Xu and Chaoyang Wu. Alemu Gonsamo made some improvement for the algorithm, and edit the English of this manuscript. Yan Shen processed and provided the ground observations.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Goovaerts, P. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *J. Hydrol.* **2000**, *228*, 113–129.
2. Wu, C.; Chen, J.M. The use of precipitation intensity in estimating gross primary production in four northern grasslands. *J. Arid Environ.* **2012**, *82*, 11–18.
3. Jia, S.; Zhu, W.; Lŭ, A.; Yan, T. A statistical spatial downscaling algorithm of TRMM precipitation based on ndvi and dem in the qaidam basin of China. *Remote Sens. Environ.* **2011**, *115*, 3069–3079.
4. Xie, P.P.; Xiong, A.Y. A conceptual model for constructing high-resolution gauge-satellite merged precipitation analyses. *J. Geophys. Res. Atmos.* **2011**, *116*, doi:10.1029/2011JD016118.
5. Ricciardelli, E.; Cimini, D.; Di Paola, F.; Romano, F.; Viggiano, M. A statistical approach for rain intensity differentiation using meteosat second generation-spinning enhanced visible and infrared imager observations. *Hydrol. Earth Syst. Sci.* **2014**, *18*, 2559–2576.
6. Mugnai, A.; Smith, E.A.; Tripoli, G.J.; Bizzarri, B.; Casella, D.; Dietrich, S.; di Paola, F.; Panegrossi, G.; Sano, P. Cdrd and pnpr satellite passive microwave precipitation retrieval algorithms: Eurotrmm/eurainsat origins and h-saf operations. *Nat. Hazards Earth Syst.* **2013**, *13*, 887–912.
7. Di Paola, F.; Casella, D.; Dietrich, S.; Mugnai, A.; Ricciardelli, E.; Romano, F.; Sano, P. Combined mw-ir precipitation evolving technique (pet) of convective rain fields. *Nat. Hazards Earth Syst.* **2012**, *12*, 3557–3570.
8. Casella, D.; Dietrich, S.; Di Paola, F.; Formenton, M.; Mugnai, A.; Porcu, F.; Sano, P. PM-GCD—A combined IR-MW satellite technique for frequent retrieval of heavy precipitation. *Nat. Hazards Earth Syst.* **2012**, *12*, 231–240.

9. Casella, D.; Panegrossi, G.; Sano, P.; Dietrich, S.; Mugnai, A.; Smith, E.A.; Tripoli, G.J.; Formenton, M.; Di Paola, F.; Leung, W.Y.H.; *et al.* Transitioning from crd to cdrd in bayesian retrieval of rainfall from satellite passive microwave measurements: Part 2. Overcoming database profile selection ambiguity by consideration of meteorological control on microphysics. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 4650–4671.

10. Di Paola, F.; Dietrich, S. Resolution enhancement for microwave-based atmospheric sounding from geostationary orbits. *Radio Sci.* **2008**, *43*, doi:10.1029/2007RS003789.

11. Puca, S.; Porcu, F.; Rinollo, A.; Vulpiani, G.; Baguis, P.; Balabanova, S.; Campione, E.; Erturk, A.; Gabellani, S.; Iwanski, R.; *et al.* The validation service of the hydrological saf geostationary and polar satellite precipitation products. *Nat. Hazards Earth Syst.* **2014**, *14*, 871–889.

12. Cimini, D.; Romano, F.; Ricciardelli, E.; Di Paola, F.; Viggiano, M.; Marzano, F.S.; Colaiuda, V.; Picciotti, E.; Vulpiani, G.; Cuomo, V. Validation of satellite opemw precipitation product with ground-based weather radar and rain gauge networks. *Atmos. Meas. Tech.* **2013**, *6*, 3181–3196.

13. Tian, Y.D.; Peters-Lidard, C.D.; Eylander, J.B.; Joyce, R.J.; Huffman, G.J.; Adler, R.F.; Hsu, K.L.; Turk, F.J.; Garcia, M.; Zeng, J. Component analysis of errors in satellite-based precipitation estimates. *J. Geophys. Res. Atmos.* **2009**, *114*, doi:10.1029/2009JD011949.

14. Gebregiorgis, A.S.; Tian, Y.; Peters-Lidard, C.D.; Hossain, F. Tracing hydrologic model simulation error as a function of satellite rainfall estimation bias components and land use and land cover conditions. *Water Resour. Res.* **2012**, *48*, W11509.

15. Gebregiorgis, A.S.; Hossain, F. Understanding the dependence of satellite rainfall uncertainty on topography and climate for hydrologic model simulation. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 704–718.

16. Joyce, R.J.; Janowiak, J.E.; Arkin, P.A.; Xie, P.P. Cmorph: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *J. Hydrometeorol.* **2004**, *5*, 487–503.

17. Mahesh, C.; Prakash, S.; Sathiyamoorthy, V.; Gairola, R. Artificial neural network based microwave precipitation estimation using scattering index and polarization corrected temperature. *Atmos. Res.* **2011**, *102*, 358–364.

18. Michaelides, S.; Levizzani, V.; Anagnostou, E.; Bauer, P.; Kasparis, T.; Lane, J.E. Precipitation: Measurement, remote sensing, climatology and modeling. *Atmos. Res.* **2009**, *94*, 512–533.

19. Grody, N.C. Precipitation monitoring over land from satellites by microwave radiometry. In Proceedings of the 1984 International Geoscience and Remote Sensing Symposium, Strasbourg, France, 27–30 August 1984; pp. 417–423.

20. Grody, N.C. Classification of snow cover and precipitation using the special sensor microwave imager. *J. Geophys. Res.Atmos.* 1991, 96, 7423–7435.

21. Ferraro, R.R. Special sensor microwave imager derived global rainfall estimates for climatological applications. *J. Geophys. Res.* **1997**, *102*, doi:10.1029/97JD01210.

22. Yao, Z.; Li, W.; Zhu, Y.; Zhao, B.; Chen, Y. Remote sensing of precipitation on the tibetan plateau using the trmm microwave imager. *J. Appl. Meteorol.* **2001**, *40*, 1381–1392.

23. Neale, C.M.U.; Mcfarland, M.J.; Chang, K. Land-surface-type classification using microwave brightness temperatures from the special sensor microwave imager. *IEEE Trans. Geosci. Remote Sens.* **1990**, *28*, 829–838.

24. Petty, G.W. Physical retrievals of over-ocean rain rate from multichannel microwave imagery. 1. Theoretical characteristics of normalized polarization and scattering indexes. *Meteorol. Atmos. Phys.* **1994**, *54*, 79–99.

25. Ferraro, R.R.; Smith, E.A.; Berg, W.; Huffman, G.J. A screening methodology for passive microwave precipitation retrieval algorithms. *J. Atmos. Sci.* **1998**, *55*, 1583–1600.

26. Sidek, O.; Quadri, S.A. A review of data fusion models and systems. *Int. J. Image Data Fusion* **2012**, *3*, 3–21.

27. Hsu, K.L.; Gao, X.G.; Sorooshian, S.; Gupta, H.V. Precipitation estimation from remotely sensed information using artificial neural networks. *J. Appl. Meteorol.* **1997**, *36*, 1176–1190.

28. Yang, X.; Lu, X.X. Estimate of cumulative sediment trapping by multiple reservoirs in large river basins: An example of the Yangtze river basin. *Geomorphology* **2014**, *227*, 49–59.

29. Chen, X.; Zong, Y.; Zhang, E.; Xu, J.; Li, S. Human impacts on the Changjiang (Yangtze) River Basin, China, with special reference to the impacts on the dry season water discharges into the sea. *Geomorphology* **2001**, *41*, 111–123.

30. Specht, D.F. Probabilistic neural networks. *Neural Netw.* **1990**, *3*, 109–118.

31. Akiwowo, A.; Eftekhari, M. Feature-based detection using bayesian data fusion. *Int. J. Image Data Fusion* **2013**, *4*, 308–323.

32. Wu, S.G.; Bao, F.S.; Xu, E.Y.; Wang, Y.-X.; Chang, Y.-F.; Xiang, Q.-L. A leaf recognition algorithm for plant classification using probabilistic neural network. In Proceedings of the 2007 IEEE International Symposium on Signal Processing and Information Technology, Cairo, Egypt, 15–18 December 2007; IEEE: New York, NY, USA, 2007; pp. 11–16.

33. Kidd, C. On rainfall retrieval using polarization-corrected temperatures. *Int. J. Remote Sens.* **1998**, *19*, 981–996.

34. Spencer, R.W.; Goodman, H.M.; Hood, R.E. Precipitation retrieval over land and ocean with the SSM/I: Identification and characteristics of the scattering signal. *J. Atmos. Ocean. Technol.* **1989**, *6*, 254–273.

35. Mashingia, F.; Mtalo, F.; Bruen, M. Validation of remotely sensed rainfall over major climatic regions in northeast Tanzania. *Phys. Chem. Earth Parts A/B/C* **2014**, *67–69*, 55–63.

36. Haile, A.T.; Rientjes, T.; Gieske, A.; Gebremichael, M. Multispectral remote sensing for rainfall detection and estimation at the source of the Blue Nile River. *Int. J. Appl. Earth Obs. Geoinf.* **2010**, *12*, S76–S82.