entropy

MDPI

Article

# Modeling Multi-Event Non-Point Source Pollution in a Data-Scarce Catchment Using ANN and Entropy Analysis

**Lei Chen, Cheng Sun, Guobo Wang, Hui Xie and Zhenyao Shen ***

State Key Laboratory of Water Environment Simulation, School of Environment, Beijing Normal University, Beijing 100875, China; chenlei1982bnu@bnu.edu.cn (L.C.); 13756892980@163.com (C.S.); wangguobo@yeah.net (G.W.); imbahui@163.com (H.X.)

* Correspondence: zyshen@bnu.edu.cn; Tel.: +86-10-5880-0398

Received: 5 May 2017; Accepted: 7 June 2017; Published: 19 June 2017

**Abstract:** Event-based runoff–pollutant relationships have been the key for water quality management, but the scarcity of measured data results in poor model performance, especially for multiple rainfall events. In this study, a new framework was proposed for event-based non-point source (NPS) prediction and evaluation. The artificial neural network (ANN) was used to extend the runoff–pollutant relationship from complete data events to other data-scarce events. The interpolation method was then used to solve the problem of tail deviation in the simulated pollutographs. In addition, the entropy method was utilized to train the ANN for comprehensive evaluations. A case study was performed in the Three Gorges Reservoir Region, China. Results showed that the ANN performed well in the NPS simulation, especially for light rainfall events, and the phosphorus predictions were always more accurate than the nitrogen predictions under scarce data conditions. In addition, peak pollutant data scarcity had a significant impact on the model performance. Furthermore, these traditional indicators would lead to certain information loss during the model evaluation, but the entropy weighting method could provide a more accurate model evaluation. These results would be valuable for monitoring schemes and the quantitation of event-based NPS pollution, especially in data-poor catchments.

**Keywords:** non-point source pollution; ANN; entropy weighting method; data-scarce; multi-events

## 1. Introduction

Non-point source (NPS) pollution has resulted in the deterioration of water bodies and has become a major environmental threat among most counties [1,2]. The quantification of the rainfall–runoff process and the resulting NPS pollutants is essential for developing mitigation strategies, which are the basis for watershed management [3]. The rainfall process is the major driving force for NPS, thus rainfall-runoff-pollutant (R-R-P) relationships have become the focus of watershed research [4,5]. Many studies have been conducted in the fields of rainfall–runoff relationships but have rarely involved the runoff–pollutant relationship, especially for the event-based estimation of NPS loads [6–8].

The NPS processes can be expressed from the event-step to long-term steps. Event-based NPS exports and the resulting change in water quality can provide detailed features of the NPS, which is more appropriate for the design of storm-based management practices [9]. Models are developed to construct the runoff–pollutant relationship, and the discrepancies of the collected measured data in different rainfall patterns would have a considerable influence on the model construction. Identifying the correlation among the series of rainfall, runoff and pollutant loads for multiple rainfall events is inevitable for NPS model construction. Although many models are well suited for offline water quality analyses, Soil and Water Assessment Tool (SWAT) is more representative than any other models [10].

However, owing to limited human resources, data scarcity has become one of the key barriers to establish the R-R-P relationship, especially for event-based process [11,12]. Thus, the application of watershed models such as SWAT for assessing NPS pollution is also limited by temporal resolution which ranges from annual to sub-hourly averages. The SWAT model usually operates continuously at a daily time step, which ensures that the long-term impacts of NPS can be quantified. Sub-daily calculations of runoff, erosion, and sediment transport are also available in new version of SWAT by sub-daily rainfall input and Green and Ampt method, though few attempts have reached to that higher temporal resolution. In the future, we would develop other more appropriate models to solve this problem. Currently, acceptable rainfall and streamflow data sets are more readily available, especially because of the recent development of data centers and satellite data observations. However, hourly or sub-hourly flow data for high-frequency time series are still limited, especially with respect to event-based hydrological studies for data-poor regions [7]. Water quality records, which are based on periodic monitoring by human resources, are thus even scarcer. Therefore, data scarcity for NPS predictions is unavoidable for multiple-rainfall event simulations. Typically, we collected samples during multiple rainfall events in the monitoring process but discarded some of events from further analysis, especially for light rainfall, for which only a few data points exist. This treatment of incomplete data would result in the loss of information, especially for multiple rainfall events among data-scarce regions.

Currently, statistical models have been widely used to estimate rainfall–runoff relationships for its ease of application without considering a large amount of delicate formulas and parameters [13]. For example, unit hydrographs (UH), as one of the most famous methods, is used to estimate a direct runoff hydrograph of a given rainfall duration. Meanwhile, statistical models are used to simulate pollutant loads based on the established runoff–pollutant relationship. For example, Park and Engel [14] developed Load Estimator (LOADEST) to predict pollutant concentration (or load) on days when flow data were measured, and the results showed that absolute values of errors in the annual sediment load estimation decreased from 39.7% to 10.8%. Meanwhile, most of the findings demonstrate that the LOADEST model could provide more accurate results and may be useful for simulating runoff–pollutant processes [15–17]. However, the LOADEST model has strict requirements on the number of data points, which should include continuous flow data and dispersed water quality data, and its calibration process is relatively complicated.

Owing to the limited measured data, the black-box model might be a substitution to construct the logical relationships between runoff and pollutant loads for multiple-events processes. The artificial neural network (ANN) with the characteristics of self-learning and adaptability has become the most commonly used tool in environmental prediction, and it is also available for poor-data regions. This method is applicable to simulate the imaginal thinking of the human brain, for which the most prominent characteristic is the parallel processing of information and distributed storage. As an example, Melesse et al. [18] used the ANNto estimate suspended sediment loads for three major rivers. The results showed that daily predictions were better than weekly predictions. Therefore, it can be seen that ANN models have flexible structures that allow multi-input and multi-output modeling. This is particularly important in streamflow forecasting where inflows at multiple locations are considered within a given catchment [19]. Though the application of ANN in the field of load production has proliferated in recent years, the impact of data scarcity on its prediction capabilities during different rainfall patterns still creates limitations [20].

Simulation evaluation is the most important step for the setup of statistical models [21]. In traditional applications, the model evaluation is usually performed using a single regression goodness-of-fit indicator, the most common of which is the point-to-point pairs (a series of single data pairs) of the predicted and measured data. However, this might lead to the loss of specific information, resulting in dubitable simulation results. In this case, a joint evaluation should be a substitute for the traditional single indicator. With the high precision and objectivity, the entropy regulates the uncertainty of different criteria from different perspectives [22]. Compared with the

traditional single indicators, it can combine different indicators to evaluate the discrepancy between causes comprehensively. For instance, Khosravi et al. [23] sought to map the flooding susceptibility using different bivariate methods, including Shannon's entropy, the statistical index and the weighting factor. Yuan et al. [24] developed an entropy method to find the weight sum of the information entropy maximum to allocate the reduction of pollutants for the main seven valleys in China. The entropy weighting method may be an efficient way to evaluate the regulation of the simulation results and to balance the strengths and weaknesses of the results. However, these studies do not provide much attention to event-based NPS predictions, especially for data-scarce catchments.

This study surveys the motivation for a methodology of action, looks at the difficulties posed by data scarcity and outlines the need for the development of possibility methods to cope with data scarcity in multiple rainfall events. The objectives of this work are: (1) to identify the impacts of different rainfall patterns on the model construction using a complete data series; (2) to simulate the scarce pollutant data in other data-scarce rainfall events; and (3) to test the application of the entropy weighting method for the evaluation of ANN.

## 2. Materials and Methods

A prediction-evaluation framework is proposed for the NPS prediction for data-scarce catchments, the flow chart of methods is shown in Figure 1. The ANN is proposed to simulate the missing data points during multiple-events, and the entropy weighting method is used as a comprehensive indicator to construct the model. As a necessary supplement, the interpolation method is used for tail correction during multiple rainfall events. Data-scarce rainfall events denote the absence of data, especially for measured flow and water quality, in a given period of time due to human mistakes during high-resolution monitoring process. Instead, complete rainfall events are defined if there are no measured flow and water quality data scarcity. The demonstration of traditional indicators is shown in Section 2.2.

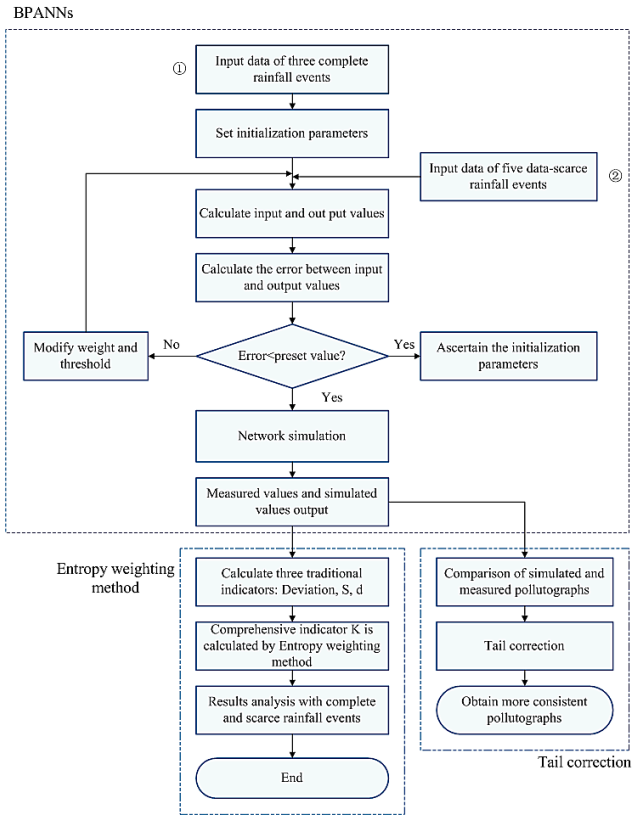



**Figure 1.** The methods presented in a flow chart.

### *2.1. The Description of the ANN*

The back propagation algorithm is a supervised learning method based on the commonly used steepest descent method to minimize global errors [25], while it is also the multilayer feedforward network based on the error back propagation algorithm [2,26]. It accumulates an abundant mapping relation of the input-output pattern and does not need to reveal mathematical equations to describe the mapping relation before calculation. The ANN may be an efficient method to adjust the weights and thresholds through back propagation to minimize the sum of the squared errors. As shown in Figure 2, the topological structure of the ANN consists of an input layer, a hidden layer and an output layer [27].

The learning mechanism of the ANN is shown in Figure 2, where $x_i$ is the input signal and $w_i$ is the weight coefficient. The outside input samples $x_1, x_2, \ldots, x_n$ are accepted into the input layer, and the network weight coefficients are adjusted during training. The discrete values, 0 and 1, are selected as the input sampling signals. By comparing the network output signals and the expected output signals to generate the error signals, the weight coefficients of the learning system can be rectified based through iterative adjustments to minimize the errors until reaching an acceptable range [28]. In this process, the expected output signals are regarded as the teacher signals, which are compared with the actual output, and the errors produced are applied to rectify the weight coefficients. At the point when the actual output values and expected values are nearly the same, the process is concluded [26]. Finally, the results are produced through and equation of $U$ based on the weight coefficients and are exported by the output layers. In the ANN training process, three prime criteria can be summarized: the error surface gradient can converge rapidly, the mean squared error is below the error of the preset level, and the correlation coefficient of the training results is more than 0.9, indicating that training results are an improvement [29]. This section briefly surveys the measurement for methodology, while the ANN should be judged for whether each indicator can or cannot reach the given standards.

In this study, multiple rainfall events are used as the input conditions. Multiple rainfall events are divided into either the training process or the simulation process based on the data conditions. To establish the black-box model, data of three complete rainfall events are first input into the layer, including light, moderate, and heavy rainfall patterns. The training results also indicate that the ANN is applicable for various rainfall patterns. In the simulation process, the flow data for all the rainfall and water quality information for the data-complete rainfall events for the same rainfall pattern are regarded as the input layer. The hidden layer contains the water quality data for the data-scarce rainfall events which correspond to all the flow and water quality data in the input layer. To obtain the output layer, the training layer feedbacks the results into the prediction interval. Finally, the output layer is simulated using the input data of the input layer.

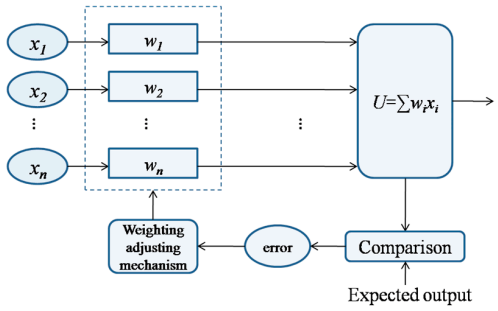



**Figure 2.** The learning mechanism of the artificial neural network (ANN).

### *2.2. The Description of the Entropy Weighting Method*

Three commonly used indicators, the mean relative error ($\bar{d}$), the standard deviation of the relative error ($S$), and the load deviation percentage (*deviation*), are selected to evaluate the simulation results [30,31]. The formulas are shown as followed:

$$deviation = \frac{O_i^{origin} - O_i}{O_i^{origin}} \times 100\% \quad (1)$$

$$\overline{d} = \sum_{i=1}^{n} (O_i - P_i)/n \quad (2)$$

$$S = \sqrt{\frac{\sum_{i=1}^{n} \left(d_i - \overline{d}\right)^2}{n-1}} \quad (3)$$

where $O_i$ is the set of measured data, $P_i$ is the set of predicted data, and $O_i{}^{origin}$ denotes the total loads of the original conditions, and is the mean value of the measured data.

Each of the three indicators represents the credibility of the measurements based on the discrepancy between the measured and simulated values. Lower indicator values indicate that the fitting between the simulated and measured data is improved, and the model is considered to have a satisfactory performance. However, single indicators have limitation on amount of information loss. Therefore, these indicators are handled with the entropy weighting method for a more comprehensive assessment of the ANN. Based on the fundamental principles of information theory, information is a measurement of the degree of order for a given system, and the entropy is a measurement of the degree of disorder [32]. The entropy weighting method serves as a mathematic method and considers the information provided by each factor [33]. Information entropy is negatively associated with the increase in information provided by different indicators, and a smaller information entropy result in higher weights for each single indicator. As an objective and comprehensive method, the entropy weighting method considers the advantages of every indicator and makes a synthetic evaluation. This principle is as follows:

Firstly, an $n \times m$ origin data matrix is established according to the selected evaluation indicators:

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{bmatrix}_{n \times m} \quad (4)$$

where $m$ denotes the evaluation indicator, and individual rows represent different evaluation objects. Therefore, matrix $X$ is known.

A second, positive matrix should be established with a transformation following same trend. The transformed matrix is

$$Y = \begin{bmatrix} y_{11} & \cdots & y_{1n} \\ \vdots & \ddots & \vdots \\ y_{n1} & \cdots & y_{nm} \end{bmatrix}_{n \times m} \quad (5)$$

Matrix $Y$ is normalized, and the ratio of each column vector $y_{ij}$ and the sum of all elements in this matrix should be normalized. The formulas for these calculations are:

$$Z_{ij} = \frac{y_{ij}}{\sum_{i=1}^{n} Y_{ij}} (j = 1,\ 2,\ \ldots,\ m) \quad (6)$$

where $Z_{ij}$ are the elements of the normalized matrix.

The operational formula in the process of generating the entropy weights of the evaluation indicators is

$$H(x_j) = -k \sum_{i=1}^{n} z_{ij} \ln z_{ij} (j = 1,\ 2, \ldots,\ m) \quad (7)$$

where $k$ is a normalizing constant, $k = 1/\ln n$, and $Z_{ij}$ is the $j$-th the probability of the element of the $i$-th evaluation unit. Entropy values of the evaluation indicators should be transformed into the weighted values:

$$w_j = \frac{1 - H(x_j)}{m - \sum_{j=1}^{m} H(x_j)} \quad j = 1, 2, \ldots, m, \tag{8}$$

where $0 \leq w_i \leq 1$ and $\sum_{j=1}^{m} w_j = 1$ are the acquired weighted values. Finally, the comprehensive weighting values for each evaluation indicator should be ensured. The weighted values of each indicator are multiplied with the corresponding indicators and summed. The evaluation model is

$$U = \sum_{j=1}^{m} w_i z_{ij} \ (j = 1, 2, \ldots, n) \tag{9}$$

where $U$ represents the comprehensive evaluation function of the entropy weights for each evaluation indicator. This function reflects the comprehensive characteristics of the evaluation objective, which avoids limiting these indicators [34].

The principle of the entropy weighting method is that information for each evaluation unit will be qualified and synthesized, while every factor is weighted to simplify the evaluation process [35]. Therefore, the weight values can be ascertained with the entropy weighting method, and we choose the *deviation*, $\overline{d}$, and $S$ as the evaluation indicators.

### *2.3. Method for Tail Correction*

Statistical models would result in tail deviation problems if data scarcity exists in this study. This problem addressed through data interpolation for the tail deviation. Therefore, linear interpolation, as a common-used method, is used to obtain the missing values of the other data points. Two values of the function $f(x)$ are used to reduce the errors in the tail of the pollutographs. This approach is relatively straightforward and is used widely in the field of mathematics or computer graphics. The error of the approximate method can be defined as follows:

$$R_T = f(x) - \rho(x) \tag{10}$$

where $\rho$ represents the linear interpolated polynomial:

$$\rho(x) = f(x_0) + \frac{f(x_1) - f(x_0)}{x_1 - x_0}(x - x_0) \tag{11}$$

As a result of Rolle's theorem, if $f(x)$ has two continuous derivatives, the error range is

$$|R_T| \leq \frac{(x_1 - x_0)^2}{8} \max_{x_0 \leq x \leq x_1} |f'(x)| \tag{12}$$

As shown in Formula (12), the approximate error of the linear interpolation increases with the function curvature.

## 3. Case Study

### *3.1. Study Areas*

As shown in Figure 3, the Zhangjiachong catchment, which is a representative area in the Three Gorges Reservoir Region (TGRR), is selected as a case study [36]. It covers a drainage area of 1.62 km$^2$, and the landscape is primarily mountainous, with an elevation between 148 m and 530 m above the Yellow Sea level. Agriculture and forests cover the majority of the total area. The main local crops are tea, corn, oil seed rape, and chestnuts [37]. The background values of nitrogen and phosphorus are

higher because the fertilizer usage is relatively high, resulting in a high risk of nutrient loss into nearby streams [38].

The average annual temperature is approximately 18 °C, and the average annual precipitation is approximately 1439 m, 80% of which occurs from May to August. Thus, soil erosion frequently occurs during wet seasons, and results in an increase in the pollutant loads with increased runoff. We consider that the variation of rainfall might impact the model accuracy. Therefore, identifying the classification of rainfall patterns should be determined before any simulations. According to the investigation results of existing rainfall data, rainfall patterns are divided into light, moderate, and heavy events. Meanwhile, based on our monitoring data, a majority of rainfall events in the Zhangjiachong catchment are considered moderate events, while heavy events are rare.

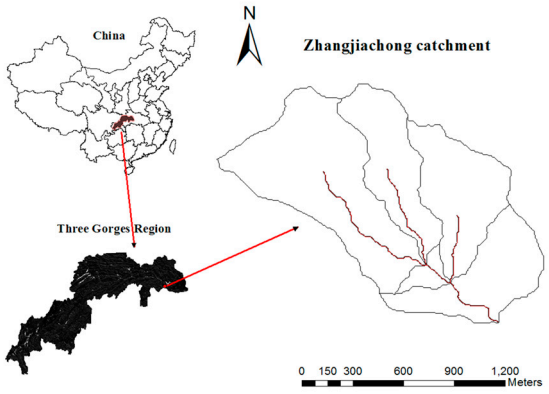



**Figure 3.** The location of the Zhangjiachong catchment.

### *3.2. Field Monitoring and Data Record*

In this study, field monitoring data were collected from 1 January 2013 to 31 December 2014 and the rainfall, streamflow and pollutant data during eight rainfall events were recorded. The data used in this study represent three complete rainfall events, which include light, middle, and heavy rainfall (21 April 2014, 24 July 2014, and 5 August 2014), and five other data-scarce events (15 April 2014, 23 August 2014, 20 July 2014, 5 July 2013, and 28 August 2013). Data-scarce rainfall events denote the absence of data, especially for measured flow and water quality, in a given period of time due to human mistakes during high-resolution monitoring process. Instead, complete rainfall events are defined if there are no measured flow and water quality data scarcity. The equations with explicit parameters are constructed through a training process with complete data of the three complete rainfall events, and the constructed ANN is used to predict the missing NPS data in the other five data-scarce rainfall events. The output layer includes pollutant load data for five data-scarce rainfall events.

The weather station (Skye Lynx Standard) provided continuous records for climate data and a float-operator sensor (WGZ-1) was located at the catchment outlet, where high-frequency sampling was recorded in approximately 15 min steps. Base flows were measured before the runoff started, and water samples were collected every 15 min in the first hour after runoff began and every 30 min over the following two hours. After water levels had stabilized, water samples were collected once every hour until the end of the event. All water samples were placed in pre-cleaned glass jars with aluminumfoil liners along the lids and stored at −20 °C during transportation to the laboratory for processing and analysis. Specifically, the total nitrogen of NPS (NPS-TN) levels were measured via Alkaline persulfate oxidation-UV spectrophoto metric with the detection limitation from 0.05 mg/L to 4.0 mg/L, while the total phosphorus of NPS (NPS-TP) levels in the samples were measured via Potassium persulfate oxidation-molybdenum blue colorimetric methods. The main instrument is ultraviolet spectrophotometer. Finally, the recorded rainfall, flow and pollutant levels were used for the following analysis.

**Table 1.** Complete data for the three rainfall events.

| 21 April 2014 | | | | 24 July 2014 | | | | 5 August 2014 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Time** | **Flow ($m^3/s$)** | **NPS-TN (mg/L)** | **NPS-TP (mg/L)** | **Time** | **Flow ($m^3/s$)** | **NPS-TN (mg/L)** | **NPS-TP (mg/L)** | **Time** | **Flow ($m^3/s$)** | **NPS-TN (mg/L)** | **NPS-TP (mg/L)** |
| 2:45 | 0.002 | 4.75 | 0.063 | 22:45 | 0.003 | 0.84 | 0.11 | 21:30 | 0.008 | 2.66 | 0.30 |
| 3:00 | 0.009 | 8.89 | 0.193 | 23:00 | 0.380 | 6.26 | 0.76 | 21:45 | 0.678 | 7.29 | 0.84 |
| 3:15 | 0.015 | 15.29 | 0.300 | 23:15 | 0.647 | 6.96 | 1.04 | 22:00 | 1.107 | 11.30 | 1.12 |
| 3:30 | 0.016 | 13.31 | 0.301 | 23:30 | 0.726 | 7.06 | 0.87 | 22:15 | 1.400 | 11.90 | 1.26 |
| 3:45 | 0.031 | 25.28 | 0.369 | 23:45 | 0.336 | 6.77 | 0.92 | 22:30 | 2.227 | 9.85 | 0.94 |
| 4:00 | 0.037 | 14.14 | 0.297 | 0:00 | 0.971 | 8.77 | 1.12 | 23:00 | 1.647 | 15.00 | 1.20 |
| 4:30 | 0.065 | 22.64 | 0.652 | 0:30 | 0.570 | 8.08 | 0.70 | 23:30 | 0.585 | 9.67 | 0.62 |
| 5:00 | 0.071 | 24.48 | 0.441 | 1:00 | 0.294 | 5.53 | 0.48 | 0:00 | 0.945 | 9.56 | 0.68 |
| 6:00 | 0.086 | 26.49 | 0.469 | 2:00 | 0.266 | 5.05 | 0.57 | 1:00 | 0.410 | 7.01 | 0.35 |
| 7:00 | 0.126 | 23.89 | 0.443 | 3:00 | 0.172 | 5.11 | 0.36 | 2:00 | 0.237 | 2.03 | 0.28 |
| 8:00 | 0.146 | 16.97 | 0.286 | 4:00 | 0.191 | 5.79 | 0.34 | 3:00 | 0.183 | 7.60 | 0.37 |
| 9:00 | 0.264 | 11.60 | 0.171 | 5:00 | 0.041 | 5.85 | 0.20 | 4:00 | 0.166 | 6.81 | 0.27 |
| 10:00 | 0.278 | 11.00 | 0.117 | 6:00 | 0.090 | 5.59 | 0.18 | 5:00 | 0.115 | 7.01 | 0.20 |
| 11:00 | 0.288 | 10.73 | 0.104 | 7:00 | 0.064 | 7.78 | 0.15 | 6:00 | 0.126 | 6.71 | 0.19 |
| 12:00 | 0.296 | 10.94 | 0.109 | 8:00 | 0.048 | 5.29 | 0.14 | 7:00 | 0.102 | 6.08 | 0.12 |
| 13:00 | 0.411 | 9.64 | 0.113 | | | | | 8:00 | 0.073 | 7.03 | 0.24 |
| 14:00 | 0.593 | 9.48 | 0.089 | | | | | 9:00 | 0.063 | 5.68 | 0.11 |
| 15:00 | 0.593 | 9.65 | 0.087 | | | | | 10:00 | 0.086 | 5.75 | 0.09 |
| 16:00 | 0.602 | 9.26 | 0.072 | | | | | 11:00 | 0.075 | 6.44 | 0.18 |
| 17:00 | 0.770 | 9.93 | 0.068 | | | | | 12:00 | 0.045 | 6.33 | 0.21 |
| | | | | | | | | 13:00 | 0.029 | 5.88 | 0.14 |

**Table 2.** Five rainfall events with data scarcity.

| 15 April 2014 (1.213 mm/h) | | | | 28 August 2013 (2.027 mm/h) | | | | 20 July 2014 (2.013 mm/h) | | | | 5 July 2013 (2.380 mm/h) | | | | 28 August 2013 (2.647 mm/h) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Time** | **Flow** | **TN** | **TP** | **Time** | **Flow** | **TN** | **TP** | **Time** | **Flow** | **TN** | **TP** | **Time** | **Flow** | **TN** | **TP** | **Time** | **Flow** | **TN** | **TP** |
| 6:00 | 0.0127 | 7.65 | 0.050 | 15:00 | 0.0375 | 1.13 | 0.13 | 19:00 | 0.0127 | 5.51 | 0.27 | 14:30 | 0.0127 | 5.03 | 0.06 | 19:00 | 0.03 | 5.59 | 0.05 |
| 6:30 | 0.024 | 12.27 | 0.412 | 15:30 | - | - | - | 19:30 | 0.0163 | 7.02 | 0.56 | 15:00 | - | - | - | 19:30 | 0.06 | 12.65 | 0.33 |
| 6:45 | 0.0375 | 11.49 | 0.366 | 16:00 | 0.0964 | 6.29 | 0.71 | 20:00 | 0.0485 | 6.92 | 0.86 | 15:30 | - | - | - | 20:00 | 1.78 | 15.24 | 1.39 |
| 7:00 | 0.0427 | 23.53 | 0.579 | 16:30 | 0.1020 | 5.57 | 0.3 | 20:30 | 0.1190 | 5.56 | 1.15 | 16:00 | 0.0127 | 8.81 | 1.50 | 20:30 | 2.22 | 16.38 | 0.84 |
| 7:15 | 0.0401 | 16.13 | 0.363 | 17:00 | 0.0964 | 5.81 | 0.32 | 21:00 | 0.0866 | 5.14 | 0.53 | 16:30 | 0.0127 | 6.37 | 2.56 | 21:00 | 1.94 | 13.88 | 0.69 |
| 7:30 | 0.0406 | 15.19 | 0.350 | 17:30 | 0.0547 | 5.2 | 0.2 | 21:30 | 0.0327 | 7.37 | 0.4 | 17:00 | 0.0375 | 1.06 | 2.11 | 21:30 | 0.80 | 14.95 | 0.44 |
| 8:00 | 0.0327 | 16.92 | 0.373 | 18:00 | 0.0427 | 4.95 | 0.14 | 22:00 | - | - | - | 17:30 | - | - | - | 22:00 | 0.39 | 13.43 | 0.55 |
| 8:30 | 0.0351 | 14.29 | 0.363 | 18:30 | - | - | - | 22:30 | 0.0375 | 5.64 | 0.24 | 18:00 | 0.4140 | - | - | 22:30 | 0.26 | 14.81 | 0.56 |
| 9:00 | 0.0327 | 11.26 | 0.291 | 19:00 | 0.0375 | 5.14 | 0.13 | 23:00 | - | - | | 18:30 | 0.4440 | - | - | 23:00 | - | - | - |
| 10:00 | - | - | - | 19:30 | - | - | - | 23:30 | 0.0182 | 6.49 | 0.19 | 19:00 | 0.2220 | 3.20 | 0.41 | 23:30 | - | - | - |
| 10:30 | 0.0351 | 8.67 | 0.228 | 20:00 | 0.0427 | 5.21 | 0.13 | | | | | 19:30 | - | - | - | 0:00 | - | - | - |
| 11:00 | 0.0127 | 7.65 | 0.050 | | | | | | | | | 20:00 | 0.1590 | 4.08 | 0.28 | 0:30 | - | - | - |
| | | | | | | | | | | | | 20:30 | - | - | - | 1:00 | 0.21 | 10.17 | 0.14 |
| | | | | | | | | | | | | 21:00 | 0.0964 | 4.26 | 0.11 | 1:30 | - | - | - |
| | | | | | | | | | | | | 21:30 | - | - | - | 2:00 | 0.15 | 9.83 | 0.19 |
| | | | | | | | | | | | | 22:00 | 0.0775 | 3.78 | 0.22 | 2:30 | - | - | - |
| | | | | | | | | | | | | 22:30 | - | - | - | 3:00 | 0.09 | 6.54 | 0.08 |
| | | | | | | | | | | | | 23:00 | 0.0616 | 3.80 | 0.13 | | | | |
| | | | | | | | | | | | | 23:30 | - | - | - | | | | |
| | | | | | | | | | | | | 0:30 | 0.0547 | - | - | | | | |
| | | | | | | | | | | | | 1:30 | 0.0427 | - | - | | | | |

Note: the units of flow are in $m^3/s$; the units of NPS-TN and NPS-TP are mg/L; - denotes that data are missing at this time.

However, flow and water quality data were limited because of the use of flow instruments via manual collection. Rainfall levels were recorded to divide the rainfall into light, moderate, and heavy events. The rainfall levels for 21 April 2014, 24 July 2014, and 5 August 2014 are 1.308 mm/h, 3.000 mm/h, and 6.054 mm/h, respectively. The flow data were replenished with unit hydrographs as the basis for the ANN. In addition, this catchment is dominated by agriculture, so fertilizer use results in deteriorated water quality. Therefore, the NPS-TN and NPS-TP are selected as the evaluation indicators. All the data for the three complete rainfall events and five typical data-scarce rainfall events are shown in Tables 1 and 2, respectively, including the rainfall intensity, flow data, and the pollutant concentration of the NPS-TN and NPS-TP. As shown in Table 1, complete data are used as the input of the ANN and represent the impacts of the rainfall patterns on the model applicability. As shown in Table 2, data scarcity of the five random rainfall events is simulated, and the impacts of the data scarcity are quantified.

## 4. Results and Discussion

### *4.1. Training Results of the ANNUsing the Complete Data*

This section demonstrates the training process with data for the three complete rainfall events, illustrating that the applicability of ANN in different rainfall patterns. The training results for the ANN areas followed (the figure is shown in the Supplementary Materials): the error surface gradient rapidly converges to a flat surface for both the NPS-TN and NPS-TP. The mean squared error of the training results for the NPS-TP prediction reaches the $10^{-3}$, $10^{-2}$, and $10^{-1}$ orders of magnitude for the light, moderate, and heavy rainfall events, respectively. However, the mean squared error for the NPS-TN prediction reaches the 1.0, $10^{-3}$, and 1.0 orders of magnitude during the light, moderate, and heavy rainfall events, respectively, indicating that all the results fall within the range of permissible errors or rapidly reach a flat surface. The correlation coefficients are more than 0.9, indicating that all the training results are good. In this respect, it can be said that the ANN is applicability to simulate the NPS for different rainfall patterns, and we extrapolated ANN for pollutant load simulations in the data-scarce rainfall events.

To better understand the simulation results, the entropy weighting method was used in the evaluation process. As shown in Table 3, K results are all higher than 0.9, indicating that there is no obvious deviation between the simulated and measured values. Meanwhile, the K values for the NPS-TP are higher than the NPS-TN for different rainfall patterns, and the K value for the light rainfall is higher than the other rainfall patterns. Therefore, it is apparent that the NPS-TP simulation is an improvement over the NPS-TN, and the simulation is better suited for the light rainfall events for both the NPS-TP and NPS-TN. It is obvious that the flow have different shear force in different rainfall patterns. The soil particles and pollutants act differently with different rainfall levels and intensities. It is possible that our monitoring scheme is more appropriate in light rainfall patterns in this experiment, and the peak data cannot be monitored during heavy rainfall patterns [39]. The NPS-TN concentration peak and flow peak appear to be consistent. When one of the flow or load peaks is missing, it is the same as both of them missing simultaneously, resulting in a poor simulation effect. However, the apparent time of the NPS-TP concentration peak and flow peak is inconsistent in different rainfall patterns. Xu et al. [40] introduced the support vector regression (SVR) model to develop a quantitative relationship between the environmental factors and the eutrophic indices compared with the ANN. The results show that the correlation coefficients of the NPS-TP are greater than those for the NPS-TN, indicating that the model effect of the NPS-TP is improved over the NPS-TN. This study verifies this conclusion with the ANN model.

**Table 3.** Evaluation of the simulation results of the pollutant loads for different rainfall patterns.

| Rainfall Events | Comprehensive Indicators K | |
|---|---|---|
| | NPS-TP | NPS-TN |
| 21 April 2014 | 0.986 | 0.953 |
| 24 July 2014 | 0.973 | 0.938 |
| 5 August 2014 | 0.958 | 0.921 |

### *4.2. Simulated Results of the ANN for Data-Scarce Events*

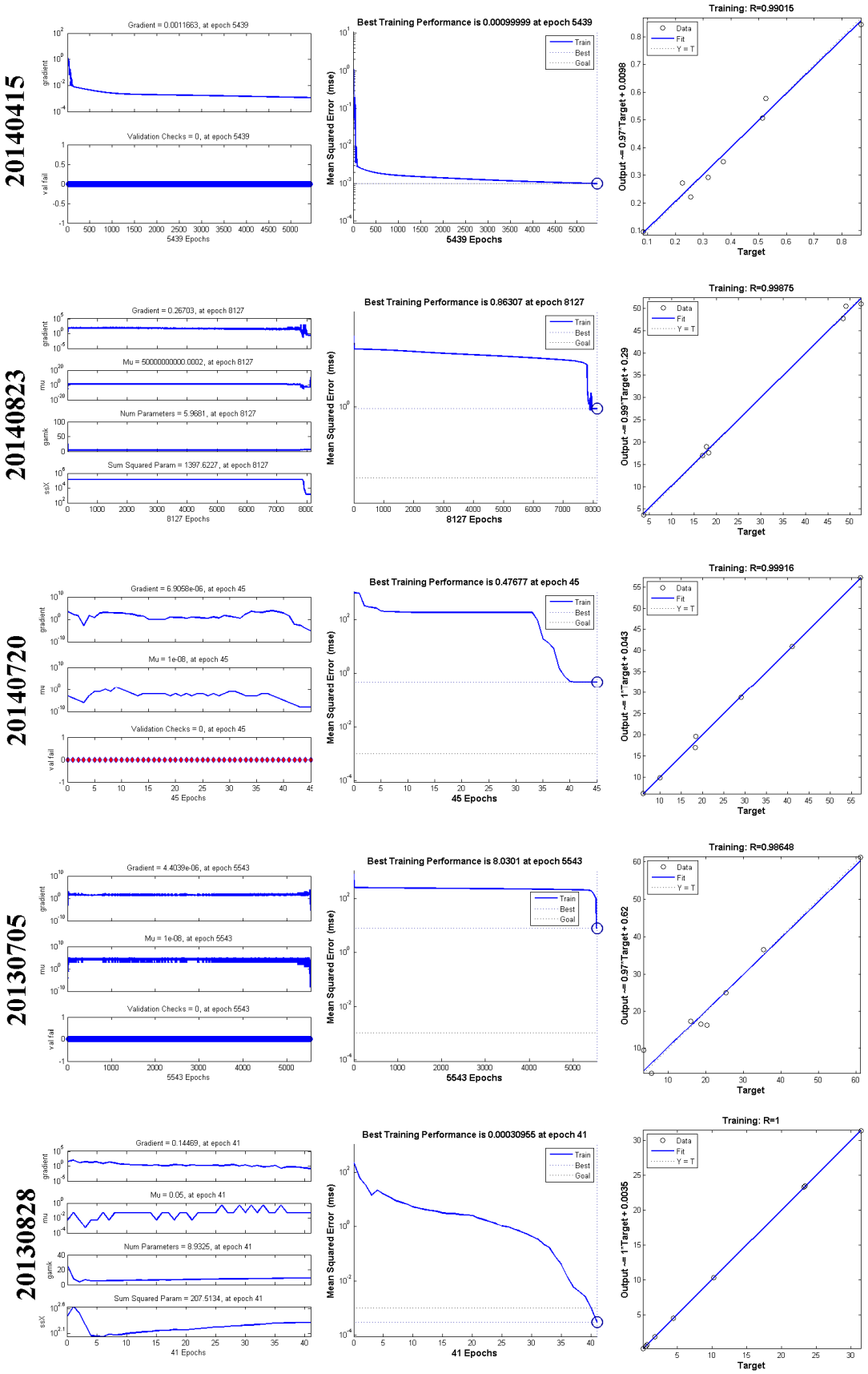
Five typical data-scarce rainfall events were used to discuss the impact of different data-scarce patterns on the NPS predictions. As shown in Figure 4, the NPS-TP training results for the NPS-TN have a faster convergence rate for the grads and lower mean squared errors. The training values for the NPS-TN are represented by an $R^2$ value that is more than 0.9, and the mean squared errors are under the permissible values or reach the flat surface rapidly. However, only one event (5 July 2013) was observed to have lower grads beyond the preset value, and its training effect was the worst because this rainfall event has peak scarcity.

The entropy values in the five data-scarce rainfall patterns are shown in Table 4. Combined with the complete data events, it is apparent that the simulated effect for 5 July 2013 has a worse fit compared with the other rainfall events, which reflects the poor training effect when there is a scarcity of peak concentration data. The peak data are the key information, and reflect the overall process of the rainfall events. However, the peak scarcity is unintentional and due to system errors. In addition, the training effect of the NPS-TP is improved over the NPS-TN.

**Table 4.** Evaluation effect of the data scarcity on the models for the five data-scarce rainfall events.

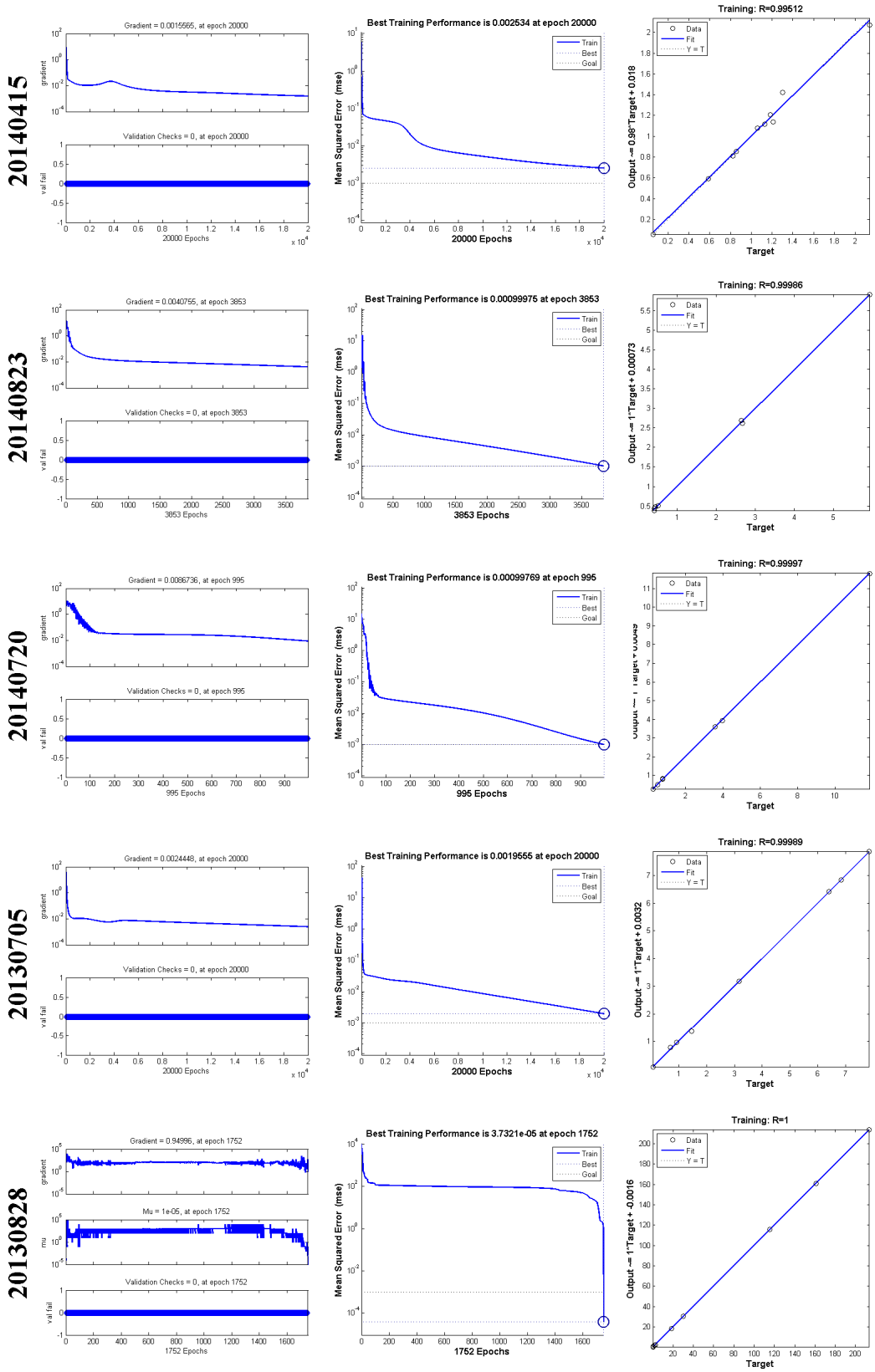
| Rainfall Events | Comprehensive Indicators K | | Traditional Indicators (NPS-TP) | | |
|---|---|---|---|---|---|
| | NPS-TN | NPS-TP | Deviation | $\overline{d}$ | $S$ |
| 15 April 2014 | 0.546 | 0.971 | 0.989 | 0.956 | 0.966 |
| 23 August 2014 | 0.937 | 0.924 | 0.934 | 0.892 | 0.943 |
| 20 July 2014 | 0.959 | 0.930 | 0.964 | 0.882 | 0.941 |
| 5 July 2013 | 0.340 | 0.948 | 0.997 | 0.908 | 0.938 |
| 28 August 2013 | 0.948 | 0.982 | 0.996 | 0.980 | 0.970 |

We further compared the evaluation results between the traditional methods and the entropy weighting method, which are shown in Table 4. As shown in the results, the rank order of the effects of the simulation results with the traditional indicators (high to low) as the following: deviations: 5 July 2013, 23 August 2013, 15 April 2014, 20 July 2014, and 23 August 2014; $\overline{d}$: 23 August 2014, 5 July 2013, 20 July 2014, 23 August 2013, and 15 April 2014; and $S$: 23 August 2014, 5 July 2013, 23 August 2013, 15 April 2014, and 20 July 2014. The application of a single indicator is limited by the indicator selection so that we cannot sum them up simply or select one of them. For instance, the effect of 5 July 2013 showed the best *deviation* but the worst $S$, which represents the rainfall amount and the average rainfall, respectively. Therefore, choosing these traditional indicators would lead to information loss during the model evaluation. Conversely, the entropy weighting method considers the advantages and characteristics of each traditional indicators and assesses the simulation results comprehensively form different perspectives [34,35]. Thus, the K values are more accurate and easier to compare.




(**a**)

**Figure 4.** *Cont.*




(**b**)

**Figure 4.** Training results for the loads in five data-scarce rainfall events: (**a**) the total nitrogen of non-point source (NPS-TN); and (**b**) the total phosphorus of non-point source (NPS-TP). Note: the pink line with dots represent the epochs are smaller.

Owing to the limited water quality data, we randomly selected 30% of the measured values as verification points, and the simulated data points were compared to the selected data to test the accuracy of the ANN during data-scarce conditions. The evaluated results are shown in Table 5, and the intuitionistic indicator is the mean percentage of the load deviation. As shown in Table 5, the effects of the training results for the runoff–pollutant load process are better in different rainfall events, and each of the load deviations is smaller. The mean percentage load deviations of the three events (15 April 2014, 23 August 2014, 28 August 2013) are higher than the other events. This is because these pollutant load data for the three individual rainfall events have peak loads nearby the flow peaks. It is apparent that the flow peak and these high values have major impacts on the training and predictive values of the ANN. In general, the simulation effects are improved, which shows that this method is feasible for estimating scarce pollutant load data.

**Table 5.** Evaluation of the predicted effects of the verification points.

| Rainfall Events | Mean Load Deviation Percentage of Verification Points (%) | |
|---|---|---|
| | NPS-TP | NPS-TN |
| 15 April 2014 | 0.150 | 0.0770 |
| 23 August 2014 | 0.153 | 0.029 |
| 20 July 2014 | 0.041 | 0.094 |
| 5 July 2013 | 0.002 | 0.038 |
| 28 August 2013 | 0.120 | 0.022 |

### *4.3. Implication for NPS Studies of Multi-Events*

Figure 5 compares pollutographs of complete rainfall events with pollutographs simulated by ANN in the same rainfall pattern (data-scarce rainfall events in light and moderate rainfall patterns). Most of the pollutographs conform to the ordinary rules (pollutographs in complete data rainfall events), and the overall tendency is consistent with the hydrographs with complete data, indicating that the method is reliable. Moreover, the model performance is worse under conditions of missing peak data, which is consistent with the abovementioned conclusions. According to the comparisons, the pollutographs of the measured points are more consistent with the ordinary rules than when the tails have missing data. Meanwhile, tail scarcity often appears in actual monitoring to reduce manpower [41]. The tails of the pollutographs have stronger linear characteristics, so a linear interpolation is used to amend the incomplete tails [42]. The pollutographs amended by linear interpolation are shown in Figure 6, indicating that the hydrographs with the tail correction are more coincident.

During the monitoring process, emphasis is placed on the discrepancy in the monitoring mechanism under different rainfall conditions. Based on the abovementioned analysis, the NPS prediction performs the best during the light rainfall events and is the worst during heavy rainfall events. Therefore, the monitoring process for the NPS can be appropriately focused on heavy rainfall conditions. Researchers should pay more attention to monitoring time to avoid peak data scarcity, especially for the NPS-TN monitoring [43]. As already suggested, peak concentration appeared after nearly five hours of runoff during light rainfall events and after nearly three hours of runoff during moderate events. Therefore, we promote peak monitoring techniques, for example, anautomatic sampler with programming, we can appropriately shorten sampling intervals for the peak lag times. Meanwhile, based on the pollutographs improved by this study, we can design the sampling scheme and avoid the risk time in order to require complete water quality data. In addition, the entropy weighting method can be effectively used to evaluate the measured and simulated data [44], showing that it can be used to comprehensively assess the discrepancies more accurately and to easily compare the results, which can be generalized to other catchments.

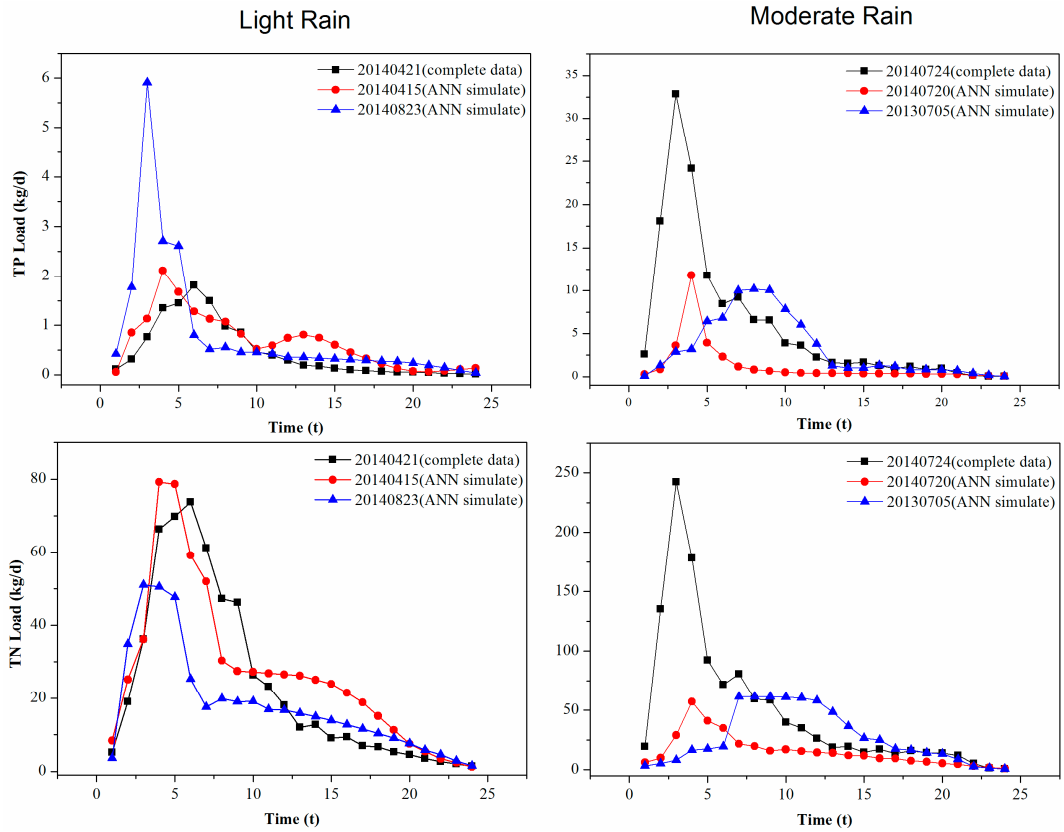



**Figure 5.** The pollutographs for different rainfall patterns.

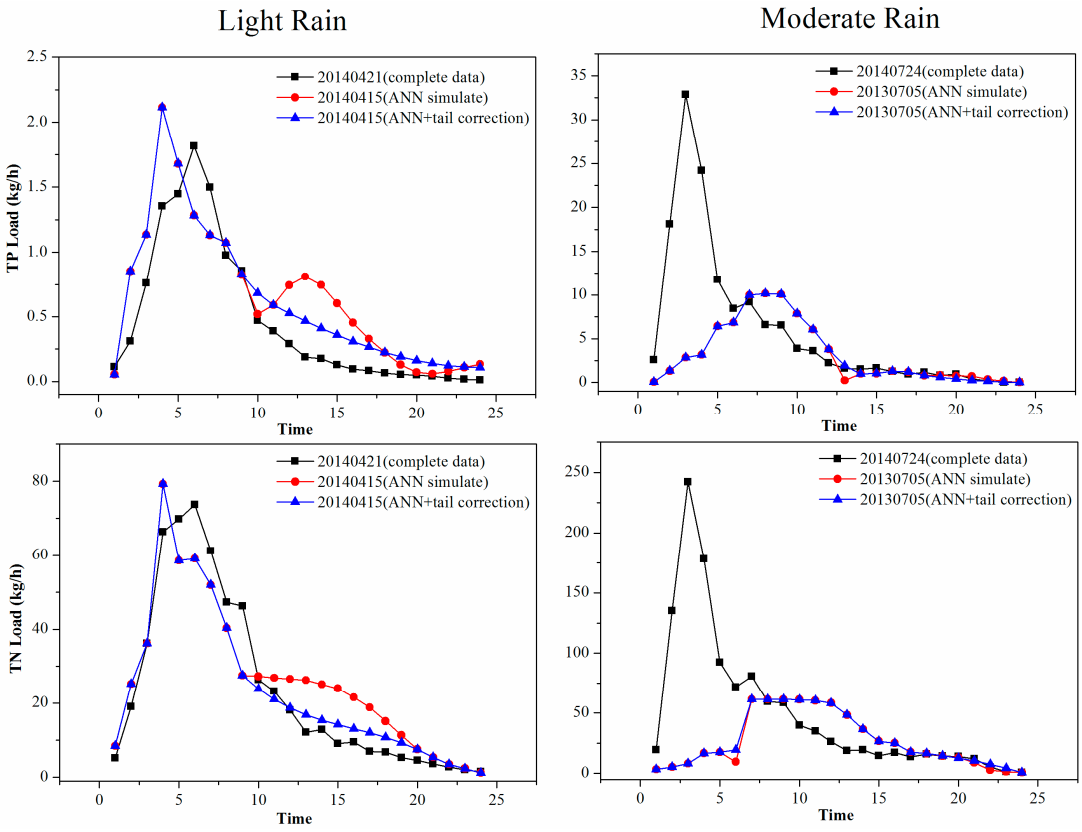



**Figure 6.** The tail amendment of the pollutographs for different rainfall patterns.

## 5. Conclusions

In this study, a new framework is proposed for the event-based NPS prediction and evaluation in data-scarce catchments. The results obtained from this study indicate that the proposed ANN had an improved performance over the NPS simulation of light rainfall events, and the NPS-TP model was always more accurate than the NPS-TN under scarce data conditions. In addition, the scarcity of the peak pollutant data has a significant impact on the model performance, so more attention should be given to the monitoring scheme of the event-based NPS studies, especially for the NPS-TN monitoring and the lag time of the peak data. Compared to the traditional indicators, the entropy weighting method can provide a more accurate ANN by considering all of the information during model evaluation. These tools could be extended to other catchments to quantify the event-based NPS pollution, especially data-poor catchments.

However, we should pay more attention to the mechanism of the NPS during multiple rainfall events because the NPS pollution was not the simple consequence of current rainfall events. Additionally, because of the computational burden, the errors and the related uncertainty of the model results were not explored, so more studies are suggested to test this new framework among more diverse regions. Meanwhile, data-driven black-box models are not good at long-term forecasting, nor are they good for examining the effect of BMPs.

**Supplementary Materials:** The following are available online at www.mdpi.com/1099-4300/19/6/265/s1, Figure S1: Training results of the loads in the different rainfall patterns: (a) NPS-TN; and (b) NPS-TP.

**Acknowledgments:** This research was funded by the National Natural Science Foundation of China (Nos. 51579011 and 51409003) and the Fund for the Innovative Research Group of the National Natural Science Foundation of China (No. 51421065).

**Author Contributions:** Lei Chen constructed the research framework and designed the study. Cheng Sun was responsible for the data analysis and code programming. Guobo Wang performed the experiments and conducted data analysis. Hui Xie and Zhenyao Shen reviewed and edited the manuscript. All of the authors have read and approved the final manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Ongley, E.D.; Zhang, X.L.; Yu, T. Current status of agricultural and rural non-point source Pollution assessment in China. *Environ. Pollut.* **2010**, *158*, 1159–1168. [CrossRef] [PubMed]
2. Li, X.F.; Xiang, S.Y.; Zhu, P.F.; Wu, M. Establishing a dynamic self-adaptation learning algorithm of the BP neural network and its applications. *Int. J. Bifurc. Chaos* **2015**, *25*, 1540030. [CrossRef]
3. Gong, Y.W.; Liang, X.Y.; Li, X.N.; Li, J.Q.; Fang, X.; Song, R.N. Influence of rainfall characteristics on total suspended solids in urban runoff: A case study in Beijing, China. *Water* **2016**, *8*, 278. [CrossRef]
4. Coulliette, A.D.; Noble, R.T. Impacts of rainfall on the water quality of the Newport River Estuary (Eastern North Carolina, USA). *J. Water Health* **2008**, *6*, 473–482. [CrossRef] [PubMed]
5. Chen, C.L.; Gao, M.; Xie, D.T.; Ni, J.P. Spatial and temporal variations in non-point source losses of nitrogen and phosphorus in a small agricultural catchment in the Three Gorges Region. *Environ. Monit. Assess.* **2016**, *188*, 257. [CrossRef] [PubMed]
6. Sajikumar, N.; Thandaveswara, B.S. A non-linear rainfall-runoff model using an artificial neural network. *J. Hydrol.* **1999**, *216*, 32–55. [CrossRef]
7. Bulygina, N.; McIntyre, N.; Wheater, H. Conditioning rainfall-runoff model parameters for ungauged catchments and land management impacts analysis. *Hydrol. Earth Syst. Sci.* **2009**, *13*, 893–904. [CrossRef]
8. Maniquiz, M.C.; Lee, S.; Kim, L.H. Multiple linear regression models of urban runoff pollutant load and event mean concentration considering rainfall variables. *J. Environ. Sci.* **2010**, *22*, 946–952. [CrossRef]
9. Chen, N.W.; Hong, H.S.; Cao, W.Z.; Zhang, Y.Z.; Zeng, Y.; Wang, W.P. Assessment of management practices in a small agricultural watershed in Southeast China. *J. Environ. Sci. Health Part A Toxic Hazard. Subst. Environ. Eng.* **2006**, *41*, 1257–1269. [CrossRef] [PubMed]
10. Sun, A.Y.; Miranda, R.M.; Xu, X. Development of multi-meta models to support surface water quality management and decision making. *Environ. Earth Sci.* **2015**, *73*, 423–434. [CrossRef]

11. Yuceil, K.; Baloch, M.A.; Gonenc, E.; Tanik, A. Development of a model support system for watershed modeling: A case study from Turkey. *CLEANSoil Air Water* **2007**, *35*, 638–644. [CrossRef]
12. Shope, C.L.; Maharjan, G.R.; Tenhunen, J.; Seo, B.; Kim, K.; Riley, J.; Arnhold, S.; Koellner, T.; Ok, Y.S.; Peiffer, S.; et al. Using the SWAT model to improve process descriptions and define hydrologic partitioning in South Korea. *Hydrol. Earth Syst. Sci.* **2014**, *18*, 539–557. [CrossRef]
13. Jeong, J.; Kannan, N.; Arnold, J.; Glick, R.; Gosselink, L.; Srinivasan, R. Development and integration of sub-hourly rainfall-runoff modeling capability within a watershed model. *Water Resour. Manag.* **2010**, *24*, 4505–4527. [CrossRef]
14. Park, Y.S.; Engel, B.A. Identifying the correlation between water quality data and LOADEST model behavior in annual sediment load estimations. *Water* **2016**, *8*, 368. [CrossRef]
15. Park, Y.S.; Engel, B.A.; Frankenberger, J.; Hwang, H. A web-based tool to estimate pollutant loading using LOADEST. *Water* **2015**, *7*, 4858–4868. [CrossRef]
16. Das, S.K.; Ng, A.W.M.; Perera, B.J.C.; Adhikary, S.K. Effects of climate and landuse activities on water quality in the Yarra River catchment. In Proceedings of the 20th International Congress on Modelling and Simulation (Modsim2013), Adelaide, Australia, 1–6 December 2013; pp. 2618–2624.
17. Chen, D.J.; Hu, M.P.; Guo, Y.; Dahlgren, R.A. Reconstructing historical changes in phosphorus inputs to rivers from point and nonpoint sources in a rapidly developing watershed in eastern China, 1980–2010. *Sci. Total Environ.* **2015**, *533*, 196–204. [CrossRef] [PubMed]
18. Melesse, A.M.; Ahmad, S.; Mcclain, M.E.; Wang, X.; Lim, Y.H. Suspended sediment load prediction of river systems: An artificial neural network approach. *Agric. Water Manag.* **2011**, *98*, 855–866. [CrossRef]
19. Tran, H.D.; Muttil, N.; Perera, B.J.C. Investigation of artificial neural network models for streamflow forecasting. In Proceedings of the 19th International Congress on Modelling and Simulation (Modsim 2011), Perth, Australia, 12–16 December 2011; pp. 1099–1105.
20. Hassan, M.; Shamim, M.A.; Sikandar, A.; Mehmood, I.; Ahmed, I.; Ashiq, S.Z.; Khitab, A. Development of sediment load estimation models by using artificial neural networking techniques. *Environ. Monit. Assess.* **2015**, *187*, 686. [CrossRef] [PubMed]
21. Dhiman, N.; Markandeya; Singh, A.; Verma, N.K.; Ajaria, N.; Patnaik, S. Statistical optimization and artificial neural network modeling for acridine orange dye degradation using in-situ synthesized polymer capped ZnO nanoparticles. *J. Colloid Interface Sci.* **2017**, *493*, 295–306. [CrossRef] [PubMed]
22. Wang, H.W.; Ai, Z.W.; Cao, Y. Information-entropy based load balancing in parallel adaptive volume rendring. In Proceedings of the International Conferences on Interfaces and Human Computer Interaction 2015, Game and Entertainment Technologies 2015, and Computer Graphics, Visualization, Computer Vision and Image Processing 2015, Las Palmas de Gran Canaria, Spain, 22–24 July 2015; pp. 163–169.
23. Khosravi, K.; Pourghasemi, H.R.; Chapi, K.; Bahri, M. Flash flood susceptibility analysis and its mapping using different bivariate models in Iran: A comparison between Shannon's entropy, statistical index, and weighting factor models. *Environ. Monit. Assess.* **2016**, *188*, 656. [CrossRef] [PubMed]
24. Yuan, Y.M.; Wei, G.A. Empirical studies of unblocked index for urban freeway traffic flow states. In Proceedings of the 2009 12th International IEEE Conference on Intelligent Transportation Systems, St. Louis, MO, USA, 4–7 Octorber 2009; pp. 1–6.
25. Kan, J.M.; Liu, J.H. Self-Tuning PID controller based on improved BP neural network. In Proceedings of the 2009 Second International Conference on Intelligent Computation Technology and Automation, Changsha, China, 10–11 Octorber 2009; pp. 95–98.
26. Chen, X.Y.; Chau, K.W. A hybrid double feedforward neural network for suspended sediment load estimation. *Water Resour. Manag.* **2016**, *30*, 2179–2194. [CrossRef]
27. Guo, Z.H.; Wu, J.; Lu, H.Y.; Wang, J.Z. A case study on a hybrid wind speed forecasting method using BP neural network. *Knowl. Based Syst.* **2011**, *24*, 1048–1056. [CrossRef]
28. Ju, Q.; Yu, Z.B.; Hao, Z.C.; Ou, G.X.; Zhao, J.; Liu, D.D. Division-based rainfall-runoff simulations with BP neural networks and Xinanjiang model. *Neurocomputing* **2009**, *72*, 2873–2883. [CrossRef]
29. Jing, J.T.; Feng, P.F.; Wei, S.L.; Zhao, H. Investigation on surface morphology model of Si3N4 ceramics for rotary ultrasonic grinding machining based on the neural network. *Appl. Surf. Sci.* **2017**, *396*, 85–94. [CrossRef]
30. Ullrich, A.; Volk, M. Influence of different nitrate-*N* monitoring strategies on load estimation as a base for model calibration and evaluation. *Environ. Monit. Assess.* **2010**, *171*, 513–527. [CrossRef] [PubMed]

31. Wilson, D.R.; Apreleva, M.V.; Eichler, M.J.; Harrold, F.R. Accuracy and repeatability of a pressure measurement system in the patellofemoral joint. *J. Biomech.* **2003**, *36*, 1909–1915. [CrossRef]
32. Ai, Y.T.; Guan, J.Y.; Fei, C.W.; Tian, J.; Zhang, F.L. Fusion information entropy method of rolling bearing fault diagnosis based on n-dimensional characteristic parameter distance. *Mech. Syst. Signal Process.* **2017**, *88*, 123–136. [CrossRef]
33. Liu, F.; Zhao, S.; Weng, M.; Liu, Y. Fire risk assessment for large-scale commercial buildings based on structure entropy weight method. *Saf. Sci.* **2017**, *94*, 26–40. [CrossRef]
34. Sun, L.Y.; Miao, C.L.; Yang, L. Ecological-economic efficiency evaluation of green technology innovation in strategic emerging industries based on entropy weighted TOPSIS method. *Ecol. Indic.* **2017**, *73*, 554–558. [CrossRef]
35. Huang, Z.Y. Evaluating intelligent residential communities using multi-strategic weighting method in China. *Energy Build.* **2014**, *69*, 144–153. [CrossRef]
36. Shen, Z.Y.; Gong, Y.W.; Li, Y.H.; Liu, R.M. Analysis and modeling of soil conservation measures in the Three Gorges Reservoir Area in China. *Catena* **2010**, *81*, 104–112. [CrossRef]
37. Shen, Z.Y.; Gong, Y.W.; Li, Y.H.; Hong, Q.; Xu, L.; Liu, R.M. A comparison of WEPP and SWAT for modeling soil erosion of the Zhangjiachong Watershed in the Three Gorges Reservoir Area. *Agric. Water Manag.* **2009**, *96*, 1435–1442. [CrossRef]
38. Shen, Z.; Qiu, J.; Hong, Q.; Chen, L. Simulation of spatial and temporal distributions of non-point source pollution load in the Three Gorges Reservoir Region. *Sci. Total Environ.* **2014**, *493*, 138–146. [CrossRef] [PubMed]
39. Gottschalk, L.; Weingartner, R. Distribution of peak flow derived from a distribution of rainfall volume and runoff coefficient, and a unit hydrograph. *J. Hydrol.* **1998**, *208*, 148–162. [CrossRef]
40. Xu, Y.F.; Ma, C.Z.; Liu, Q.; Xi, B.D.; Qian, G.R.; Zhang, D.Y.; Huo, S.L. Method to predict key factors affecting lake eutrophication—A new approach based on Support Vector Regression model. *Int. Biodeterior. Biodegrad.* **2015**, *102*, 308–315. [CrossRef]
41. Wagner, P.D.; Fiener, P.; Wilken, F.; Kumar, S.; Schneider, K. Comparison and evaluation of spatial interpolation schemes for daily rainfall in data scarce regions. *J. Hydrol.* **2012**, *464–465*, 388–400. [CrossRef]
42. Croke, B.; Islam, A.; Ghosh, J.; Khan, M.A. Evaluation of approaches for estimation of rainfall and the unit hydrograph. *Hydrol. Res.* **2011**, *42*, 372–385. [CrossRef]
43. Ryu, J.; Jang, W.S.; Kim, J.; Jung, Y.; Engel, B.A.; Lim, K.J. Development of field pollutant load estimation module and linkage of QUAL2E with watershed-scale L-THIA ACN model. *Water* **2016**, *8*, 292. [CrossRef]
44. Li, P.Y.; Qian, H.; Wu, J.H. Groundwater quality assessment based on improved water quality index in Pengyang County, Ningxia, Northwest China. *J. Chem.* **2010**, *7*, S209–S216.