High Temporal Resolution Rainfall Information Retrieval from Tipping-Bucket Rain Gauge Measurements

Yang Songa,*, Dawei Hanb, Miguel A. Rico-Ramirezb

aDepartment of Civil Engineering, University of Bristol, Bristol BS8 1TR, UK

Abstract

Disdrometer can play a vital role in restoring detailed rainfall process by providing rainfall at a high temporal resolution. Rainfall rate derived from the widely used “Tipping-Bucket rain gauge” usually neglects its temporal variation especially during the low rainfall intensity periods. This study explores a heuristic artificial neural networks (ANN) approach along with the conventional Cubic Spline Algorithm (CSA) and Multivariate Linear Regression method (MLR) for high temporal resolution rainfall rate retrieval for the period of 2007 to 2009 at Chilbolton, U.K. The Supervised Levenberg-Marquardt backpropagation algorithm and the K-folds cross-validation method are integrated in a feed-forward neural network as to implicitly detect complex nonlinear relationships and to avoid model overfitting. Results indicate ANN is performing equivalently well with CSA after training, however, with poor generalisation in test due to low correlation between input and target data, as well as the curse of dimensionality in optimum model complexity selection. MLR can be an alternative approach in rainfall rate estimation but it highly depends on the data quality.

Keywords: Rainfall rate, Disdrometer, Tipping-Bucket rain gauge, ANN, CSA, MLR, Pattern Classification;

* Corresponding author. Tel.: +447521608135.
E-mail address: ys14464@bristol.ac.uk
1. Introduction

Precipitation is a basic input for practically all scientific studies dealing with the hydrological cycle. However, precipitation is extremely difficult to measure accurately due to its intermittent nature, spatial and temporal variability and sensitivity to environmental conditions [1]. Rain gauges are the most worldwide used devices for in-situ point measurements of precipitation intensity and duration [1,2], especially for the Tipping-Bucket rain gauge since it cannot only accurately measure rainfall intensity from low-to-intermediate level, but also recording remotely with reliability and suitability [3]. Weather radars are remote sensing equipment which is widely employed in areal precipitation estimation in the hydrological and meteorological community at high resolution in time and space [4,5]. Many studies [6,7,8,9,10,11,12,13,14,15,16,17] incorporate rain gauges and weather radar in processing high-resolution merged products embedded with higher accuracy than rain gauge or radar itself in rainfall fields. However, constraints of reliable estimation of “true rainfall” still exist due to spatial and especially temporal variabilities between rain gauges and radar.

Weather radars can provide instantaneously and high temporal resolution precipitation measurement from interpreting radar reflectivity into rainfall intensity on the ground level [11,18]. In contrast, rain gauges accumulate rainfall continuously over a time period of interest [15]. However, the rainfall intensity derived from Tipping-Bucket rain gauges often neglects the precipitation variations especially for rainfall events occur in short duration [19]. Since the high temporal resolution of rainfall is vital in improving radar rainfall intensity calibration and decreasing the uncertainty of precipitation measurements as well as crucial for the hydrological response of flash flood in small urban catchments [20,21,22], it is expected that high-resolution rainfall rates at time scale can be derived from Tipping-Bucket rain gauge measurements.

Prior studies have demonstrated high temporal resolution rainfall rate retrieval based on Tipping-Bucket rain gauges [19,23,24,25,26]. [24,25] have applied the cubic spline algorithm to fit the accumulated rainfall amount collected by Tipping-Bucket rain gauges during rainfall events and differentiated the cubic spline to derive the rainfall rates at one-minute time scale. The cubic spline algorithm shows its advantages over traditional methods such as linear or quadratic approach because of its high accuracy and easy implementation [24,27]. However, cubic spline algorithm may not be feasible for rainfall data which is sensitive to the smoothness of the third or higher derivatives due to its piecewise continuous property. In addition, negative rainfall rates can be derived when large rain gradients exist at low rainfall rate. Therefore, rainfall events should be properly defined with reasonable assumptions [25]. Moreover, [26] employed regression scheme to enhance visual perception of rainfall rates that simply divide tipping-bucket volume by the time between tips. Since regression analysis is one of the statistical methods that can be used to model the relationship between input(s) and output variables [28]. Thus, it can be an alternative in rainfall rate estimation.

With the development of real-time control systems, a variety of optical and electronic instruments have been invented [2]. The disdrometer is an integral component to count raindrop size distributions and compute the velocity of falling hydrometeors at high temporal scales, during which rainfall intensity can be computed [29]. Therefore, the disdrometer has been widely used in rain gauge, radar and satellite-borne remote sensing research and operations [2,25,30,31,32,33,34,35,36,37,38]. The reason is the disdrometer not only shows a great potential in the rain gauge, radar adjustment, calibration, and reflectivity monitoring but also identifies the key source of errors in radar rainfall estimation [37]. The Joss-Waldvogel (JW) disdrometer is generally considered as the reference instrument for drop size distribution measurement and quantification at the ground surface [39,40]. Consequently, rainfall intensity obtained from disdrometor can be regarded as the reference rainfall rate.

The objective of this study is to retrieve high temporal resolution rainfall rate from Tipping-Bucket rain gauge based on the JW disdrometer data. Recent advancement has been achieved in the applied ANN (artificial neural networks) approach in water content, rainfall rate and radar reflectivity estimation based upon rain drop size distributions [41]. This is a heuristic and prospective method implicitly detecting complex nonlinear relationships between input and target values and it might be implemented in restoring high temporal resolution rainfall rate from Tipping-Bucket rain gauge data. Consequently, in this study, 1-minute rainfall rate estimation model is built and assessed between a JW disdrometer and a Tipping-Bucket rain gauge based on 3 years’ (2007 to 2009) dataset, by using the ANN approach in Chilbolton for calibration and validation. Moreover, the cubic spline algorithm and multiple linear regression methods are also implemented and compared with the ANN approach by Nash-Sutcliffe efficiency (NSE) and Root Mean Square Error (RMSE).
2. Study Area and Dataset

In this study, the data is collected from Chilbolton in the Southern England (Fig. 1). The disdrometer data is measured by an impact type Joss–Waldvogel disdrometer RD-69 and it is available from April 2003 to September 2014. Data from 2007 to 2009 is selected in this study because it is relatively complete compared with other time periods. An RW Munro 0.2mm Tipping Bucket rain gauge is collocated with disdrometer and the data is matching the same period as the disdrometer.

3. Research Methodology

3.1. Disdrometer and Tipping-Bucket Rain Gauge Mechanisms

The disdrometer consists of a transducer, a processor and an analogue to digital converter. The transducer converts the vertical momentum of a falling drop into drop diameter [42]. The detailed derivation process of rainfall intensity can be found in [42,43,44]. Rainfall rate estimated by the tipping-bucket rain gauge is simply by dividing the each tip volume by the time period between consecutive tips.

3.2. Artificial Neural Networks

Artificial neural networks (ANN) are generally described as systems of interconnected neurons, which transmit and exchange information between each other in the input, hidden and output layers, along with the weight of each connection (Fig. 2). Weights are adjusted through an error back propagation process during training in which the calculated outputs could be approximated by the target values [41]. Neural networks can be considered as being capable of computing other responses to new entries and this is defined as the test process. Data can be divided into
different groups using the ratios (90%:10% to 50%:50%) for training and testing based on data quantity [45,46]. This study splits the training and test data into 80%:20% to verify the model performance.

K-fold cross-validation is selected to guarantee that the model can provide accurate estimation on both the training and test datasets since it can honestly assess the true accuracy of the system [47]. The basic principle is first to partition the training set successively into k equal sized folds (subsamples). Afterwards, a single subsample is retained as the validation data and the k−1 remaining subsamples are used as the training data. Finally, the cross-validation process is repeated k times and each of the k subsamples is used once as the validation data [48,49]. The Levenberg-Marquardt backpropagation method is used as the training algorithm, since it has the fastest backpropagation and highly recommended as the first-choice supervised algorithms with the feed-forward neural network [50,51]. According to [52] and [53] one hidden layer is employed in this study.

Bias and variance are calculated through training iterations. However, it is a dilemma of minimizing bias and variance concurrently that stop supervised learning algorithms from generalizing beyond the training set. Fig. 2 depicts the typical behaviour of the training and validation errors with variations of model complexity. The training error (green line) tends to decrease along with the increase of model complexity. In contrast, high variance (blue line) indicates that well-trained dataset can have the risk of overfitting to noise [49]. The total error (red line) is the sum of bias and variance. The minimum total error as well as the optimum model complexity is found at the intersection of the vertical dotted line and red line. It represents that the model can not only accurately capture the regularities in the training data, but also generalize well to the test data.

![Fig. 2. (a) Schematic of a typical Multi-Layer Artificial Neural Network; (b) Bias and Variance Trade-off.](image)

In this study, the neural network model setups are configured as one hidden layer and one output layer. Tipping-Bucket rainfall rate, cubic spline rainfall rate and each tip duration are considered as 3 input variables. Sensitivity test for the number of each input variable is applied based on a regression analysis between the input and target variables. 

![Fig. 3. Regression Analysis between Input and Target Variables.](image)
consecutive Tipping-Bucket rainfall rates are employed in the regression analysis to predict the target variable (rainfall rate from the Disdrometer) captured in a black circle which is corresponding to the fourth input data, the process is repeated by moving along the input variable’s time series. In this study, the number of each input variable ranges from 1 to 5, therefore, the total input variables vary from 3 to 15, and only one target variable is considered. Moreover, rainfall rate is chosen at the central position of each tip duration. The hidden layer size is increased from 1 to 5 and five folds are used for cross-validation. Estimated rainfall rate is obtained in the output layer. The agreement of rainfall rates, derived from ANN, CSA and MLR in terms of the reference rainfall rate measured by the Disdrometer, is assessed by using Nash-Sutcliffe Efficiency (NSE) and Root Mean Square Error (RMSE) as performance indicators. NSE is dimensionless and the unit of RMSE is mm/h.

4. Results and Discussion

Table 1 and Table 2 present the model performance of rainfall rates simulated by ANN, CSA and MLR from the training and test sets respectively from 2007 to 2009 at Chilbolton. All values are obtained from the optimum model complexities. Table 1 indicates that ANN and CSA are performing equivalently well in the training dataset as NSE is higher than 0.850 and RMSE is below 2.400. Both methods outperform MLR which has around 0.1 less in NSE and 0.8 more in RMSE. However, it can be seen that increasing inputs in ANN and MLR does not show apparent improvement in the model performance as the more inputs are added, and the smaller corresponding correlation can be derived with the target. The test results in Table 2 reveal that compared with ANN and CSA, MLR has the best performance in rainfall rate estimation and especially integrated with 3 inputs that have the highest NSE(0.841) and lowest RMSE(2.212). In contrast, ANN has a poor performance in rainfall rate estimation as NSE is about 0.1 lower and RMSE is 0.5 higher than CSA. The reason is due to the input rainfall rate data derived from Tipping-Bucket rain gauge is sensitive to the integration period especially at short integration periods [37]. Fig. 4 shows the scatter plots of the correlation coefficient of rainfall amount in minute and hour time integrations between the Tipping-Bucket and disdrometer for the entire time period at Chilbolton. The statistic indicator is raised apparently from 0.78 to 0.95 since time scale increase can substantially decrease errors [19]. However, it should be specified that better results of ANN can be found in other model configurations since it is due to the issue of “Curse of Dimensionality”: increase of input will lead to the exponential rise of dimensionality, and thus require more samples to stop neural networks memorizing the data [54]. As a result, the generalization of ANN in optimum model complexity setups is also a limitation while applying the neural network method.

5. Conclusions

This study has conducted comparisons of the artificial neural networks approach, cubic spline algorithm and multi-linear regression method in high temporal resolution rainfall rate estimation. Rainfall rates can be derived from the
Joss-Waldvogel disdrometer and tipping bucket rain gauge from 2007 to 2009 at Chilbolton, UK. The results show that ANN is performing as good as CSA after training, however, its generalization is poor in rainfall rate estimation, and this is not only because of the low correlation of rainfall amount at 1-minute time scale between the Tipping-Bucket rain gauge and disdrometer, but also the optimum model complexity selection due to the curse of dimensionality. MLR can be an alternative approach in rainfall rate estimation though it highly depends on the data quality and quantity.

Table 1. Statistical performances between the simulated and the Disdrometer rainfall rates in training dataset from 2007 to 2009 at Chilbolton.

<table>
<thead>
<tr>
<th>NSE</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANN</td>
<td>CSA</td>
</tr>
<tr>
<td>3</td>
<td>0.876</td>
</tr>
<tr>
<td>6</td>
<td>0.859</td>
</tr>
<tr>
<td>9</td>
<td>0.878</td>
</tr>
<tr>
<td>12</td>
<td>0.860</td>
</tr>
<tr>
<td>15</td>
<td>0.853</td>
</tr>
</tbody>
</table>

Table 2. Statistical performances between the simulated and the Disdrometer rainfall rates in test dataset from 2007 to 2009 at Chilbolton.

<table>
<thead>
<tr>
<th>NSE</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANN</td>
<td>CSA</td>
</tr>
<tr>
<td>3</td>
<td>0.702</td>
</tr>
<tr>
<td>6</td>
<td>0.747</td>
</tr>
<tr>
<td>9</td>
<td>0.697</td>
</tr>
<tr>
<td>12</td>
<td>0.702</td>
</tr>
<tr>
<td>15</td>
<td>0.728</td>
</tr>
</tbody>
</table>

Acknowledgements

The author would like to acknowledge the British Atmospheric Data Centre for providing the dataset to accomplish this study.

References

1199


