Peer-reviewed version available at Water 2019, 11, 977; doi:10.3390/w11050977

# Improving Monsoon Precipitation Prediction using Combined Convolutional and Long Short Term Memory Neural Network

Qinghua Miao[1], Baoxiang Pan[2,*], Hao Wang[1,3]

Kuolin Hsu[2], Soroosh Sorooshian[2]

1 State Key Laboratory of Hydroscience and Engineering, Department of Hydraulic Engineering, Tsinghua University, Beijing, 100084, China

2 Center for Hydrometeorology and Remote Sensing, University of California, Irvine, CA, 92617

3 Department of Water Resources, China Institute of Water Resources and Hydropower Research, Beijing, 100038, China

Submitted to ***Water***

*Corresponding author address:

**Baoxiang Pan,** Email: baoxianp@uci.edu

Peer-reviewed version available at Water 2019, 11, 977; doi:10.3390/w11050977

**Abstract:**

Precipitation downscaling is widely employed for enhancing the resolution and accuracy of precipitation products from general circulation models (GCMs). In this study, we propose a novel statistical downscaling method to foster GCMs' precipitation prediction resolution and accuracy for monsoon region. We develop a deep neural network composed of convolution and Long Short Term Memory (LSTM) recurrent module to estimate precipitation based on well-resolved atmospheric dynamical fields. The proposed model is compared against GCM precipitation product and classical downscaling methods in the Xiangjiang River Basin in South China. Results show considerable improvement compared to the ECMWF-Interim reanalysis precipitation. Also, the model outperforms benchmark downscaling approaches, including 1) quantile mapping, 2) support vector machine, and 3) convolutional neural network. To test the robustness of the model and its applicability in practical forecast, we apply the trained network for precipitation prediction forced by retrospective forecasts from ECMWF model. Compared to ECMWF precipitation forecast, our model makes better use of the resolved dynamical field for more accurate precipitation prediction at lead time from 1 day up to 2 weeks. This superiority decreases along forecast lead time, as GCM's skill in predicting atmospheric dynamics being diminished by the chaotic effect. At last, we build a distributed hydrological model and force it with different sources of precipitation inputs. Hydrological simulation forced with the neural network precipitation estimation shows significant advantage over simulation forced with the original ERA-Interim precipitation (with NSE value increases from 0.06 to 0.64), and

Peer-reviewed version available at Water 2019, 11, 977; doi:10.3390/w11050977

the performance is just slightly worse than the observed precipitation forced simulation (NSE=0.82). This further proves the value of the proposed downscaling method, and suggests its potential for hydrological forecasts.

**Key Words**: Precipitation downscaling; Convolutional Neural Networks; Long Short Term Memory Networks; Hydrological simulation

## 1. Introduction

Precipitation is a primary forcing in hydrological systems (Nijssen and Lettenmaier 2004). Obtaining accurate and reliable precipitation data at relevant spatial and temporal scales is crucial for efficient water resources management and timely warning of precipitation-related natural hazards, such as flood and drought. To sustain a reasonably long lead-time for the above-mentioned applications, it is imperative to employ precipitation prediction techniques.

For short-term range up to climate range, numerical weather/climate modeling is perhaps the only reliable tool for predictions. Through the past decades, numerical models have achieved impressive progresses in predicting atmospheric dynamics and physics (Bauer et al., 2015). Here, dynamics refer to atmospheric state variables (i.e., density, pressure, temperature, and velocity) that are explicitly described by atmospheric primitive equations and resolved by numerical partial differential equation solvers, while physics refer to the unresolved processes that are diagnosed from the resolved variables based on empirical parameterization schemes. Precipitation results from complex processes that are mostly parameterized. Compared to models' relatively

Peer-reviewed version available at Water 2019, 11, 977; doi:10.3390/w11050977

satisfactory skills in resolving the atmospheric dynamics, models' precipitation estimation suffers from multiplicate sources of errors (Tapiador et al., 2018), and the skill has been described as "dreadful" (Stephens et al., 2010). The uncertainties for precipitation prediction generally stem from the following aspects: 1) models' dynamical forcings are of limited resolution for making detailed representation of cloud microphysics; 2) we usually do not have direct observation for the initial distribution state for cloud hydrometeors of liquid, solid, or mixed phases; 3) the evolution and interaction of precipitating cloud hydrometeors are not well described, due to our limited understandings or computation resources. Models' deficiencies in each of the above aspects quickly reveal themselves in the precipitation product (Tapiador et al., 2018). Many studies have revealed that the accuracy of the rainfall prediction in GCMs (such as ECMWF and NCEP) is far from enough to be used directly in East Asian monsoon region (Kang et al. 2002; Wang et al. 2004). Besides the "dreadful" effect, the resolution of the computing grids is also usually too coarse for hydrological simulations (Xu 1999; Ghosh, 2010; Meehl et al., 2007).

To improve the precipitation estimations of GCMs, hydrologists have developed various downscaling methods, including the dynamical downscaling and statistical downscaling (Benestad, 2007; Benestad and Haugen, 2007; Christensen et al. 1997, 2007; Hanssen et al. 2005; Prudhomme et al. 2002; Xu, 1999). Dynamical downscaling usually includes running a regional climate model with the initial and boundary conditions provided by GCMs. The massive computational cost and the requirement of local conditions severely limited its application in many regions. Statistical

Peer-reviewed version available at Water 2019, 11, 977; doi:10.3390/w11050977

downscaling establishes statistical links between large scale weather and local observation. Despite some limitations, such as the stationarity assumption in the predictor–predictand relationship and requirement for long observation records, statistical downscaling method is straightforward and computationally efficient, thus can be easily applied to different regions.

There are many statistical downscaling method proposed by past researchers. The simplest form is linear regression, which estimates target predictand using an optimized linear combination of local circulation features (Murphy et al., 2000; Schoof and Pryor, 2001; Li and Smith, 2009). The features are usually represented as the leading Principal Components (PC) of the moisture, pressure and wind field. While the leading PCs represent the inner linear structure of the circulation field at climate scale, they might not be directly related to the predictand at weather scale. For instance, frontal precipitation is closely related to its corresponding cyclone geometries, such as depression intensity, coverage and distance. These geometries are highly varied from event to event, not all of them could be well illustrated through the leading eigenvectors of the circulation field.

Some other methods estimate precipitation based on non-linear features of relevant circulation field, such as the self-Organizing Map (SOM) (Hope, 2006). SOM clusters synoptic circulation field into different categories, with each category defining a spatial rainfall pattern. However, similar charge for principal component regression applies here as well, since these features are not designed based on predictor-predictand connection but the inner structure of the predictor, which does not necessarily relate to

Peer-reviewed version available at Water 2019, 11, 977; doi:10.3390/w11050977

weather scale precipitation distribution.

Another category of machine learning algorithms uses kernel tricks to implicitly transfer raw input data into feature space, from which the learning algorithm could better extract useful information for a given target. This is achieved through applying a kernel function to estimate two points' distance in the feature space by transforming their dot-product in the input space. Kernel trick allows customizing features toward specific target by selecting kernel functions and their parameters. Relevant applications include kernel regression (Kannan and Ghosh, 2013) and support vector machine (Tripathi et al., 2006; Pan and Cong, 2016). The design of kernels rely heavily on modeler's prior knowledge. For the problem here, it is difficult to design kernel functions that explicitly consider the precipitation related influences of depression intensity, coverage or distance for different cyclone events or convective activities.

The requirement for recognizing key circulation features of different appearances and positions lead us to adopt deep Neural Networks. Artificial Neural Networks (ANN) have also been widely applied to precipitation downscaling problems in the past (Guhathakurta, 2008; James, 2010; Norton, 2011; Schoof and Pryor, 2010). However, conventional ANNs tends to get trapped in poor local minima, and are relatively worse or no better than other downscaling methods. Recent progress in ANN such as the Convolutional Neural Networks (CNNs) and recurrent neural networks (RNNs), have gain great success in many application like speech recognition, visual object recognition and object detection, etc. Works by Vandal et al. (2017) and Pan et al. (2019) also have proved the effectiveness of CNNs in precipitation downscaling in United

Peer-reviewed version available at Water 2019, 11, 977; doi:10.3390/w11050977

States. Shi et al. (2015) proposed a new method for radar precipitation nowcasting by combining the CNNs with Long Short Term Memory Networks (LSTM).

In this study, we attempt to improve the precipitation estimations using the state of art deep learning methods. Instead of correcting biases for numerical model's precipitation estimations, we propose to predict the rainfall based on model resolved circulation dynamics. In other words, we try to replace the unresolved processes beyond the original numerical simulation with the deep neural networks. This is motivated by the fact that, although the predictive skills for both precipitation and circulation dynamics are diminishing along forecast lead time, the primitive variables are generally more reliable and sustains a longer usable forecast range.

After the Introduction, Section 2 presents the study area and datasets. Section 3 introduces the downscaling methods and hydrological model that are used in this study. The results of precipitation estimation as well as its performance evaluation are described in Section 4. Finally, a brief summary and the major conclusions are provided in Section 5.

## 2. Study area and datasets

### 2.1 Study area

The Xiangjiang River, a tributary of the Yangtze River with a drainage area of 63980 $km^2$ at the Hengshan hydrological station, was selected as the study area (see Figure 1). This basin is located in southeastern Hunan Province in South China and extends from longitudes of 109.27°E to 114.99°E from latitudes of 23.98°N to 28.64°N. The climate of this region is humid subtropical monsoon, with a mean annual

Peer-reviewed version available at Water 2019, 11, 977; doi:10.3390/w11050977

precipitation of approximately 1366 mm. The precipitation in this region exhibits high seasonal and inter-annual variability and mainly occurs between April and September. The annual average runoff depth is approximately 822 mm. The area has complex topography, with elevations ranging from 30 m to 2097 m above sea level. The headwater regions are characterized by steep mountain slopes and deep fluvial valleys and consequently suffer from flash flooding. The lower portion of the river flows through the floodplain where the outlet station Hengshan is located.

[Figure 1 is to be inserted about here]

**2.2 Datasets**

Data used to train and validate the downscaling methods includes observed rainfall data and the predictor for precipitation estimation. The observation data is the China Gauge-Based Daily Precipitation Analysis product developed by the National Meteorological Information Center (Shen and Xiong, 2016), with a temporal-spatial resolution of one day and 0.25°, and can be downloaded from the website (http://data.cma.cn). Products from the European Centre for Medium Range Weather Forecasts (ECMWF) Interim Reanalysis (ERA-Interim) (Dee et al. 2011) are selected as the predictors for precipitation estimation. It has a temporal-spatial resolution of one day and 0.75° .The potential predictor candidates used in this study include the following simulated atmospheric circulation variables: the mean sea level pressure (MLS), the total column water (TCW), the convective available potential energy (CAPE); and also the geo-potential height (GH), the U wind component (UW), the Vertical velocity (VV), the air temperature (T), potential vorticity (PV) at

Peer-reviewed version available at Water 2019, 11, 977; doi:10.3390/w11050977

500/700/850/925/1000 hpa. Detail description of this variables can be found in the website (https://apps.ecmwf.int/codes/grib/param-db). The final predictors were determined through a trial and error method. Data from longitudes of 106°E to 125°E, and from latitudes of 20°N to 33°N are extracted for predictors. Range of predictand extend from 115°E to 120°E and from 24°N to 28°N.

To further validate the downscaling methods, we also used the ECMWF subseasonal to seasonal (S2S) prediction project database (hindcasts) (Vitart et al., 2017), which contains the same predictors with ERA-Interim. The ECMWF hindcasts model is initialized with realistic estimates of their observed states, hereafter iteratively predicts the weather for a preset extension without any boundary constrains. It restart on every Monday and Thursday from 1995 to 2016 to forecast next 46 days weather evolution, using 11 ensemble members. It is coupled with ocean model but not sea ice model. Together there are 1869 hindcast experiments during out validation period.

The other data used in this study include geographical information, which was used to build the distributed hydrological model; meteorological data, which were used as input data for the hydrological model; and discharge data, which were used to calibrate and validate the hydrological model. Catchment topography is represented using a digital elevation model (DEM) with a spatial resolution of 90 m and the DEM data were downloaded from the SRTM Database (http://srtm.csi.cgiar.org). The soil map was obtained from the China Dataset of Soil Properties for Land Surface Modeling (Dai et al. 2013). Land use/cover data were obtained from the Environmental and Ecological Science Data Center of West China (http://westdc.westgis.ac.cn/) and have a resolution

Peer-reviewed version available at Water 2019, 11, 977; doi:10.3390/w11050977

of 100 m. Daily discharge data from the Hengshan hydrological station are available from 2007 to 2013 and were obtained from the Hydrological Year Book to calibrate and validate the hydrological model. Daily meteorological data were obtained from the China Administration of Meteorology and includes precipitation, mean, maximum and minimum air temperatures, sunshine duration, wind speed, and relative humidity data. The meteorological data were used to estimate potential evaporation by using the Penman equation (Penman 1948), which was also used in the hydrological model.

## 3. Methods

### 3.1 Downscaling Methods

#### Convolutional Neural Networks

The convolutional neural networks (CNNs) is a special type of the Deep Neural Networks. For a regular neural network, a statistical connection between the inputs and the outputs is constructed through hierarchical connected layers of neurons. Each neuron is a computing unit that receives some inputs, performs a dot product and optionally follows a non-linear transformation. For supervised learning problems (i.e., classification and regression), a loss function is defined by comparing the network's output estimations with observations. The network parameters are trained through minimizing the loss function using gradient descent, which is known as backpropagation training.

Different from the full-connected networks, CNNs involves two special matrix operators: a convolutional layer and a pooling layer. Units in convolutional layers are only connected to specific local patches through a set of leant filters. All units in a single

Peer-reviewed version available at Water 2019, 11, 977; doi:10.3390/w11050977

layers share the same filters. In this way, it greatly reduces the number of parameters in networks and allows the networks to be deeper and more efficient. Usually a non-linear function (such as rectified linear unit or hyberbolic tangent, etc) is applied after convolution oprtators (Krizhevsky et al., 2012). Then the pooling layers are used to merge semantically similar local features into one (Lecun et al., 2015). This is due to the fact that relative positions of features that make up the motifs may vary somewhat, thus coarsing the positions of each feature can help to detect reliably motifs. Typical pooling layers partition a feature map into a set of non-overlapping rectangles and output the maximum or the average value for each sub-region (Deep Learning Tutorials).

In addition, to reduce overfitting problems, the dropout and batch-normalization methods are also adopted in this study. Dropout helps reduce overfitting by randomly setting some weight parameters or outputs of the hidden layers to zero with predefined probability during training process (Hinton et al., 2012). Batch-normalization alleviates internal covariate shift by normalizing layer inputs (Ioffe and Szegedy, 2015). The CNNs are implemented in tensorflow (Abadi et al., 2016) under python platform. Different predictors are fed as different channels in the inputs. The Mean Square Error between the simulated and observed precipitation is used as loss function, and the Adam gradient-based optimizer is used to minimize the loss function. Figure 2 illustrates the architecture of the CNN networks used in this study.

[Figure 2 is to be inserted about here]

**Combination of CNN and Long Short Term Memory Networks**

Peer-reviewed version available at Water 2019, 11, 977; doi:10.3390/w11050977

The long short term memory networks (LSTM) (Hochreiter et al., 1997) is a special type of recurrent neural networks (RNNs). RNNs contain a feed back connections that allows past information to affect the current output, thus is very effective for tasks involving sequential inputs (Lecun et al., 2015). As an extension of the conventional RNNs, LSTM introduces a special so-called memory cell, which acts like an accumulator, to learn long-term dependency in a sequence, and make the optimization much easier. This cell is self-connected and will copy its own real-valued state and accumulate the external input. Simultaneously each cell is controlled by three multiplicative units –the input, output and forget gates—to determine whether to forget past cell status or to deliver output to the last state, which allow LSTM store and access information over long periods. Following the work of Graves (2013), the formulation are shown as follows:

$$
\begin{aligned}
i_t &= \sigma\left(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i\right) \\
f_t &= \sigma\left(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f\right) \\
c_t &= f_t c_{t-1} + i_t \tanh\left(W_{xc}x_t + W_{hc}h_{t-1} + b_c\right) \\
o_t &= \sigma\left(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o\right) \\
h_t &= o_t \tanh\left(c_t\right)
\end{aligned}
\tag{251}
$$

Where $i, f, o$ represent the input, forget and output gate; $c$ is the memory cell; $\sigma$ is the logistic sigmoid function; h is a hidden vector; $W$ and $b$ are gate matrix and bias term.

CNNs is good at dealing with spatially related data while the LSTM is good at temporal signals. Combination of these two methods can make use of the best of both worlds. To this end, we simply use the convolutional layers to extract the spatial features

Peer-reviewed version available at Water 2019, 11, 977; doi:10.3390/w11050977

of the raw input and then feed them to the LSTM networks (hereinafter referred to as ConvLSTM). In this study, predictors from the past 7 days are used to estimate daily rainfall. The structure of the convolutional layers are just the same as previously mentioned. And 400 hidden layers are set up in the LSTM, which is also implemented in tensorflow (Abadi et al., 2016) under python platform.

**Support Vector Machine**

The Support Vector Machine (SVM) was first developed by Vapnik (1995) for binary classification. The principle of SVM algorithm is to find the optimal separation hyperplane between two classes by maximizing the boundary margin between the closest points of the class on the boundaries (Sehad, 2017). These points are called support vectors.

In SVM regression, the input $X$ is first mapped into a higher dimension feature space, and then a linear model can be constructed as follows (Cover, 1965; Smola, 1996):

$$f(X,w)=\sum_j w_j g_j(X)+b$$

Where $g_j$ denotes a set of nonlinear transformations, $w$ and $b$ are model parameters to be calibrated. Defining the $\varepsilon$-insensitive loss function $L_\varepsilon\left(y,f(X,w)\right)$ (Vapnik, 1995):

$$L_\varepsilon\left(y,f(X,w)\right)=\begin{cases}0 & if \quad |y-f(X,w)|<\varepsilon \\ |y-f(X,w)|-\varepsilon & otherwise\end{cases}$$

Then the empirical risk can be calculated as:

Peer-reviewed version available at Water 2019, 11, 977; doi:10.3390/w11050977

$$R_{emp}(w)=\frac{1}{n}\sum_{i=1}^{n}L_{\varepsilon}\left(y_i,f\left(X_i,w\right)\right)$$

Following Haykin (2003)'s regularization theory, by introducing (non-negative) slack variables $\xi_i$, $\xi_i^*$ to measure the deviation of training samples outside $\varepsilon$-insensitive zone, the parameters $w$ and $b$ are estimated by minimizing the cost function:

$$\min\frac{1}{2}\|w\|+C\sum_{i=1}^{N}\left(\xi_i+\xi_i^*\right)$$

$$s.t.\begin{cases}y_i-f\left(X,w\right)\le\varepsilon+\xi_i\\-y_i+f\left(X,w\right)\le\varepsilon+\xi_i^*\\\xi_i\ge 0\\\xi_i^*\ge 0\end{cases}$$

where C is a positive real constant. This optimization problem can be solved by the method of Lagrangian multipliers (Haykin, 2003):

$$w=\sum_{i=1}^{N}\left(\alpha_i-\alpha_i^*\right)g(\mathrm{X}_i)$$

where $\alpha_i$ and $\alpha_i^*$ are the Lagrange multipliers, which are positive real constants.

In this study, the SVM model is implemented in sklearn (Pedregosa et al., 2011) under python platform. The training of SVM includes selecting the kernel function, and determining the model parameters C and Gamma. These parameters are optimized through the grid search mechanism (Baesens et al., 2002), and C=10， Gamma=0.001 and radial basis function are used in this study.

**Quantile Mapping Method**

The quantile mapping (QM) method (Panofsky and Brier, 1968) is a relatively

Peer-reviewed version available at Water 2019, 11, 977; doi:10.3390/w11050977

simple approach that has been successfully used in hydrologic studies (e.g., Boé et al., 2008). It uses the cumulative frequency curve of the observed precipitation to correct the simulated rainfall so that the corrected rainfall will have the same cumulative frequency curve with the observed one. Figure 3 illustrates how quantile mapping method works. For each grid, calculate the cumulative frequency function of the simulated precipitation ( $CF_{sim}(p)$ ) and the observed precipitation ( $CF_{obs}(p)$ ), respectively. Then for a specific precipitation in the validation period $\mathrm{Pre}_{vali}$, we can calculate its frequency on $CF_{sim}(p)$ as $CF_{sim}^{-1}(\mathrm{Pre}_{vali})$. And then the corresponding precipitation on cumulative frequency function of observed rainfall is just the corrected precipitation $\mathrm{Pre}_{corr} = CF_{obs}(CF_{sim}^{-1}(\mathrm{Pre}_{vali}))$.

[Figure 3 is to be inserted about here]

### 3.2 Statistical evaluation based on gauge rainfall data

To qualitatively evaluate the downscaling methods, the following metrics were adopted: relative bias (RB), and the root mean square error (RMSE), which were used to show the error and bias of the simulated precipitation compared with observed rainfall data (Ebert et al. 2007; Schaefer 1990); the correlation coefficient (CC), which aims to show the consistency between the predicted rainfall and the observed rainfall. The metrics are calculated as follows:

$$RB = \frac{\sum_{i=1}^{N}(P_i - G_i)}{\sum_{t=1}^{T} G_i}$$

Peer-reviewed version available at Water 2019, 11, 977; doi:10.3390/w11050977

315 $$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(P_i - G_i\right)^2}$$

316 $$CC = \frac{\sum_{i=1}^{N}(G_i - \bar{G})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^{N}(G_i - \bar{G})^2 \sum_{i=1}^{N}(P_i - \bar{P})^2}}$$

317 where $P_i$ and $G_i$ denote the predicted rainfall and observed gauge rainfall,
318 respectively.

319 **3.3 Evaluation through hydrological modeling**

320 **Description of the Distributed Hydrological Model and Model Validation**

321 The hydrological model used in this study is a distributed geomorphology-based
322 hydrological model (GBHM) developed by Yang et al. (1998, 2002, 2004). In the
323 GBHM, the study basin is divided into a number of sub-catchments linked by the river
324 network and ordered by the Horton-Strahler scheme. Then, grids within a sub-
325 catchment are grouped into several flow intervals according to the flow distance to the
326 outlet. The runoff generated from the grids within a flow interval contributes to the
327 main stream with the same flow distance, and each grid is represented by a number of
328 topographically similar "hillslope-valley" systems, which is the basic unit of the
329 hydrological simulation (Yang et al. 2002, 2015).

330 The GBHM mainly consists of a hillslope module and a kinematic wave flow
331 routing module (Yang et al. 2002, 2004). In the hillslope module, the GBHM simulates
332 the hydrological processes, including interception, evapotranspiration, infiltration,
333 overland flow, unsaturated flow and groundwater flow. Evapotranspiration is calculated
334 as evaporation from water stored in the canopy, on the surface and from the soil surface

Peer-reviewed version available at Water 2019, 11, 977; doi:10.3390/w11050977

in addition to transpiration from the root zone. The topsoil is divided into several layers according to depth, and the vertical soil water movement is described using the Richards equation. Overland flow is described using the one-dimensional kinematic wave equation. Subsurface flow along the hillslope occurs when the soil water content exceeds field capacity. The groundwater aquifer (corresponding to each grid) is discretized and treated as an individual storage compartment. The water exchange between the groundwater and river channel is expected to be steady and is estimated using Darcy's law (Yang et al. 2002). Most model parameters are defined according to their physical meaning, either based on in situ measurements or regional databases. Only a few parameters must be calibrated, such as the hydraulic conductivity of the groundwater.

In this study, 170 sub-catchments are divided with a grid resolution of 2 km, as suggested by Yang et al. (2001) and Miao et al. (2016). The GBHM simulates the hydrological processes at the hourly time step, and is calibrated for the period of 2007-2010, and is validated for the period of 2011-2013. The Nash and Sutcliffe coefficient (NSE) and relative bias (RB) are adopted to evaluate the model performance and are defined as follows:

$$NSE = 1 - \frac{\sum_{t=1}^{T}(Q_{\mathrm{obs}}^{t} - Q_{sim}^{t})^{2}}{\sum_{t=1}^{T}(Q_{obs}^{t} - \overline{Q_{obs}})^{2}}$$

$$RB = \frac{\sum_{t=1}^{T}(Q_{obs}^{t} - Q_{sim}^{t})}{\sum_{t=1}^{T}Q_{obs}^{t}}$$

Peer-reviewed version available at Water 2019, 11, 977; doi:10.3390/w11050977

where $Q_{obs}^{t}$ and $Q_{sim}^{t}$ denote the observed and simulated discharge and $\overline{Q_{\mathrm{obs}}}$ denotes the average values of the observed discharges during the simulation period T. Table 1 contains the calibration and validation results obtained from using gauge rainfall input data for the model. The NSE values for the calibration and validation period are greater than 0.8, and the absolute values of RB are less than 0.05, indicating that GBHM has good performance in the study basin.

[Table 1 is to be inserted about here]

## 4. Results and Discussion

### 4.1 Precipitation Estimation Performance With different Predictors

In the past studies, many predictors have been used for precipitation downscaling, such as geopotential height (e.g., Kidson and Thompson 1998; Zorita and von Storch 1999), sea level pressure (Cavazos 1999), geostrophic vorticity (Wilby et al. 1998), or wind speed (Murphy 1999). The choice of the predictors vary across different regions, characteristics of the large-scale atmospheric circulation, seasonality and geomorphology (Anandhi et al. 2009). So in this section, we designed four experiments to evaluate the forecast performance of different predictors. As listed in Table 2, Experiment 0 represents the original ECMWF- Interim precipitation; Experiment 1 uses the mean sea level pressure as predictor; Experiment 2 uses the geopotential height at 500/700/850/925/1000hpa as predictors; Experiment 3 uses the geopotential height at 500/700/850/925/1000hpa as well as the total column water as predictors; Experiment 4 uses as much climatology variables as possible as predictors, including all the circulation variables described in section 2.2. Experiments 1-4 all use the CNN

Peer-reviewed version available at Water 2019, 11, 977; doi:10.3390/w11050977

networks as downscaling method. Models were trained during the period from 1979 to 2002, and were validated from 2003 to 2016.

Table 2 lists the metrics of results of these experiments (Note that the indexes are calculated for each grid respectively, and the average value for these indexes is shown in Table 2; same in Table 3 and Figure 5). Precipitation estimations are plotted against observed ones in Figure 4. Results of experiment 0 shows a relatively low correlation coefficient with observed data (with CC of 0.29) (for validation period, the same hereafter), along with an overestimation of the precipitation (with RB of 12.75%), the root-mean-square error is 11.48mm/day. These metrics indicate a bad performance of original ERA-interim rainfall. Results of experimental 1-4 all outperform original ERA-interim rainfall. Specifically, CC, RB and RMSE values for Experiment 1 is 0.54, 5.53%, 8.86 mm/day. All metrics shows an improvement over Experiment 0, but far from enough; and the scatter plots show that it severely underestimates most high-intensity rainfall. The CC, RB and RMSE values for Experiment 2 is 0.66, 4.08%, and 7.93, and the scatter plot indicates the model could well simulate the high-intensity rainfall. In Experiment 3, CC value further increases to 0.69. Experiment 4 gives the highest CC value of 0.72.

[Table 2 is to be inserted about here]

[Figure 4 is to be inserted about here]

Overall, using as much meteorological variables as possible is conducive to improving the accuracy of downscaling rainfall. Among all meteorological variables, the geopotential height might be the most useful one. Considering both accuracy and

Peer-reviewed version available at Water 2019, 11, 977; doi:10.3390/w11050977

complexity of model, we suggest that the combination of geopotential height and total water vapor might be reasonable.

### 4.2 Precipitation Estimation Performance of different Methods

In this section, we compare the performance of different downscaling methods. Quantile mapping method, CNN networks, SVM, and ConvLSTM networks are used as downscaling methods for Experiment 5-8, respectively. Experiment 5 uses ERA-interim precipitation as predictor, and Experiment 6-8 use geopotential height at 500/700/850/925/1000hpa and total column water as predictors.

Metrics and scatter plots are shown in Table 3 and Figure 5, respectively. Experiment 5 use the traditional Quantile mapping method. Its improvement upon original ERA-interim precipitation is very limited. The CC value is only 0.54, and the RMSE value is 10.01mm/day. However, it is worth to mention that only in Experiment 5, performance in validation period is comparable to training period, which may be due to its simple model structure and fewer model parameters. While for other complex models, performance in training period evidently exceeds validation period, which reminds us that we should be cautious whether the overfitting phenomena would lead to bad results in operation. Experiment 6 is the same as Experiment 3 in section 4.1, and we will not repeat it here. The SVM is used as training method in Experiment 7, the CC, RB and RMSE values for which are 0.65, -5.05% and 7.91 mm/day. Although its evaluation indexes are comparable to Experiment 6, its scatter plots show that it could barely simulate the high intensity rainfall in the study area (although some study claims good performance of SVM in rainfall downscaling in other regions), which

Peer-reviewed version available at Water 2019, 11, 977; doi:10.3390/w11050977

greatly limits its availability in hydrological simulation. Alternatively, if we change to draw the area mean precipitation, as shown in figure 6, the SVM methods can give reasonable results. Performance of Experiment 8 (use ConvLSTM) gives the best evaluation indexes, with CC value of 0.74, RB value of 1.73% and RMSE value of 7.17 mm/day.

[Table 3 is to be inserted about here]

[Figure 5 is to be inserted about here]

[Figure 6 is to be inserted about here]

Overall, the performance of these downscaling methods gradually increases according to the order of Quantile mapping, SVM, CNN networks, and ConvLSTM networks.

### 4.3 Application of Methods in Precipitation Forecast

To further estimate the robustness of the network, in this part we applied the trained network in precipitation forecasting based on the S2S ECMWF hindcast products. For simplicity, only the geopotential height at 500/700/850/925/1000hpa and total water vapor content are used as the predictors in this section (Note that only geopotential is available in S2S ECMWF hindcasts, but it can be transferred to geopotential height by multiply by a constant value of g = 9.80665), and ConvLSTM networks is used as the training model. For the 11 ensemble models, we first obtain the adjusted precipitation using the outputs of each model respectively, and then calculate their average as the final result.

Figure 7 illustrate the predictive skill of S2S-ECMWF precipitation and the

Peer-reviewed version available at Water 2019, 11, 977; doi:10.3390/w11050977

adjusted precipitation in the validation period as a function of forecast lead time. We can see that the performance of the corrected precipitation always outperforms the direct output of S2S-ECMWF. For the Day 1 forecast, the correlation coefficient between the corrected precipitation and the observed precipitation is 0.58, slightly lower than the result of experimental 3 in section 4.1 (with CC of 0.69), but significantly better than the S2S-ECMWF precipitation (with CC of 0.36). The difference of the correlation coefficient is 0.22. The CC values decrease sharply with the increase of the lead time. For the Day 15 forecast, the CC value of the corrected precipitation and S2S ECMWF drop to 0.21 and 0.19 respectively, with a difference of only 0.02. In summary, the corrected precipitation is superior to the S2S-ECMWF throughout all the lead time, but the superiority gradually decrease with the increase of the lead time.

[Figure 7 is to be inserted about here]

For the original ECMWF precipitation prediction, the uncertainty mainly comes from two sources. The first is dynamical deviation from the true atmosphere evolution, and the second is parameterization error due to imperfect description of unresolved scale physical processes. Compare to the original ECMWF precipitation prediction, the downscaled precipitation can eliminate the parameterization error to some extent: for the one hand, it uses the observed precipitation, but the GCM is not calibrated locally; for the other hand, the straightforward ConvLSTM network uses a top-down parameterization paradigm, and might be more efficient when properly calibrated. When the lead time is short, both of two kinds of uncertainty sources cannot be ignored, thus performance of the ConvLSTM downscaling method presents a significant

Peer-reviewed version available at Water 2019, 11, 977; doi:10.3390/w11050977

advantage over S2S-ECMWF precipitation prediction. As forecast lead time extends, the first kind of error (dynamical deviation) gradually dominates due to the chaotic effect, and the superiority of the ConvLSTM downscaling method gradually disappears.

### 4.4 Application in Hydrological Simulation

After calibration and validation, the GBHM is run using the observed rainfall, the original ERA-interim rainfall and the downscaled rainfall (using geopotential height at 500/700/850/925/1000hpa and total column water as predictors, ConvLSTM networks as downscale method, i.e., Experiment 8 in 4.2) as input data, respectively. All simulations use the hydrological state at the end of 2010 as initial condition, which is obtained from continuous running GBHM using gauge rainfall as input.

Table 4 gives the evaluation indexes and figures 8 compares the streamflow simulation results using three sets of rainfall input at Hengshan station from 2011 to 2016, respectively. The simulation using observed rainfall data agrees well with the observed discharge, with NSE value of 0.82 and RB value of 7%. However, the original ERA-interim rainfall forced simulation is almost completely useless, with a NSE value of 0.06. It severely overestimates most flood peaks as shown in figure 8, and overestimate 24% of the total runoff. The simulation using the adjusted rainfall provides reasonable result, with NSE value of 0.64 and RB value of -10%. It is close to the observed rainfall forced simulation, and far better than the original ERA-interim rainfall forced simulation. Hydrological systems are non-linear systems, uncertainties in precipitation inputs may be magnified when it is transferred to runoff, which makes the correction of the raw predicted rainfall more necessary.

Peer-reviewed version available at Water 2019, 11, 977; doi:10.3390/w11050977

[Table 4 is to be inserted about here]

[Figure 8 is to be inserted about here]

## 5. Summary and conclusions

In this study, we proposed a new method for precipitation downscaling by combining CNN and LSTM, based on model resolved circulation dynamics. This method was tested for precipitation estimation or prediction in Xiangjiang River Basin in south China, which located in the East Asian monsoon region. The results show that this method has advantages over the traditional quantile mapping method or SVM based method. The downscaled rainfall was further evaluated through a distributed hydrological model. The major conclusions can be summarized as follows:

(1) 4 Experiments using different circulation predictors are designed to test the effectiveness of those predictors. Results show that using as much meteorological variables as possible is conducive to improving the rainfall estimation. While using mean sea level alone can only provide limited improvement. Among all meteorological variables, the geopotential height might be the most important one. Considering both accuracy and complexity of model, we suggest that the combination of geopotential height and total water vapor might be reasonable.

(2) Another 4 Experiments are designed to compare the different performance of Quantile Mapping method, SVM, CNN networks, and ConvLSTM networks. Precipitation estimated by ConvLSTM networks give the best performance (with highest correlation coefficient of 0.73), along with CNN, SVM and

Peer-reviewed version available at Water 2019, 11, 977; doi:10.3390/w11050977

Quantile Mapping method, with correlation coefficient of 0.69, 0.65, 0.54 respectively. We also found that method based on SVM could not predict those very high intensity rainfall.

(3) The trained ConvLSTM networks are applied to S2S-ECMWF hindcast datasets to further test its robustness. We find the corrected precipitation is superior to the original S2S-ECMWF precipitation all the time, but the superiority gradually decreased with the increase of the lead time. We think the improvement mainly comes from the use of observed data and the effective networks, which can reduce the parameterization error. But when the lead time extends, the parameterization error becomes subordinate. Thus the superiority of proposed method gradually fades away.

(4) Different rainfall inputs are fed into the distributed hydrological model. The original EAR-Interim rainfall shows little usage in hydrological simulation with a NSE of 0.06 and RB of 24%. While the corrected rainfall forced simulation improves the NSE to 0.64 and reduce RB to -10%, which is comparable to the simulation forced by observed rainfall. This further proves the value of the proposed method.

## References

Abadi M, Agarwal A, Barham P, et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems[J]. 2016.

Peer-reviewed version available at Water 2019, 11, 977; doi:10.3390/w11050977

529 Anandhi A, Srinivas VV, Nanjundiah RS, Nagesh Kumar D (2008) Downscaling precipitation to
530 river basin in India for IPCC SRES scenarios using support vector machine. Int J Climatol
531 28:401 – 420.
532 Anandhi A , Srinivas V V , Kumar D N , et al. Role of predictors in downscaling surface temperature
533 to river basin in India for IPCC SRES scenarios using support vector machine[J].
534 International Journal of Climatology, 2009, 29(4):583-603.
535 Baesens B , Viaene S , Gestel T V , et al. An empirical assessment of kernel type performance for
536 least squares support vector machine classifiers[C]. International Conference on
537 Knowledge-based Intelligent Engineering Systems & Allied Technologies. IEEE, 2002.
538 Bauer P, Thorpe A , Brunet G . The quiet revolution of numerical weather prediction [J]. Nature,
539 2015, 525(7567):47-55.
540 Benestad R. Novel methods for inferring future changes in extreme rainfall over Northern
541 Europe[J]. Climate Research, 2007, 34(3):195-210.
542 Benestad R. and J. Haugen, 2007: On complex extremes: Flood hazards and combined high spring-
543 time precipitation and temperature in Norway. Climatic Change, 85, 381–406,
544 doi:10.1007/s10584-007-9263-2.
545 Boé J., Terray L., Habets F., Martin E.. Statistical and dynamical downscaling of the Seine basin
546 climate for hydro-meteorological studies. International Journal of climatology, 2007, 27:
547 1643-1655.
548 Cavazos, T., 1999: Large-scale circulation anomalies conducive to extreme precipitation events and
549 derivation of daily rainfall in northeastern Mexico and southeastern Texas. J. Climate, 12,
550 1506–1523.

Peer-reviewed version available at Water 2019, 11, 977; doi:10.3390/w11050977

551 Christensen J H, Machenhauer B, Jones R G, et al. Validation of present-day regional climate
552 simulations over Europe: LAM simulations with observed boundary conditions[J]. Clim
553 Dyn, 1997, 13(7):489-506.
554 Cover, T.M., 1965. Geometrical and statistical properties of systems of linear inequalities with
555 applications in pattern recognition. IEEE Transactions on Electronic Computers EC-14,
556 326–334.
557 Dai, Y., W. Shangguan, Q. Duan, B. Liu, S. Fu, and G. Niu, 2013: Development of a China dataset
558 of Soil hydraulic parameters using Pedotransfer functions for Land surface modeling. *J.*
559 *Hydrometeor*, 14, 869–887. doi:10.1175/JHM-D-12-0149.1.
560 Dee D P, Uppala S M, Simmons A J, et al. The ERA‐Interim reanalysis: configuration and
561 performance of the data assimilation system[J]. Quarterly Journal of the Royal
562 Meteorological Society, 2011, 137(656):553-597.
563 Graves A . Generating Sequences With Recurrent Neural Networks[J]. Computer Science, 2013.
564 Ghosh S. SVM-PGSL coupled approach for statistical downscaling to predict rainfall from GCM
565 output[J]. Journal of Geophysical Research Atmospheres, 2010, 115(D22).
566 Guhathakurta P. Long lead monsoon rainfall prediction for meteorological sub-divisions of India
567 using deterministic artificial neural network model[J]. Meteorology & Atmospheric
568 Physics, 2008, 101(1-2):93-108.
569 Hanssen‐Bauer, I., C. Achberger, R. Benestad, D. Chen, and E. Forland (2005), Statistical
570 downscaling of climate scenarios over Scandinavia, Clim. Res., 29(3), 255‐268.
571 Haykin, S., 2003. Neural Networks: A comprehensive foundation. Fourth Indian Reprint, Pearson
572 Education, Singapore, pp. 842.

Peer-reviewed version available at Water 2019, 11, 977; doi:10.3390/w11050977

573 Hochreiter S, Schmidhuber J. Long Short-Term Memory[J]. Neural Computation, 1997, 9(8):1735-
574 1780.
575 Hope, P. K. (2006), Projected future changes in synoptic systems influencing southwest western
576 australia, Climate Dynamics, 26(7-8), 765–780.
577 Ioffe, S., and C. Szegedy (2015), Batch normalization: Accelerating deep network training by
578 reducing internal covariate shift, in International conference on machine learning, pp. 448–
579 456.
580 James W Taylor. A quantile regression neural network approach to estimating the conditional
581 density of multi-period returns. Journal of Forecasting, 19(4):299–311, 2000.
582 Kidson, J. W., and C. S. Thompson, 1998: A comparison of statistical and model-based downscaling
583 techniques for estimating local climate variations. J. Climate, 11, 735–753.
584 Kang, I. S., and Coauthors, 2002: Intercomparison of the climatological variations of Asian summer
585 monsoon precipitation simulated by 10 GCMs. Climate Dyn., 19, 383–395.
586 Kannan, S., and S. Ghosh (2013), A nonparametric kernel regression model for downscaling
587 multisite daily precipitation in the mahanadi basin, Water Resources Research, 49(3),
588 1360–1385.
589 Lecun Y , Bengio Y , Hinton G . Deep learning.[J]. Nature, 2015, 521(7553):436.
590 Li, Y., and I. Smith (2009), A statistical downscaling model for southern australia winter rainfall,
591 Journal of Climate, 22(5), 1142–1158.
592 Li, Z., D.W. Yang, B. Gao, Y. Jiao, Y. Hong, and T. Xu, 2015: Multiscale hydrologic applications
593 of the latest satellite precipitation products in the Yangtze River basin using a distributed

Peer-reviewed version available at Water 2019, 11, 977; doi:10.3390/w11050977

hydrologic model. *J. Hydrometeor.*, **16**, 407-426, doi: http://dx.doi.org/10.1175/JHM-D-14-0105.1..

Meehl, G. A., et al., Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, chap. Global Climate Projections, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2007.

Miao Q , Yang D , Yang H , et al. Establishing a rainfall threshold for flash flood warnings in China’s mountainous areas based on a distributed hydrological model[J]. Journal of Hydrology, 2016, 541(541):371-386.

Murphy, J., 1999: An evaluation of statistical and dynamical techniques for downscaling local climate. J. Climate, 12, 2256–2284.

Murphy, J., et al. (2000), Predictions of climate change over Europe using statistical and dynamical downscaling techniques, International Journal of Climatology, 20(5), 489–501.

Nijssen, B. and D.P. Lettenmaier, 2004: Effect of precipitation sampling error on simulated hydrological fluxes and states: anticipating the global precipitation measurement satellites. J. Geophys. Res., 109, D02103. doi:10.1029/2003JD003497.

Norton C W, Chu P S, Schroeder T A. Projecting changes in future heavy rainfall events for Oahu, Hawaii: A statistical downscaling approach[J]. Journal of Geophysical Research Atmospheres, 2011, 116(D17):-.

Pan, B., and Z. Cong (2016), Information analysis of catchment hydrologic patterns across temporal scales, Advances in Meteorology, 2016.

Peer-reviewed version available at Water 2019, 11, 977; doi:10.3390/w11050977

615 Pan, Baoxiang & Hsu, Kuolin & AghaKouchak, Amir & Sorooshian, Soroosh. (2019). Improving
616 Precipitation Estimation Using Convolutional Neural Network. Water Resources Research.
617 10.1029/2018WR024090.
618 Panofsky, H. A., and G. W. Brier (1968), Some Application of Statistics to Meteorology, 224 pp.,
619 Pa. State Univ., University Park, Pa
620 Penman, H. L., 1948: Natural evaporation from open water, bare soil and grass. Proc. of the Roy.
621 Soc. Lond. Ser. A. Math. Phys. Sci., 193, 1doi:10.1098/rspa.1948.0037.
622 Prudhomme, C., N. Reynard, and S. Crooks (2002), Downscaling of global climate models for flood
623 frequency analysis: Where are we now?, Hydrol. Processes, 16(6), 1137–1150,
624 doi:10.1002/hyp.1054.
625 Schoof, J. T., and S. Pryor (2001), Downscaling temperature and precipitation: A comparison of
626 regression-based methods and artificial neural networks, International Journal of
627 climatology, 21(7), 773–790.
628 Scikit-learn: Machine Learning in Python, Pedregosa *et al.*, JMLR 12, pp. 2825-2830, 2011.
629 Sehad M, Lazri M, Ameur S. Novel SVM-based technique to improve rainfall estimation over the
630 Mediterranean region (north of Algeria) using the multispectral MSG SEVIRI imagery[J].
631 Advances in Space Research, 2016, 59(5):1381-1394.
632 Shen Y, Xiong A, 2016: Validation and comparison of a new gauge-based precipitation analysis
633 over mainland China. Int. J. Climatol., 36, 252-265. doi: 10.1002/joc.4341.
634 Shi X, Chen Z, Hao W, et al. Convolutional LSTM Network: a machine learning approach for
635 precipitation nowcasting[C]// International Conference on Neural Information Processing
636 Systems. 2015.

Peer-reviewed version available at Water 2019, 11, 977; doi:10.3390/w11050977

637 Smola, A. J. (1996), Regression Estimation with Support Vector Learning Machines, Technische
638 Universitat Munchen, Munich, Germany.
639 Stephens G L , L'Ecuyer T , Forbes R , et al. Dreary state of precipitation in global models[J].
640 Journal of Geophysical Research Atmospheres, 2010, 115(D24):-.
641 Tapiador F J , Roca, Rémy, Genio A D , et al. Is precipitation a good metric for model performance?
642 [J]. Bulletin of the American Meteorological Society, 2018.
643 Tripathi, S., V. Srinivas, and R. S. Nanjundiah (2006), Downscaling of precipitation for climate
644 change scenarios: a support vector machine approach, Journal of hydrology, 330(3- 4),
645 621–640.
646 Vandal, T., E. Kodra, S. Ganguly, A. Michaelis, R. Nemani, and A. R. Ganguly (2017), Deepsd:
647 Generating high resolution climate change projections through single image super-
648 resolution, in Proceedings of the 23rd ACM SIGKDD International Conference on
649 Knowledge Discovery and Data Mining, pp. 1663–1672, ACM.
650 Vapnik, V.N., 1995. The Nature of Statistical Learning Theory. Springer Verlag, New York.
651 Vitart, F. (2004), Monthly forecasting at ecmwf, Monthly Weather Review, 132(12), 2761–2779.
652 Vitart, F., C. Ardilouze, A. Bonet, A. Brookshaw, M. Chen, C. Codorean, M. Déqué, L. Ferranti, E.
653 Fucile, M. Fuentes, et al. (2017), The subseasonal to seasonal (s2s) prediction project
654 database, Bulletin of the American Meteorological Society, 98(1), 163–173.
655 Wang, B., I. S. Kang, and J. Y. Lee, 2004: Ensemble simulation of Asian–Australian monsoon
656 variability by 11 AGCMs. J. Climate, 17, 803–818.
657 Wilby, R. L., and T. M. L. Wigley, 2000: Precipitation predictors for downscaling: Observed and
658 general circulation model relationships. Int. J. Climatol., 20, 641–661.

Peer-reviewed version available at Water 2019, 11, 977; doi:10.3390/w11050977

659 Yang, D W, T Koike, and H Tanizawa, 2004: Application of a distributed hydrological model and
660 weather radar observations for flood management in the upper Tone River of Japan.
661 *Hydrological Processes*, **18**, 3119-3132. doi:10.1002/hyp.5752.
662 Yang, D, S Herath, and K Musiake, 2001: Spatial resolution sensitivity of catchment
663 geomorphologic properties and the effect on hydrological simulation. *Hydrol. Process.*, **15**,
664 2085-2099. doi:10.1002/hyp.280.
665 Yang, D., S. Herath, and K. Musiake, 2002: A Hillslope-based hydrological model using catchment
666 area and width functions. *Hydrol. Sci. J.*, **47**, 49-65. doi:10.1080/02626660209492907.
667 Yang, D.W., B. Gao, Y. Jiao, H.M. Lei, Y.L. Zhang, H.B. Yang, and Z.T. Cong, 2015: A distributed
668 scheme developed for eco-hydrological modeling in the upper Heihe river. *Sci. China Earth*
669 *Sci.*, **58**, 36–45. doi:10.1007/s11430-014-5029-7.
670 Yang, D.W., S. Herath, and K Musiake, 1998: Development of a geomorphology-based
671 hydrological model for large catchments. *Proc. Hydraul. Eng.*, **42**, 169–174.
672 doi:10.2208/prohe.42.169.
673

Peer-reviewed version available at Water 2019, 11, 977; doi:10.3390/w11050977

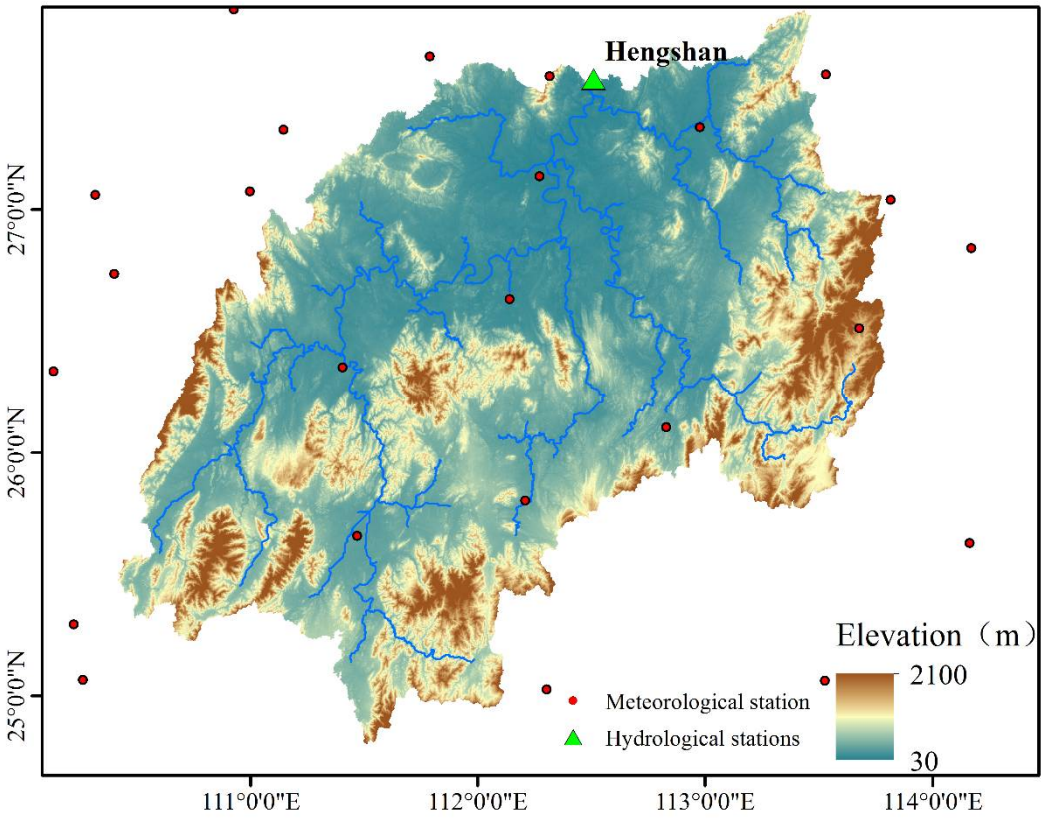



Figure 1. Study basin and locations of the hydrological station and meteorological stations

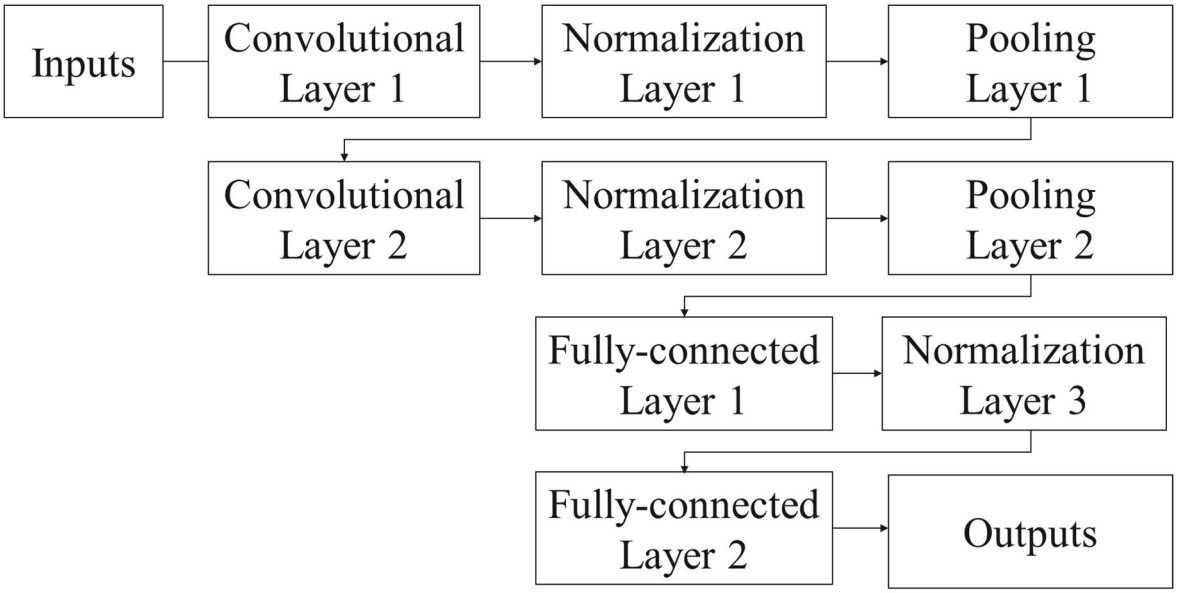
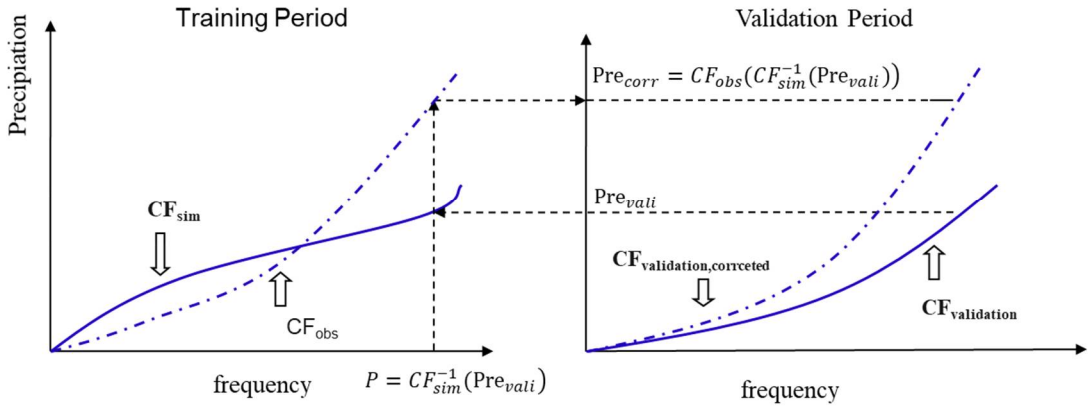


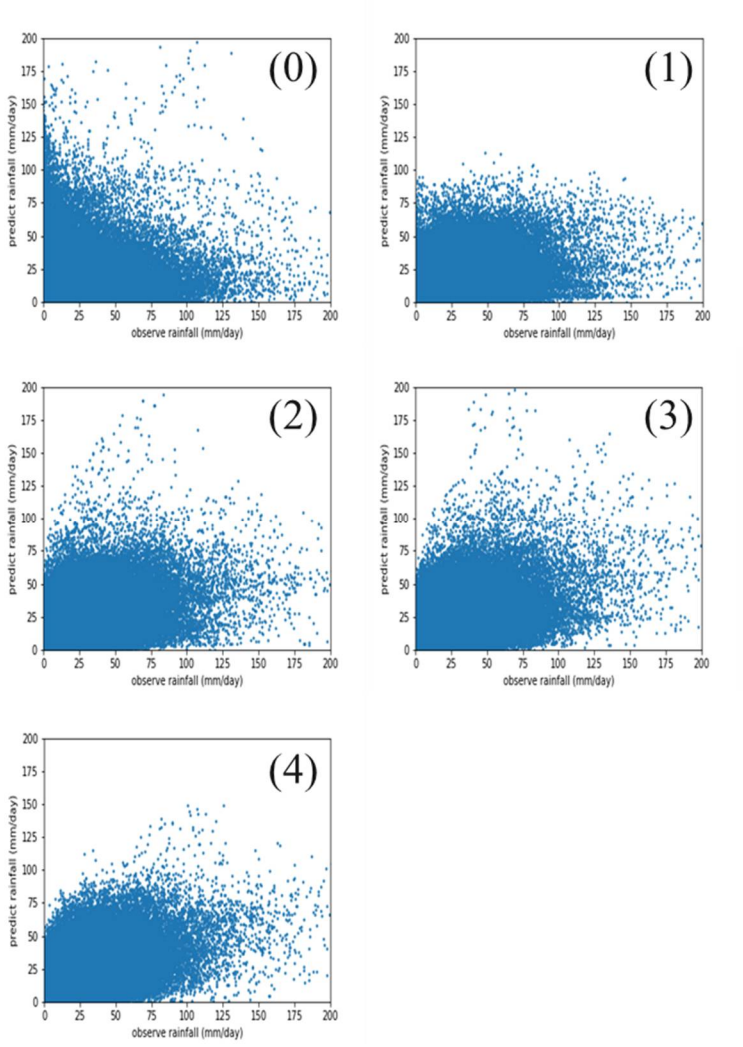

Figure 2. Architecture of the CNN networks used in this study

Peer-reviewed version available at Water 2019, 11, 977; doi:10.3390/w11050977

683
684



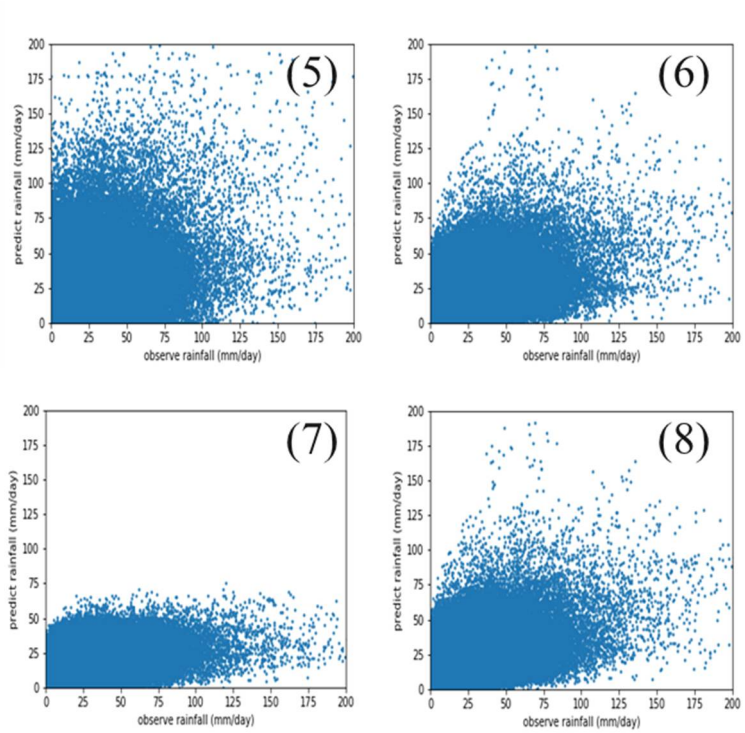

685
686 Figure 3 Illustration of the quantile mapping method
687



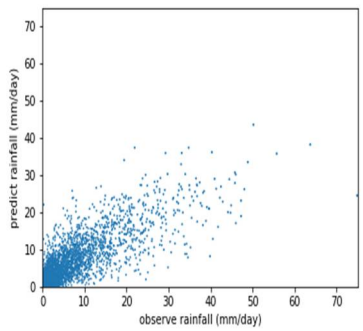

688
689 Figure 4 Scatterplots of estimated precipitation using different predictors against observed
690 precipitation (daily scales): (0) Original ERA-Interim precipitation, (1) Sea level mean pressure,

Peer-reviewed version available at Water 2019, 11, 977; doi:10.3390/w11050977

691 (2) Geopotential Height, (3) Geopotential Height filed and total column water, and (4) all
692 circulation variables described in section 2.2




693
694 Figure 5 Scatterplots of estimated precipitation using different downscaling methods against
695 observed precipitation (daily scales): (5) Quantile mapping, (6) CNN networks, (7) SVM, and
696 (8) ConvLSTM networks
697
698
699




700
701 Figure 6 Similar to figure 5 (7), but for area mean precipitation
702
703

Peer-reviewed version available at Water 2019, 11, 977; doi:10.3390/w11050977

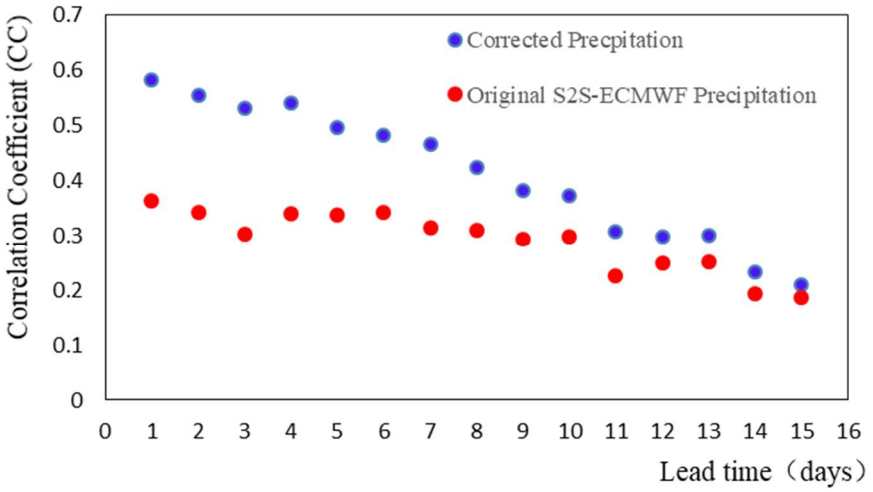



Figure 7 Predictive skill of S2S ECMWF precipitation and the corrected precipitation in the validation period as function of forecast lead time

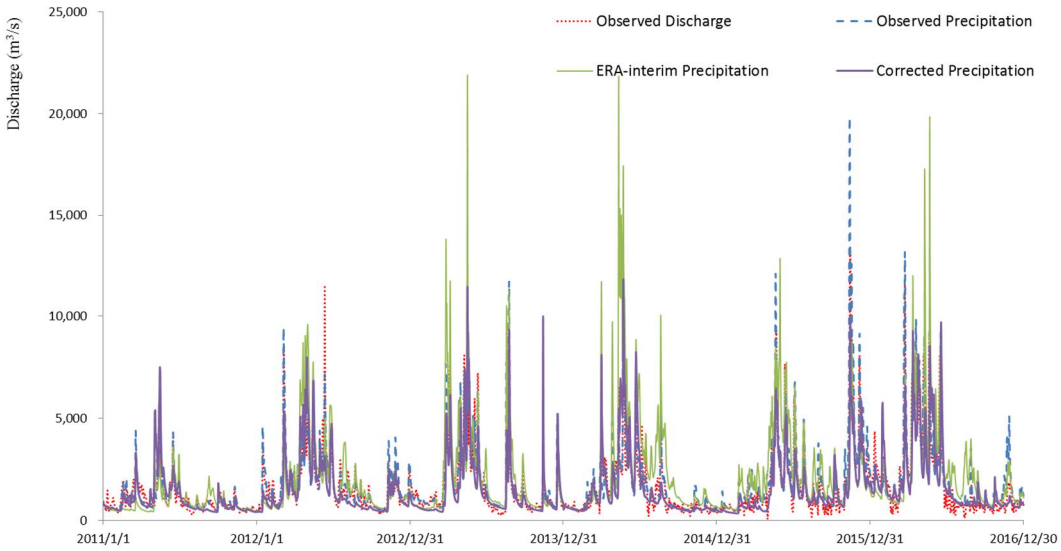



Figure 8 Comparison of the hydrographs at Hengshan station among the observation (reference), observed precipitation-driven simulation, ERA-interim precipitation -driven simulation and corrected precipitation-driven simulation

Peer-reviewed version available at Water 2019, 11, 977; doi:10.3390/w11050977

716 Table 1. GBHM performance of daily discharge simulation during the calibration and validation

717 periods at Hengshan Station

| Period | NSE | RB |
|---|---|---|
| Calibration period | 0.89 | 0.04 |
| Validation period | 0.88 | 0.02 |

718 Table 2. Performance of downscaled precipitation using different predictors

| Exp | Predictors | Training Period (1979-2002) | | | Validation Period (2003-2016) | | |
|---|---|---|---|---|---|---|---|
| | | CC | RB (%) | RMSE (mm/day) | CC | RB (%) | RMSE (mm/day) |
| 0 | p | 0.31 | 16.53 | 11.16 | 0.29 | 12.75 | 11.48 |
| 1 | msl* | 0.69 | 4.73 | 7.42 | 0.54 | 5.53 | 8.86 |
| 2 | gp* | 0.79 | 4.46 | 6.2 | 0.66 | 4.08 | 7.93 |
| 3 | gp,tcw* | 0.85 | 2.27 | 5.36 | 0.69 | 1.87 | 7.54 |
| 4 | gp,tcw,tem,uw,vw, cape,vv,pv* | 0.94 | 3.08 | 3.4 | 0.72 | 6.92 | 7.28 |

719 Note: p represents original ERA-interim precipitation; msl represents mean sea level pressure; gp
720 represents geopotential height; tcw represents total column water; tem represents air temperature;
721 cape represents convective available potential energy; uw represents u wind component; vw
722 represents v wind component; vv represents vertical velocity; pv represents potential vorticity.

723 Table 3. Performance of downscaled precipitation using different downscaling methods

| Exp | Method | Training Period (1979-2002) | | | Validation Period (2003-2016) | | |
|---|---|---|---|---|---|---|---|
| | | CC | RB (%) | RMSE (mm/day) | CC | RB (%) | RMSE (mm/day) |
| 0 | ERA-interim | 0.31 | 16.53 | 11.16 | 0.29 | 12.75 | 11.48 |
| 5 | QM | 0.54 | 0.21 | 9.77 | 0.54 | -2.82 | 10.01 |
| 6 | CNN | 0.85 | 2.27 | 5.36 | 0.69 | 1.87 | 7.54 |
| 7 | SVM | 0.70 | 2.31 | 7.33 | 0.65 | -5.05 | 7.91 |
| 8 | ConvLSTM | 0.85 | 2.27 | 5.36 | 0.73 | 1.73 | 7.17 |

724 Table 4. Performance of simulations with different precipitation forcing

| Precipitation Inputs | NSE | RB(%) |
|---|---|---|
| Observed Precipitation | 0.82 | 7 |
| Original ERA-interim Precipitation | 0.06 | 24 |
| Corrected Precipitation | 0.64 | -10 |

725