Nierji reservoir flood forecasting based on a Data-Based Mechanistic methodology

Guozhen Wei[a,*], Wlodek Tych[b], Keith Beven[b], Bin He[a], Fanggui Ning[c] and Huicheng Zhou[a]

[a] *School of Hydraulic Engineering, Dalian University of Technology, Dalian, 116024,P.R.China*

[b] *Lancaster Environment Centre, Lancaster University, Lancaster LA1 4YQ, UK*

[c] *Songliao Water Resources Commission,Changchun,130021, P.R.China*

[*] *Corresponding author at: School of Hydraulic Engineering, Dalian University of Technology, Dalian, 116024, P.R.China*

Abstract

The Nierji Basin, in the north-east of China, is one of the most important basins in the joint operation of the entire Songhua River, containing a major reservoir used for flood control. It is necessary to forecast the flow of the basin during periods of flood accurately and with the maximum lead time possible. This paper presents a flood forecasting system, using the Data Based Mechanistic (DBM) modeling approach and Kalman Filter data assimilation for flood forecasting in the data limited Nierji Reservoir Basin (NIRB). Examples are given of the application of the DBM methodology using both single input (rainfall or upstream flow) and multiple input (rainfalls and upstream flow) to forecast the downstream discharge for different sub-basins. Model identification uses the simplified recursive instrumental variable (SRIV) algorithm, which is robust to noise in the observation data. The application is novel in its use of stochastic optimisation to define rain gauge weights and identify the power law

nonlinearity. It is also the first application of the DBM methodology to flood forecasting in China. Using the methodology allows the forecasting with lead times of 1-day, 2-day, 3-day, 4-day, 5-day with 98%, 97%, 96%, 96% and 93% forecast coefficient of determination respectively, which is sufficient for the regulation of the reservoirs in the basin.

Key words: flood forecasting, DBM, Kalman filter, SDP, large basin

1 Introduction

Flood forecasting is a particularly interesting and challenging application of hydrological theory. It is interesting because of the considerable operational importance in providing timely and accurate forecasts with sufficient lead time to facilitate decision making during flood events that might have considerable impacts on people and damage to infrastructure. It is also challenging because it is just during such flood events that we expect to have the greatest uncertainties associated with both inputs and flow data, and with the representation of hydrological processes. Unlike hydrological simulation, however, data assimilation can be used in the flood forecasting to constrain the forecast uncertainties and improve forecast accuracy. This is advantageous when we expect the next event to be different in both form and data uncertainties from those in the past (that might be used to calibrate a model). These specific aspects of flood forecasting have led to a variety of operational approaches from the use of conceptual models (e.g. Franz et al. 2003; Schaake et al., 2007); neural network models (e.g. Han et al., 2007; Chang et al., 2007); simple storage-outflow models based directly on data (e.g. Lambert 1972), and linear transfer function models using input transforms such as the Data-Based Mechanistic (DBM) methodology used here (the form of Hammerstein model of Young, 2002; Young et al., 2014). Forecasting methods using ensembles of inputs from numerical weather prediction systems that would allow forecast lead times longer than the natural time delay of a basin have also been reviewed by Cloke and Pappenberger (2009).

Currently, many different flood forecasting models including lumped conceptual models, semi-distributed models, and distributed models are used in China. The most popular conceptual models are Xinanjiang Model and Dahuofang Model. The Xinanjiang model developed by Zhao (1984) is suitable for both humid and semi-humid regions, and has been widely used in Southern China (Zhao, 1992; Cheng et al., 2006; Yao et al., 2014; Lu and Li, 2015). The Dahuofang model (Wang, 1996; Wang et al., 2012), developed by the Dalian University of Technology and the Office of State Flood Control and Drought Relief, is more effective for arid areas. Semi-distributed models have also been commonly applied. In particular, TOPMODEL, developed by Beven and Kirkby (1979), has been applied in many basins in China, including arid areas (Peng et al., 2017), humid areas (Xiao et al., 2017) and semi-humid areas (Li et al., 2015). With the advent of the information age, more and more distributed models have been developed and used in the country. Raster modelling concepts have been introduced into the Xinanjiang Model, resulting in the Grid-Xinanjiang distributed Model with good results (Zhi-Jia et al., 2007; Yao et al., 2009; Yao Ji et al., 2012; Yao et al., 2012; Yao et al., 2014). Many other models have also been used and modified for Chinese basins, such as VIC (Guo et al., 2009; Xue et al., 2016; Li et al., 2016a, b); TOPKAPI (Liu, 2004; Liu, 2004; Liu et al., 2005; Liu et al., 2016; Liu et al., 2016); and HEC-HMS (Oleyiblo and Zhi-Jia, 2010). Other than the GIS and DEM data, distributed models require spatial fields of input variables (e.g. precipitation, radiation, and surface air temperature) which means that they are less suitable for areas with sparse *in situ* networks, such as the Nierji Basin that is the subject of this paper.

Many existing models have been tested, but without achieving high accuracy in the study area (e.g. Liu et al., 2012; Wei et al., 2015). This was a reason to test the application of the DBM methodology in this type of data-sparse application. DBM models have a number of advantages in that they can be derived directly from the available data (even when only a small number of events are available); they have a physically mechanistic interpretation; and they are readily implemented within a data assimilation framework.

One feature of flood forecasting relative to hydrological simulation is that the methods used, such as DBM, are not required to maintain mass balance. In fact, given the uncertainties in the hydrological data associated with extreme events it might be disadvantageous to impose mass balance constraints. That is also why data assimilation can be so useful in forecasting. A number of data assimilation strategies have also been proposed from direct insertion of latest discharge values (as in the ISO model of Lambert, 1972); adaptive gain methods that can be applied to either model outputs or to a statistical model of residuals from a deterministic model (Smith et al., 2012); and ensemble Kalman filter and particle filter methods (Moradkhani et al., 2005;

Weerts and El Sarafy, 2006). Most recent applications of these methods involve estimation of forecast uncertainties. Other methods for uncertainty estimation used in flood forecasting include neural networks and quantile regression of model residuals (Brath et al., 2002; Weerts et al., 2011) and the Bayesian Forecasting System of Krzysztofowicz (2002, Reggiani and Weerts, 2008; Herr and Krzysztofowicz, 2010). Many of these methods are reviewed in Sene et al. (2014).

The methodology adopted in this study is the DBM forecasting system developed at Lancaster University by Peter Young and his colleagues (Lees et al., 1994; Young, 2002). The DBM methodology has been applied to a number of UK catchments (Lees et al., 1994; Romanowicz et al., 2006, 2008; Leedal et al. 2010; Smith et al., 2013a, 2014) and elsewhere (Alfieri et al., 2011; Smith et al. 2013b). As the name suggests, DBM models are derived from data, using a combination of linear transfer functions and state dependent parameter estimation to identify appropriate nonlinear transforms of the input variables (Young and Beven, 1994; Beven et al., 2011). The models can be derived from relatively few events, but this means that they will be necessarily approximate when applied to extreme events that may be outside the range of the calibration data. Thus, the calibrated models are used with a simple data assimilation algorithm to help improve the forecasts in real time. Post-event analysis of the changing gains during an event can then be used to provide information about the effective nonlinearities for a new event. The model itself can then be updated as more information from extreme events (or major catchment changes such as the building of reservoirs) becomes available.

The accuracy and information content of the input data are important in flood forecasting applications of the DBM methodology. What is required is an estimate of the inputs from the available raingauges that provides the most effective forecast. Many methods for deriving the weights of rain gauges have been developed such as methods based on spatial statistics (Griffith,1993); thin plate smoothing splines (Hutchinson, 1998); Thiessen polygons method (e.g. Thiessen 1911; Panigrahy et al., 2005); and a variety of distance weighted methods (e.g. Yang et al., 2003). However, these methods will be less useful in basins like the Nierji that is the subject of this study with an extremely uneven spatial distribution of rain gauges and high rainfall spatial variability. Therefore, we use a stochastic optimisation approach which is simple and effective in determining the relative weights of rain gauges to optimise the forecasting performance as part of the model calibration process.

The DBM methodology has previously been applied in China in modelling changes in Leaf Area Index (e.g. Chen et al. 2012; Guo et al., 2014; Zhou et al., 2017). The present study is, however, the first application of the DBM methodology with the implementation of a full Kalman filter and a stochastic optimisation approach to finding raingauge weights in the identification of the input nonlinearity, to flood forecasting in China.

2 Study Area and Datasets

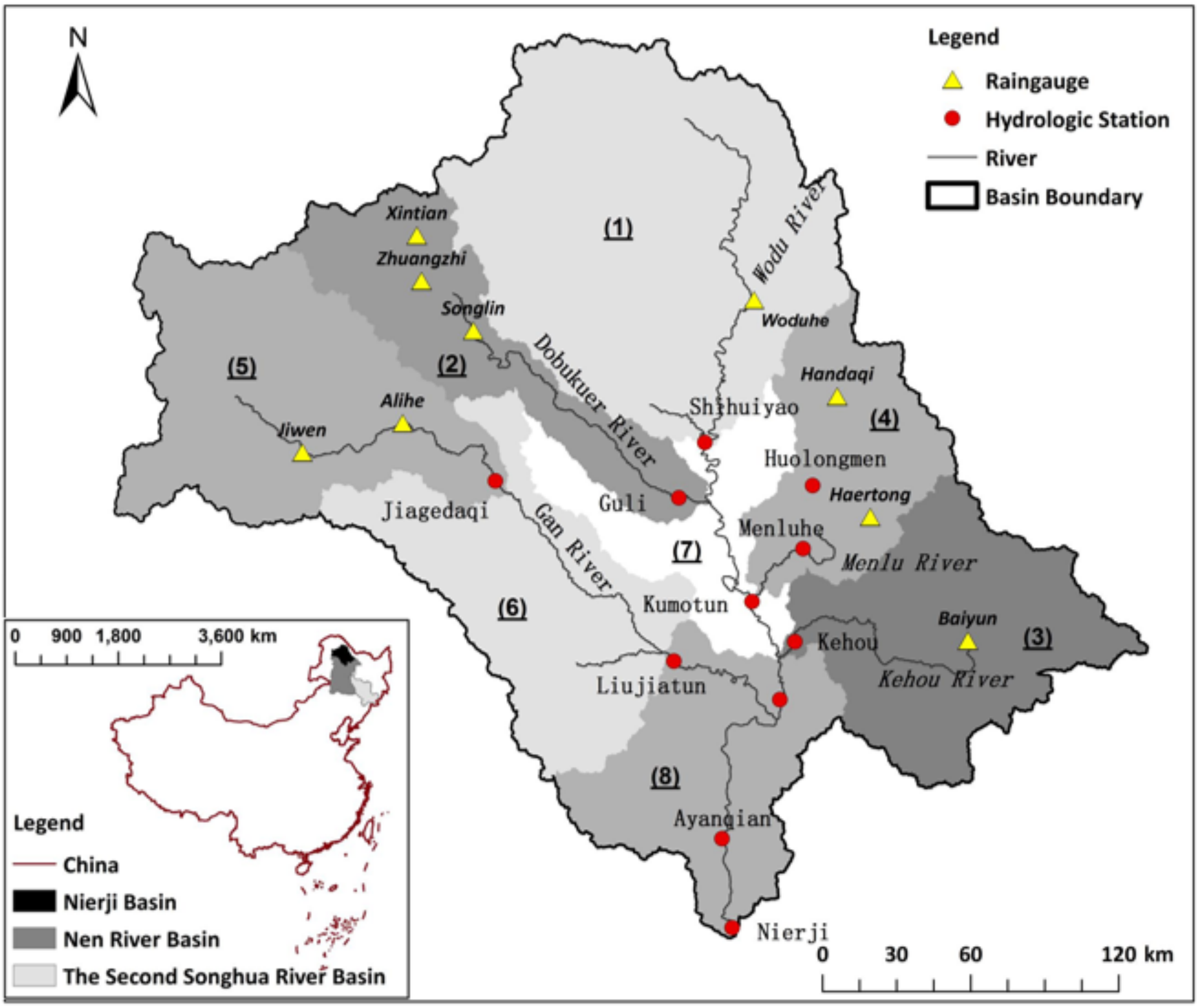

Figure 1. The Nierji Basin upstream of the Nierji Reservoir: raingauge and discharge gauging sites and the location in China ((1) (2) … (8) are all the sub-basins' number (see detail in section 2.2).)

2.1 Study Area

The Nierji Reservoir Basin (NIRB) that is the subject of this study is located in the larger Nen River Basin (NRB). It is the upstream basin above the Nierji Reservoir and spans the Inner Mongolia and Heilongjiang Provinces. The area of the basin is 66382 km^2, accounting for 22.35% of the NRB (Figure 1). The basin originates in the Dailinghuli Mountains, Daxinganling, where it goes through Nen River County, and enters Nehe City and Nierji Town from north to south, with a length of 782 km. Left bank main tributaries include the Wodu River, the Menlu River and the Kehou River, while the right bank main tributaries include the Dobukuer River and the Gan River. In the section from the source to Kumotun the valley bottom is narrow with a width of 1 km, while the section below Kumotun in the middle reaches has a 5 km-11 km wide valley bottom. The Nen River and the Second Songhua River flow into the Main Songhua River, which flows through the capital city of Heilongjiang Province, Haerbin,

The average annual runoff of this basin is $104.7 \times 10^9\ m^3$, accounting for 45.7% of the flow from the whole NRB. The average annual precipitation in the basin is 400-500 mm, with more in the upper reaches than in the lower reaches, and more in the mountainous areas than in the flat areas. The basin belongs to the north temperate monsoon climate area with a long, cold and dry winter, hot and rainy summer, dry and windy spring, and rapid cooling short autumn. As shown in Figure 1 the existing hydrological station network over the basin is unevenly distributed. In some areas, there are few or no stations (rain gauges and discharge stations) such as the Gan River tributary and upstream of Shihuiyao Station.

The outlet of the basin is the Nierji reservoir, which is located near Nierji Town, 32 km downstream of the Ayanqian hydrological station. It is a large reservoir that mainly provides flood control, and storage for urban, industrial and agricultural water supplies. Nierji Reservoir is an important flood control structure for the Nen River Basin with a total storage capacity of $86.1 \times 10^9\ m^3$. The limiting level for flood control is 213.37 m, with a static maximum water storage level of 218.15 m, while the normal reservoir water level is at 216.00 m.

It is known that summer rains in the NRB can be frequent and heavy. There have been numerous rainfall events that have caused severe floods and serious floods in downstream cities in the NRB such as Qiqihaer and Fulaerji. How to use existing engineering to control the floods is a significant management problem. Thus, accurate forecasting to control the Nierji Reservoir, which is one of the three most important control structures in the whole Songhua River Basin, would improve the utilization of the reservoirs to achieve optimal flood reduction in the areas at risk in downstream cities including the capital city Haerbin.

2.2 Data Sets

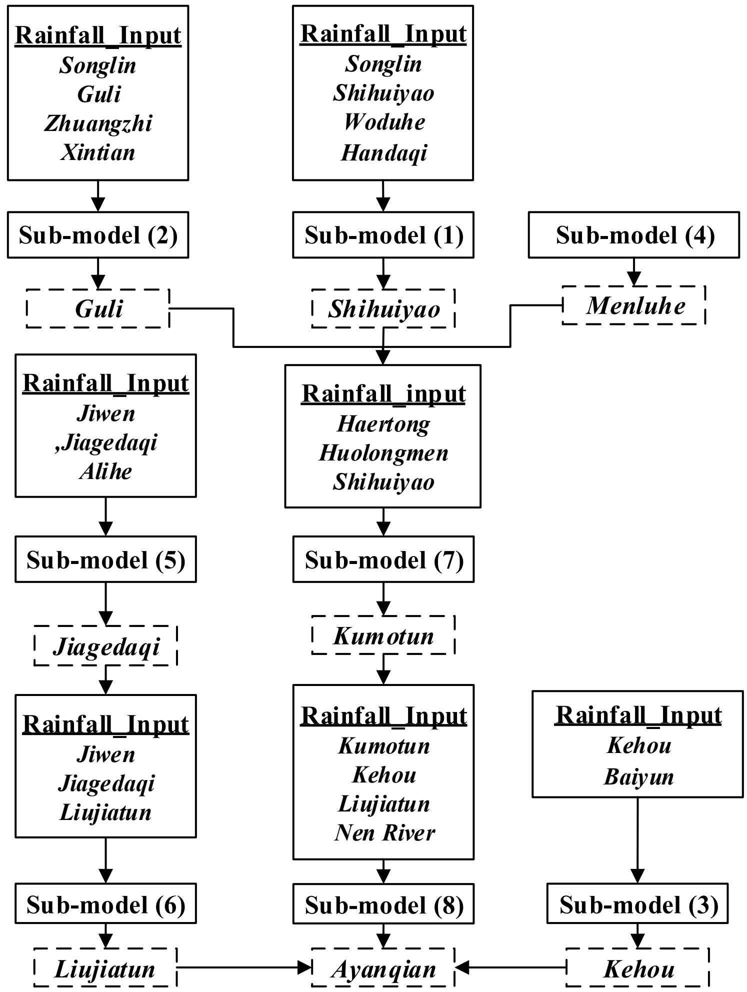

Figure 2. The relationships among the sub-models (the solid box represents the name of sub-basin and the corresponding rainfall gauges, (1) (2) … (8) are the sub-basin number; the dashed box represents the outlet name of sub-model)

According to the discharge stations at Shuihuiyao, Guli, Kumotun, Jiagedaqi, Liujiatun Kehou, Menlu and Ayanqian, the flood forecasting system has been divided into 8 sub-models as follows (Figure 2):

(1) using rainfall (Shihuiyao,Woduhe,Songlin and Handaqi) to forecast the discharge at Shihuiyao gauge;

(2) using rainfall (Songlin,Guli,Zhuangzhi and Xintian) to forecast the discharge at Guli gauge;

(3) using rainfall (Kehou and Baiyun) to forecast the discharge at Kehou gauge;

(4) For this particular sub-model, no observed flow data are available for the Menlu sub-basin, so the Kehou station is used to represent Menlu, scaled by the difference is area. The two stations are not only adjacent but also similar in size, and comparing the rainfall data of these two sub-basins suggests that they are similar in response even in large flood events.

(5) using rainfall (Jiwen,Jiagedaqi and Alihe) to forecast the discharge at Jiagedaqi gauge;

(6) using rainfall (Jiwen,Jiagedaqi and Liujiatun) and the forecasting discharge at Jiagedaqi gauges to forecast the discharge at Liujiatun;

(7) using rainfall (Haertong,Huolongmen and Shihuiyao) and the forecasted discharge of Shihuiyao, Menluhe and Guli gauges to forecast the discharge at Kumotun gauge;

(8) using rainfall (Kumotun,Kehou,Liujiatun and Nen River) and the forecasting discharge at Kumotun, Liujiatun and Kehou gauges to forecast the discharge at Ayanqian, which is used to represent the input discharge of Nierji Reservoir.

3 Methodology

The reason why we chose DBM for flood forecasting in the complex NIRB is that it allows the system to be represented with few parameters to describe the relationship between rainfall and flow, or for flow routing from upstream to downstream stations. The DBM approach allows the model structure to be defined by the available data, including both a linear transfer function and nonlinear input transform if required. The framework of the DBM forecasting system is illustrated in Figure 3.

Figure 3 the structure of forecasting algorithm. (See text for explanation of abbreviations)

Within the methodology, the nonlinear transform box can then be used to apply the form of any input nonlinearity required. Here we use a simple form of stochastic optimization to identify the best weights on rain gauges and parameter of the nonlinear input transform for each sub-basin model within the forecasting process (see below). This proved to be the best way of defining an optimal effective input for forecasting. The Simplified Recursive Instrumental Variable (SRIV) algorithm is used to identify the transfer function for the DBM model. SRIV is fast, robust to data errors, and just needs a few iterations to converge (see Young, 2011). For updating the forecasts in real time, the transfer function has been put into a data assimilation strategy to improve accuracy and constrain uncertainty. The Kalman filter has been chosen as a data assimilation strategy here because it is not assumed that the uncertainty on the transfer function parameters and forecast error is constant, but parameter vectors and associated covariance matrix are continuously updated. All of the DBM methods used here have been implemented in the CAPTAIN Toolbox for Matlab (Taylor et al, 2007).

3.1 Method of estimation of the raingauge weights and input nonlinear transform

Due to the extremely uneven spatial distribution of rain gauges and rainfall spatial variability, traditional Thiessen polygons and rainfall averaging methods cannot be used for observed rainfall in the studied basins. To define the best available forecasting model, therefore, different weights on the available rainfall stations have been considered in model calibration for each sub-model in the NIRB. A set of samples were formed by weighting each rainfall station that might contribute to each rainfall-runoff sub-model. In each case 1000 sets of sample weights are randomly selected constrained to a total of 100% by uniform random sampling of approach. This is a form of stochastic optimization of the rain gauge weights, where the normal procedure would have been to use Thiessen Polygons. Forecasting performance with weights chosen in this way compares favorably with the Thiessen weights approach and a simple average.

As is well known, input-output relationships in hydrology are non-linear in many situations. How to identify the non-linearity between the input and the effective input is the first step of DBM and is very necessary to the accuracy of the whole model. State Dependent Parameter estimation (SDP) is a way of identifying the nature of the nonlinear transform required in the DBM modelling methodology based on recursive estimation of the gains on an initial estimate of the transfer function (see, for example, Young and Beven, 1994; Beven et al., 2011). Different methods can be used to represent the form of that nonlinearity, including a power law, radial basis functions, piecewise cubic hermite data interpolation and so on (e.g. Beven et al., 2011). In this paper, a power law function of the current discharge, which has been suggested by past SDP identifications (as in Young and Beven 1994), has been adopted to transform the observed input to an effective input because it is simple to use and has a reasonable physical explanation in that the current discharge can be taken as an index of the wetness of the catchment. The power law function has the simple form:

$$\breve{\boldsymbol{P}}(\boldsymbol{i})=\boldsymbol{P}(\boldsymbol{i}) \times \boldsymbol{Q}^{\beta}(\boldsymbol{i}-\boldsymbol{\delta}) \quad (1)$$

Where $\breve{\boldsymbol{P}}(\boldsymbol{i})$ is the effective input at the $\boldsymbol{i}^{th}$ time step, $\boldsymbol{P}(\boldsymbol{i})$ is the observed input at the $\boldsymbol{i}^{th}$ time step, δ is the pure time delay between the observed output and the effective input and β is the power law parameter. In this paper it has been found that the channel flow routing model components required only linear transfer functions. $\boldsymbol{P}(\boldsymbol{i})$ is the observed rainfall and $\breve{\boldsymbol{P}}(\boldsymbol{i})$ is the effective rainfall input to the transfer function. The nonlinearity is only considered in the rainfall-runoff model components. Identification of the power value β in Equation (1) was achieved by uniform random sampling of 100 values between 0 and 1, and using those values to generate model outputs with an initial estimate of the transfer function. The optimal rain gauge weights and power law coefficient values in the DBM models for each sub-basin were then found by evaluation of the Young Information Criterion (YIC) and coefficient of determination (see Section 4 below).

3.2 Fitting the transfer function

After defining the nonlinear transform between the input and effective input, this section will introduce how to fit the transfer function. The linear transfer function can be described as follows:

$$\boldsymbol{Q}(i)=\frac{\boldsymbol{B_1}(\boldsymbol{z}^{-1})}{\boldsymbol{A}(\boldsymbol{z}^{-1})}\widetilde{\boldsymbol{P_1}}(i-\delta(1))+\cdots+\frac{\boldsymbol{B_N}(\boldsymbol{z}^{-1})}{\boldsymbol{A}(\boldsymbol{z}^{-1})}\widetilde{\boldsymbol{P_N}}(i-\delta(\mathrm{N}))+\boldsymbol{\varepsilon_i} \quad (2)$$

$$\boldsymbol{B}_k(z^{-1}) = b_k(0) + b_k(1)z^{-1} + \cdots + b_k(m(k))z^{-m(k)} \quad (k=1, 2, \ldots, \mathrm{N}) \tag{3}$$

$$\boldsymbol{A}(z^{-1}) = 1 + a(1)z^{-1} + \cdots + a(n)z^{-n} \tag{4}$$

$$z^{-1}\breve{\boldsymbol{P}}_k(i) = \breve{\boldsymbol{P}}_k(i-1) \tag{5}$$

Where $\boldsymbol{Q}(i)$is the discharge at the $\boldsymbol{i}^{\text{th}}$ time step, *k* is the index of an input, $\breve{\boldsymbol{P}}_K(i-\delta(k))$ is the k^{th} effective input at the $(i-\delta(k))$th sample, a(1), a(2)… a(n) are the transfer function denominator parameters, b(0), b(1)… b(m(*k*)) are the numerator parameters, m(k) represents the numerator order of the k^{th} input, δ (k) means the lead time of the k^{th} input, *N* is the number of effective inputs, and z^{-1} is the backward shift operator. Sub-models (1),(2),(3),(4) and (5) (see section 4) have only rainfall input data so that they are all single input models (*N*=1). Sub-models (6), (7) and (8) all include both rainfall and upstream flow inputs so that they are defined as multiple input models (*N*>1). The parameterisations of single input and multiple inputs models are different and so they are summarised and discussed separately in section 4 for clarity.

Stability of the transfer parameter estimates can be examined by examining plots of their variation in the time step by time step recursive estimation, while the physical acceptability of the transfer function can be assessed by plotting the response to a unit effective input. A variance-covariance matrix of the parameter estimates is also produced. The speed of estimation allows many different transfer functions to be evaluated. Model choice is based on the Young Information Criterion (YIC) that is a combination of model fit, and parameter uncertainty.

$$\mathbf{YIC} = \mathbf{log}_e \frac{\sigma^2_{residuals}}{\sigma^2_{obs}} + \mathbf{log}_e\{\boldsymbol{NEVN}\} \quad (6)$$

Where $\sigma^2_{\boldsymbol{residuals}}$ is the variance of the model residuals, $\sigma^2_{\boldsymbol{obs}}$ is the variance of the observed flow, and NEVN is the normalized error variance norm. The first term is similar to the coefficient of determination and is a measure the feasibility of the identified model. The term will become more negative with the decease of $\sigma^2_{\boldsymbol{residuals}}$. The second term is used to penalize the degree of over-parameterization. Normally, with an increase in model complexity, the dynamics of the system could be described more accurately. However, if the model is over-parameterized, this increase is associated with the increase of uncertainty in the parameter estimates and consequent rapid rise in the YIC. Results are also presented in terms of the forecast coefficient of determination for a lead time *t*, defined as

$$\boldsymbol{R_t^2} = 1 - \frac{\boldsymbol{\sigma^2_{residuals}}}{\boldsymbol{\sigma^2_{obs}}}$$

This is identical to the Nash-Sutcliffe efficiency measure often used to evaluate simulation model results when used the calibration process, but is also used later to assess the forecasting results at different lead times with the forecast residuals with and without data assimilation.

3.3 Kalman Filter

Data assimilation in real time is important in forecasting, where this is possible. The Kalman filter has been chosen as the data assimilation methodology in this application. The Kalman filter was developed by Kalman (1960) to define a way of updating model parameters and uncertainty as new observations become available. The standard Kalman Filter is suitable for linear systems, so it can be used in the forecasting system in conjunction with the transfer function identified for each model component in the catchment, once any nonlinear transform has been applied to the inputs. Equation (2) can be converted to a state-space equation as follows:

$$\boldsymbol{x}(\boldsymbol{i}) = \mathbf{F}\boldsymbol{x}(\boldsymbol{i}-\mathbf{1}) + \boldsymbol{G_1}\widetilde{\boldsymbol{P_1}}(\boldsymbol{i}-\boldsymbol{\delta}(1)) + \cdots + \boldsymbol{G}_{\mathrm{N}}\widetilde{\boldsymbol{P}_{\mathrm{N}}}\big(\boldsymbol{i}-\boldsymbol{\delta}(\mathrm{N})\big) + \boldsymbol{D\eta}(\boldsymbol{i}-\mathbf{1}) \tag{7}$$

$$\boldsymbol{y}(\mathbf{i}) = \boldsymbol{Cx}(\boldsymbol{i}) + \boldsymbol{\xi}(\boldsymbol{i}) \tag{8}$$

Where $\boldsymbol{X}(\boldsymbol{i})$ is the state vector; $\boldsymbol{y}(\mathbf{i})$ is the vector of observations; $\mathbf{F}$, $\mathbf{G}$, $\mathbf{C}$ and $\mathbf{D}$ are the model matrices which are derived from Equation (2); $\widetilde{\boldsymbol{P_k}}(\boldsymbol{i})$ is the effective inputs, *k*=1,2,…, N, N is the number of input; $\boldsymbol{\delta}(k)$ is the lead time of the k^{th} input, $\boldsymbol{G}_k\widetilde{\boldsymbol{P}_k}\big(\boldsymbol{i}-\boldsymbol{\delta}(k)\big)$ is the term to allow the $\widetilde{\boldsymbol{P}_k}(\boldsymbol{i}-\boldsymbol{\delta}(k))$ to affect the output. The variables $\boldsymbol{\xi}(\boldsymbol{i})$ and $\boldsymbol{\eta}(\boldsymbol{i})$ are assumed to follow independent Gaussian distributions with zero mean and time-variable covariance matrices $\boldsymbol{H}(\boldsymbol{i})$ and $\boldsymbol{R}(\boldsymbol{i})$ respectively. $\mathbf{P}(\mathbf{i})$ is the error covariance matrix of the state vector $\boldsymbol{x}(\boldsymbol{i})$. The nonlinear power identified in model calibration is assumed constant; the other model parameters are included in the Kalman filter.

Equations (7) and (8) are implemented at each time step as follows:

Prediction:

$$\boldsymbol{x}(\boldsymbol{i}|\boldsymbol{i}-\mathbf{1}) = \mathbf{F}\boldsymbol{x}(\boldsymbol{i}-\mathbf{1}) + \boldsymbol{G_1}\widetilde{\boldsymbol{P_1}}(\boldsymbol{i}-\boldsymbol{\delta}(\mathbf{1})) + \cdots + \boldsymbol{G}_{\mathrm{N}}\widetilde{\boldsymbol{P}_{\mathrm{N}}}\big(\boldsymbol{i}-\boldsymbol{\delta}(\mathrm{N})\big) \tag{9}$$

$$\mathbf{P(i|i-1)} = \mathbf{FP}(\boldsymbol{i}-\boldsymbol{\delta})\mathbf{F}^{T} + \boldsymbol{DH}(\boldsymbol{i})\ \mathbf{D}^{T} \tag{10}$$

$$\boldsymbol{y}(\mathbf{i}) = \boldsymbol{CX}(\boldsymbol{i}|\boldsymbol{i}-\mathbf{1}) \tag{11}$$

Correction:

$$\boldsymbol{x}(\mathbf{i}) = \boldsymbol{x}(\mathbf{i|i-1}) + \mathbf{P(i|i-1)}\boldsymbol{C}^{T}[\boldsymbol{R}(\mathbf{i}) + \boldsymbol{C}\mathbf{P(i|i-1)}\boldsymbol{C}^{T}]^{-1}\{\boldsymbol{y}(\mathbf{i}) - \boldsymbol{C}\mathbf{x(i|i-1)}\} \tag{12}$$

$$\mathbf{P(i)} = \mathbf{P(i|i-1)} - \mathbf{P(i|i-1)}\boldsymbol{C}^{T}[\boldsymbol{R}(\mathbf{i}) + \boldsymbol{C}\mathbf{P(i|i-1)}\boldsymbol{C}^{T}]^{-1}\boldsymbol{C}\mathbf{P(i|i-1)} \tag{13}$$

4 Results

In this section the results are presented for the 8 sub-basin models. Given the rainfall regime in the NIRB, there is not a significant flood event each year. For each sub-basin, the five years of data (1984, 1985, 1988, 1989, 1998), which had significant flood peaks, are used for model calibration, while the 2013 flood data are used for validation. Two types of sub-models are differentiated: those involving only a single weighted rainfall input, and those that have both rainfall and upstream discharge inputs. The results are presented both with and without data assimilation.

4.1 Design of rain gauge spatial weighting

The results of the stochastic optimization of the rainfall weights for each of the sub-models are shown in Table 1. Table 2 shows how the results compare with Thiessen polygon and simple averaging of the available rain gauges in each sub-model, in terms of the YIC and R_t^2 statistics of fitting the DBM model for the calibration data. The optimized weights show somewhat better results (more negative YIC and R_t^2 closer to 1) than the other methods. The differences are more marked in the sub-models having only rainfall as an input (1,2,4,5).

Table 1. The optimal weight of rain gauges in different sub-models determined by stochastic optimisation

Sub-models	Rain gauges Name (the Best weight)
(1)	Shihuiyao(0.13) Woduhe(0.37) Songlin(0.50)
(2)	Songlin(0.4) Guli(0.1) Zhuangzhi(0.4) Xintian(0.1)
(3) (4)	Kehou(0.75) Baiyun(0.25)
(5)	Jiwen(0.5) Jiagedaqi(0.13) Alihe(0.37)
(6)	Jiwen(0.17) Jiagedaqi(0.33) Liujiatun(0.5)
(7)	Haertong(0.3) Shihuiyao(0.3) Kumotun(0.4)
(8)	Kumotun(0.22)Kehou(0.09) Liujiatun(0.44) Nen River(0.34)

Table 2. A comparison of using different methods of defining raingauge weights (method of section 2.2, Thiessen polygons method and averaging method) in different sub-models

Sub-model	The best weight		Thiessen polygons		Averaging	
	YIC	R_T^2	YIC	R_T^2	YIC	R_T^2
(1)	-7.67	0.79	-6.5	0.67	-7.5	0.78
(2)	-9.45	0.87	-9.38	0.86	-9.3	0.86
(3)(4)	-8.34	0.79	-8.21	0.77	-7.94	0.75
(5)	-8.94	0.81	-8.92	0.81	-8.83	0.81
(6)	-7.79	0.9	-7.75	0.9	-7.68	0.9
(7)	-7.76	0.98	-7.52	0.97	-7.62	0.97
(8)	-8.73	0.98	-8.67	0.98	-8.64	0.98

4.2 Sub-models using a single input to forecast downstream discharge

The model structure of a single input single output model is defined by the following triplet:

$$[\mathrm{n\ m\ d}],$$

Where n is the denominator order (Equation (4)); m represents the numerator order of the input (Equation (3)); and d is the lead time of the input (δ in Equation (2)).

The outlet discharges for the 4 sub-models at sub-basins (1), (2), (4) and (5) are forecast using single input models. The outputs from sub-basin (3) are estimated by simple area of scaling from the output from (4). Because of the similarity of the modelling process at these sub-basins, Shihuiyao (sub-basin 1) is used as a representative site to show the modelling process and results. Assuming a power law nonlinearity and optimizing the exponent, we obtained the nonlinear relationship between rainfall and discharge as follows:

$$\boldsymbol{\breve{P}_1(i)} = \boldsymbol{P_1}(\boldsymbol{i}) \times \boldsymbol{Q_1^{0.3}}(\boldsymbol{i}-\mathbf{1}) \tag{14}$$

Where $\boldsymbol{Q_1}(\boldsymbol{i}-\mathbf{1})$ represents the discharge at Shihuiyao at the $(\boldsymbol{i}-\mathbf{1})^{\mathrm{th}}$ sample. $\boldsymbol{\breve{P}_1}$ represents the effective input of sub-basin 1, and $\boldsymbol{P_1}$ is the observed input of sub-basin 1.

The structure of the transfer function is [1 1 2], i.e. a simple first order DBM model with parameters identified by the SRIV algorithm as:

$$\boldsymbol{Q_1}(\boldsymbol{i}) = \frac{1.573}{1-0.8221z^{-1}}\breve{\boldsymbol{P}}_{\boldsymbol{1}}(\boldsymbol{i}-\boldsymbol{2}) \tag{15}$$

Following the procedure for identifying rain gauge weights outlined above, the combination minimizing the YIC criterion (-7.67) is chosen. This is a first order model without data assimilation which, when used with the power law transformation (14), gives a forecast coefficient of determination R_t^2 of 79%. For the same model used within the Kalman filter framework, the 2-day ahead forecast has a coefficient of determination of nearly 88% for the calibration period, and accounts for 92% of the observed variance in the 2013 forecasting period. As Figure 4 (a) shows, the discharge of the model without data assimilation is relatively inaccurate in the small flood events in 1984 and 1985 and the extremely large flood event in 1988. The reason for this is the limited rainfall gauge information in this sub-basin which has just 4 rainfall gauges on the two sides of the basin while the central area has no rainfall monitoring station. However, it is noteworthy that the Kalman filter could improve the accuracy of discharge by assimilating the latest flow data, which reduces the effect of the uncertainty resulting from the limited rainfall information to some extent. Regarding the uncertainty in the different weights for the rainfall gauges, the yellow curves in Figure 4 show clearly how the forecast uncertainty is significantly reduced after using the Kalman Filter and the accuracy of the forecasts has increased. As Figure 5 (a) shows, the observation points are all nearly in the CB shading which is the 95% confidence limits for the forecasts using the final DBM model and best weights of rainfall gauges found from the sampling. The values of flood peak in observation and forecast are 1690 m^3/s and 1711 m^3/s, respectively. A value of R_t^2 in the 2013 forecasting period of 92% is good enough for operational use in the future. The observation is larger than the forecast value following the forecast peak (2013/08/14), which may be still a result of the limited rainfall gauge availability.

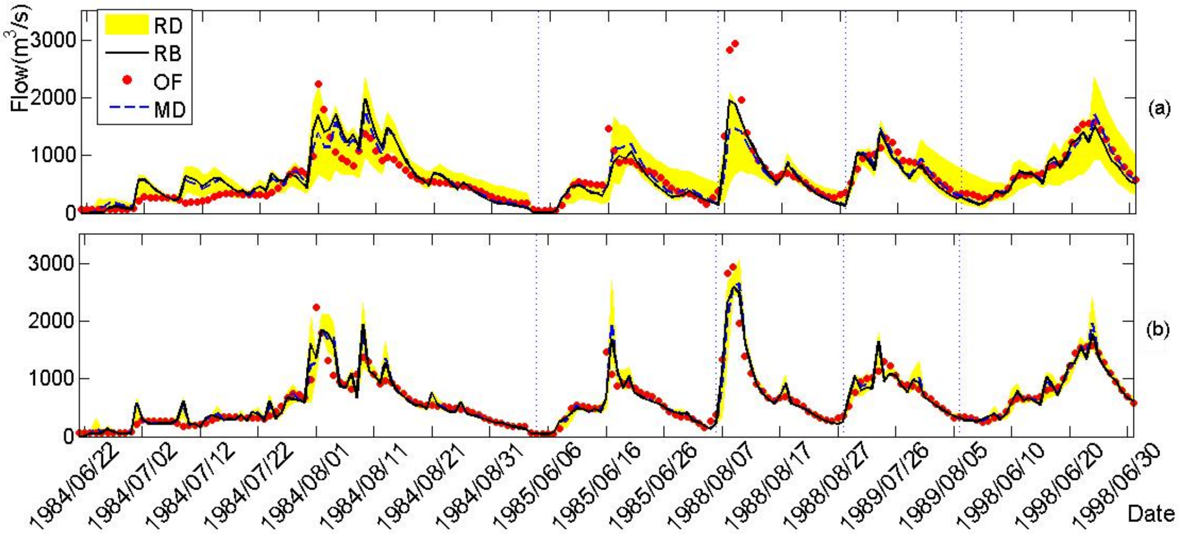

Figure 4. (a) 2-day ahead flood forecasts at Shihuiyao (sub-basin 1) in 1984,1985,1988,1989,1998 without Kalman Filter data assimilation; (b) 2-day ahead forecasts at Shihuiyao in 1984,1985,1988,1989,1998 with Kalman Filter data assimilation. RD shows the forecasting result using all 1000 sets of rainfall gauge weights; RB shows the forecasting result of the best weight set; OF is the observed flow value; MD is the mean forecasting result of the 1000 sets of rain gauge weights.

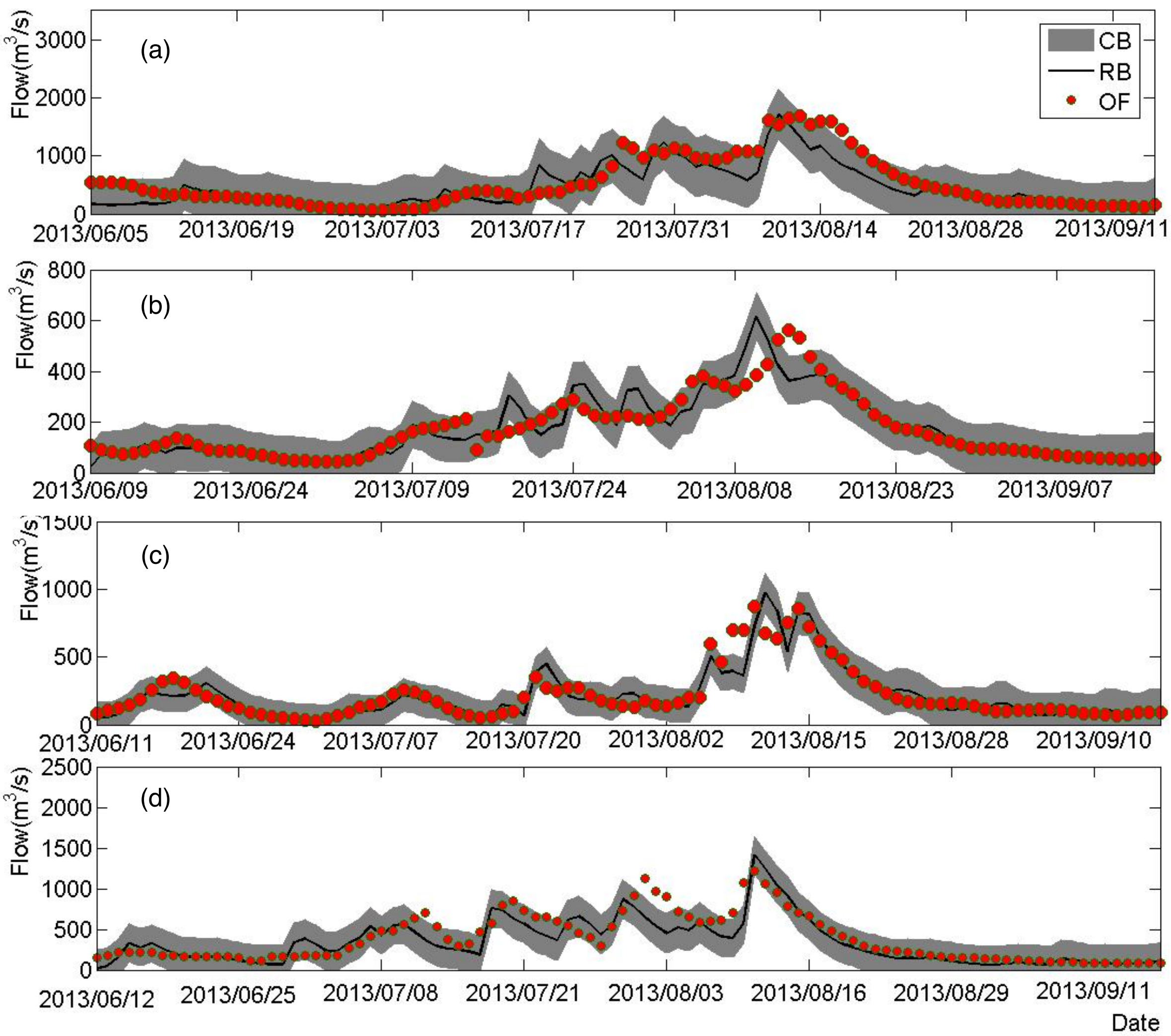

Figure 5. Forecasting results using Kalman Filter data assimilation (a) the 2-day ahead forecast for Shihuiyao in 2013 flood event ; (b) the 2-day ahead forecast for Guli in 2013 flood event;(c) the 2-day ahead forecast for Kehou in 2013 flood event;(d) the 3-day ahead forecast for Jiagedaqi in 2013 flood event. CB shading shows the 95% confidence limits for the forecasts using the final DBM model using the best weights of rainfall gauges; RB shows the forecasting result of the best weight set; OF is the observed flow value; MD is the mean forecasting result of the 1000 sets of rain gauge weights.

Table 3. Summary of model result of single input sub-models (p(1) p(2) p(3) in the model structure [p(1) p(2) p(3)] represents denominator order, numerators order, time delays respectively; the Sub-models are defined in Section 2.2 and a and b are defined in section 3.2).

Sub-model	Outlet Name	Area(km^2)	Power law	Structure	A	b
(1)	Shihuiyao	17205	0.3	[1 1 2]	[0 0 1.573]	[1 -0.8221]
(2)	Guli	5490	0.7	[1 1 2]	[0 0 0.05704]	[1 -0.8832]
(3),(4)	Kehou	7310	0.68	[1 1 2]	[0 0 0.2297]	[1 -0.8389]
(5)	Jiagedaqi	9575	0.2	[1 1 3]	[0 0 0 2.554]	[1 -0.8057]

Table 4. Evaluation measures for the single input models. The Sub-models are defined in Section 2.2.

Sub-model	Outlet Name	Calibration (1984 1985 1988 1989 1998)				Validation (2013)	
		Lead Time	YIC	without Kalman Filter	with Kalman Filter	Lead Time	
				R_T^2	R_T^2		R_T^2
(1)	Shihuiyao	2	-7.67	0.79	0.88	2	0.92
(2)	Guli	2	-9.45	0.82	0.89	2	0.89
(3),(4)	Kehou	2	-8.34	0.79	0.92	2	0.89
(5)	Jiagedaqi	3	-8.94	0.81	0.9	3	0.89

The Guli, Kehou, and Jiagedaqi sub-basins analyses resulted in similar model characteristics as Shihuiyao. The results are summarized in Tables 3 and 4. The coefficients of determination of each sub-basin for their own lead time ahead forecasts in the 2013 validation period are above 89%. Figure 5 show the results for this period graphically, showing that most observations are covered by the 95% confidence limits of the forecast, which suggests that the uncertainty in assessing the rainfall inputs is the main uncertainty in the whole forecasting process.

4.3 Multiple inputs including rainfall and upstream discharges to forecast the downstream discharge.

In the case of a model structure with multiple inputs, the general form of a DBM model can be defined by

$$[\text{den } \{\text{m}(1)\ \text{m}(2)\ \ldots \text{m}(N)\}\{\delta(1)\ \delta(2)\ \ldots \delta(N)\}] \tag{16}$$

Where N is the number of inputs, den is the denominator order; m (k) represents the numerator order of the k^{th} input; and $\delta(k)$ means the lead time of the k^{th} input.

The outlets of 3 sub-models at Kumotun, Liujiatun and Ayanqian represent the use of multiple inputs. The model for Kumotun (sub-model 7) will be discussed in some detail. The modelling process at Liujiatun and Ayanqian is very similar.

The discharge from the upstream stations of Shihuiyao, Guli and Menlu, together with the rainfall data for the sub-basin between the Shihuiyao station and the Kumotun station are applied to forecast the discharge of Kumotun station. Therefore, there are four inputs in the DBM model and one output (the discharge at the Kumotun station). The resulting transfer function model structure is

$$[1\ \{1\ 1\ 1\ 1\}\{3\ 1\ 1\ 1\}],$$

and the form of the TF decomposition is:

$$\boldsymbol{Q_4}(\boldsymbol{i}) = \frac{0.3606}{1-0.3204z^{-1}}\breve{\boldsymbol{P}}_{\boldsymbol{2}}(\boldsymbol{i}-\boldsymbol{3}) + \frac{0.8022}{1-0.3204z^{-1}}\boldsymbol{Q_1}(\boldsymbol{i}-\boldsymbol{1}) + \frac{0.4676}{1-0.3204z^{-1}}\boldsymbol{Q_2}(\boldsymbol{i}-\boldsymbol{1})$$

$$+\frac{0.6679}{1-0.3204z^{-1}}\boldsymbol{Q_3}(\boldsymbol{i}-\boldsymbol{1}) \tag{17}$$

Where $\boldsymbol{Q_1}(\boldsymbol{i})$ represents the discharge at Shihuiyao at the $(\boldsymbol{i})^{\text{th}}$ sample, and the lead time from Shihuiyao to Kumotun is 1 day; $\boldsymbol{Q_2}$ represents the discharge of Menlu with a 1-day lead time; $\boldsymbol{Q_3}$ represents the discharge of Guli with a 1-day lead time; $\boldsymbol{Q_4}$

represents the discharge at Kumotun; and $\breve{\boldsymbol{P}}_2$ represents the effective input of the 3 weighted rain gauges (Haertong, Huolongmen and Shihuiyao) with a 3-day lead time, $\boldsymbol{P}_2$ is the observed input of the 3 weighted rain gauges. The best nonlinear power law function of rainfall was identified1

$$\breve{\boldsymbol{P}}_2(\boldsymbol{i}) = \boldsymbol{P}_2(\boldsymbol{i}) \times \boldsymbol{Q}_4^{0.3}(\boldsymbol{i}-\boldsymbol{1}) \quad (18)$$

As shown in Figure 6 and Table 5, the forecasting results without the Kalman filter and the forecasting results with the Kalman filter are both very good in calibration. The RD is very small which means the rain gauge weightings in this case has a small influence in the accuracy of forecasting the flood peak of forecasting flood. This is also reflected in the numerator of the effective rainfall in the TF decomposition (Equation14), which is smaller than other inputs. The performance of forecasting process in the 2013 validation period is shown in Figure 7. The coefficient of determination is nearly 95%.

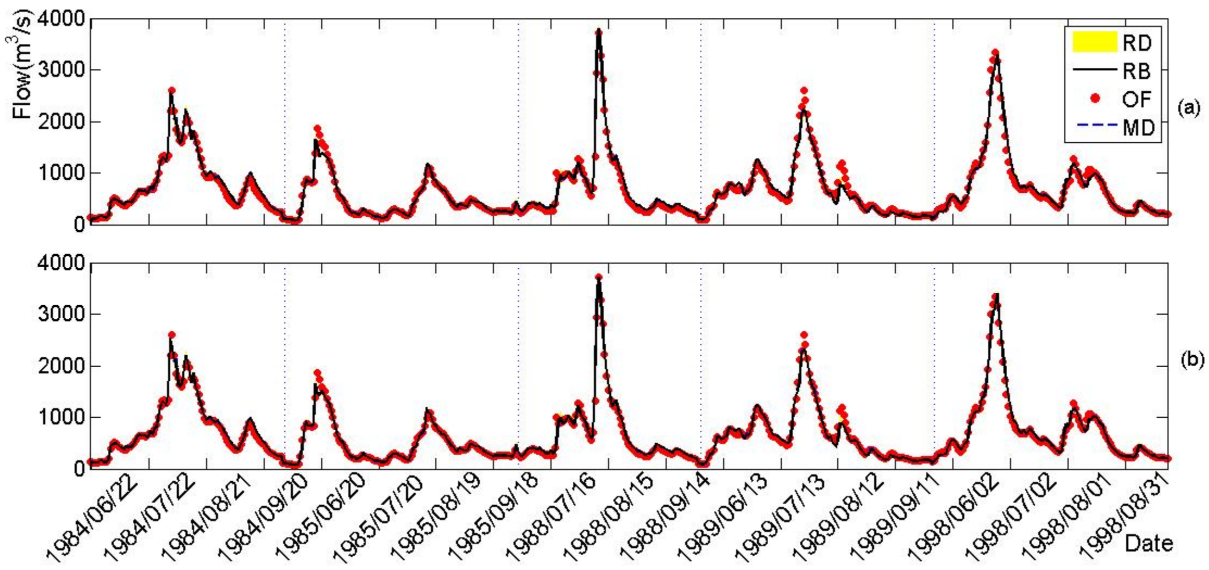

Figure 6(a) 3-day ahead flood forecast for Kumotun without data assimilation in 1984,1985,1988,1989,1998 with different rainfall weights; (b) 3-day ahead flood events forecast of

Kumotun in 1984,1985,1988,1989,1998 with different rainfall weights with Kalman Filter data assimilation. The meaning of each variable is as defined in Figure 5.

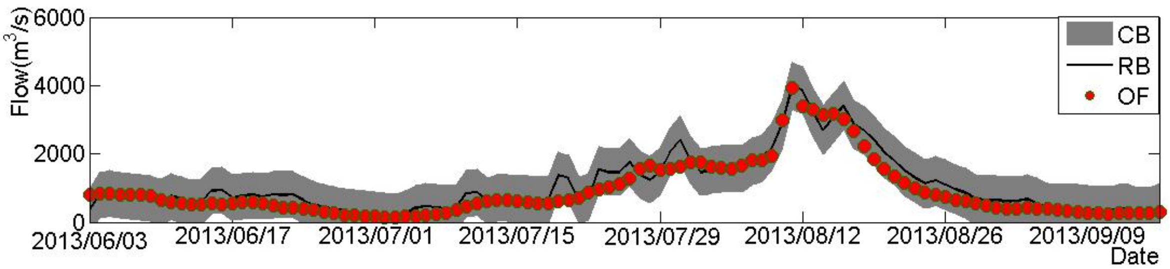

Figure 7. The 3-day ahead forecasting result with Kalman Filter data assimilation for Kumotun in 2013 flood event. The meaning of each variable is as defined in Figure 5.

Table 5. Summary of model result of multiple inputs sub-models (p(1) p(2) p(3) in the model structure [p(1) p(2) p(3)] represents denominator order, numerators order, time delays respectively; the Sub-models are defined in Section 2.2.)

Sub-model	Outlet Name	Area(km^2)	Power law	Structure
(6)	Liujiatun	10090	1	[1 1 1 3 1]
(7)	Kumotun	9534	0.3	[1 1 1 1 1 3 1 1 1]
(8)	Ayanqian	7178	0.9	[1 1 1 1 1 4 3 1 2]

Table 6. Evaluation measures for the single input models. The Sub-models are defined in Section 2.2.1

Sub-model	Outlet Name	Calibration (1984 1985 1988 1989 1998)				Validation (2013)	
		Lead Time	YIC	without Kalman Filter	with Kalman Filter	Lead Time	
				R_T^2	R_T^2		R_T^2
(6)	Liujiatun	1	-7.79	0.9	0.92	3	0.92
(7)	Kumotun	1	-7.76	0.97	0.98	3	0.95
(8)	Ayanqian	1	-8.67	0.98	0.98	4	0.96

The structures of the sub-models for Liujiatun and Ayanqian are as follows:

Liujiatun:

$$\boldsymbol{Q_6}(\boldsymbol{i}) = \frac{0.01415}{1-0.6482z^{-1}}\breve{\boldsymbol{P}}_{\mathbf{3}}(\boldsymbol{i}-\mathbf{3}) + \frac{0.4759}{1-0.6482z^{-1}}\boldsymbol{Q_5}(\boldsymbol{i}-\mathbf{1}) \qquad (19)$$

Where $\boldsymbol{Q_5}$ represents the discharge of Jiagedaqi, and the lead time from Jiagedaqi to Liujiatun is 1 day. $\breve{\boldsymbol{P}}_3$ represents the effective input from rainfall observation(effective rainfall) in the basin above Jiagedaqi with the 3-day lead time and $\boldsymbol{P_3}$ is the observed input in the basin. The nonlinear function of rainfall (effective rainfall nonlinearity) was estimated as:

$$\breve{\boldsymbol{P}}_{\mathbf{3}}(\boldsymbol{i}) = \boldsymbol{P}_{\mathbf{3}}(\boldsymbol{i}) \times \boldsymbol{Q}_{\mathbf{6}}^{\mathbf{1.00}}\ (\boldsymbol{i}-\mathbf{1}) \qquad (20)$$

Ayanqian:

$$\boldsymbol{Q_9}(\boldsymbol{i}) = \frac{0..01018}{1-0.4278z^{-1}}\breve{\boldsymbol{P}}_{\mathbf{4}}(\boldsymbol{i}-\mathbf{4}) + \frac{0.7408}{1-0.4278z^{-1}}\boldsymbol{Q_7}(\boldsymbol{i}-\mathbf{3}) + \frac{0.7281}{1-0.4278z^{-1}}\boldsymbol{Q_6}(\boldsymbol{i}-\mathbf{1})$$

$$+\frac{0.5019}{1-0.4278z^{-1}}\boldsymbol{Q_4}(\boldsymbol{i}-\mathbf{2}) \qquad (21)$$

Where $\boldsymbol{Q_7}$ represents the discharge of Kehou, and the lead time from Kehou to Ayanqian is 3-day. where $\boldsymbol{Q}_6$ represents the discharge of Liujiatun, and the lead time from Liujiatun to Ayanqian is 1-day. where $\boldsymbol{Q_4}$ represents the discharge of Kumotun, and the lead time from Kumotun to Ayanqian is 2-day. $\breve{\boldsymbol{P}}_4$ and $\boldsymbol{P_4}$ represent the effective input and the observed input from rainfall observation between Kumotun and Ayanqian with the 4-day lead time. The nonlinearity for the effective rainfall for this data set is found as:

$$\breve{\boldsymbol{P}}_{4}(\boldsymbol{i}) = \boldsymbol{P}_{4}(\boldsymbol{i}) \times \boldsymbol{Q}_{9}^{0.9}\,(\boldsymbol{i}-\mathbf{1}) \tag{22}$$

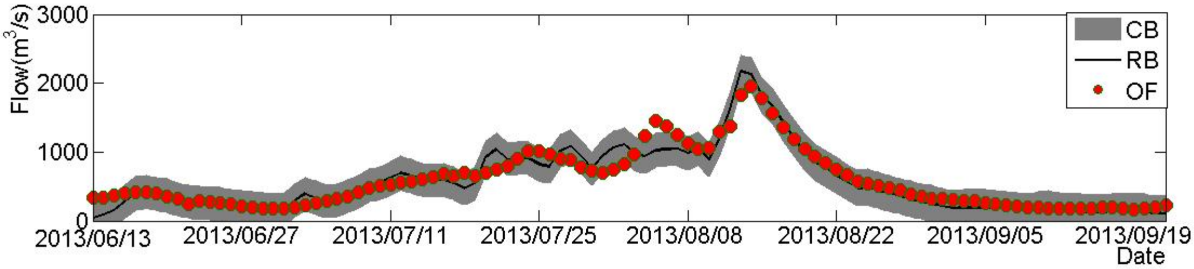

Figure 8. The 3-day forecasting result with Kalman Filter data assimilation for Liujiatun in 2013 flood event. The meaning of each variable is as defined in Figure 5.

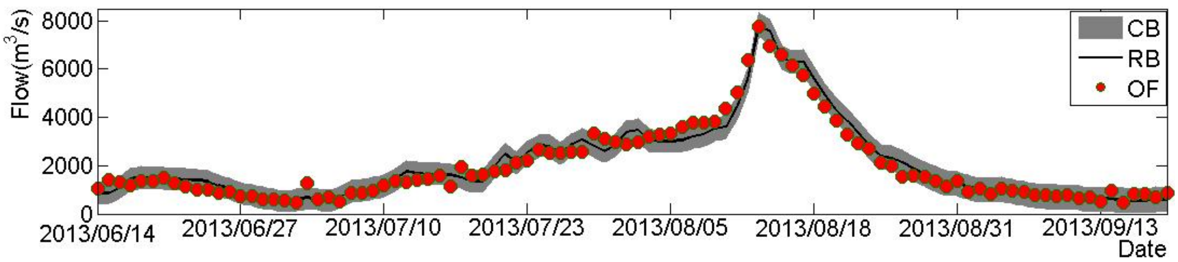

Figure 9. The 4-day forecasting result with Kalman Filter data assimilation for Ayanqian in 2013 flood event. The meaning of each variable is as defined in Figure 5.

Here, the parameters and model structures for the multiple input sub-models have been summarized in Table 5 and Table 6. Lijiatun has two inputs including rainfall between Jiagedaqi and Liujiatun and the discharge of Jiagedaqi with 92% coefficients of determination in the calibration period and 92% in out–of-sample forecasting for 2013. Ayanqian has four inputs including rainfall between Kumotun and Ayanqian and the discharge of Liujiatun, Kehou and Kumotun and the coefficient of determination R_t^2 in calibration and out-of-sample forecasting are 98% and 96% respectively. It is noteworthy that the lead time and value of the final forecast are very accurate at the peak of the flood, but are at times less accurate for relatively low flow levels. However, the performance of the forecasting results in Ayanqian is nearly 96%, which is adequate for the operational forecasting of inputs to the Nierji reservoir.

In additional, the 1 day, 2 days, 3 days and 5 days forecasting results are shown in Figure 10 with coefficients of the determination of 98%, 97%, 96% and 93% respectively. When the forecasting lead time is four days, we need to use extended forecasts from Kumotun and Kehou, beyond the identified time delays, which could certainly increase the uncertainty of the result. With an increase in the lead time from 2 days to 4 days, the uncertainty and the accuracy of the flood forecasting result of Ayanqian Station, which is also the outlet of Nierji Basin, is decreasing from 98% to 93%. However, this is still sufficiently accurate to provide useful information for forecasting further downstream for the whole Songhua River and consequently improve decision making about flood management and warnings.

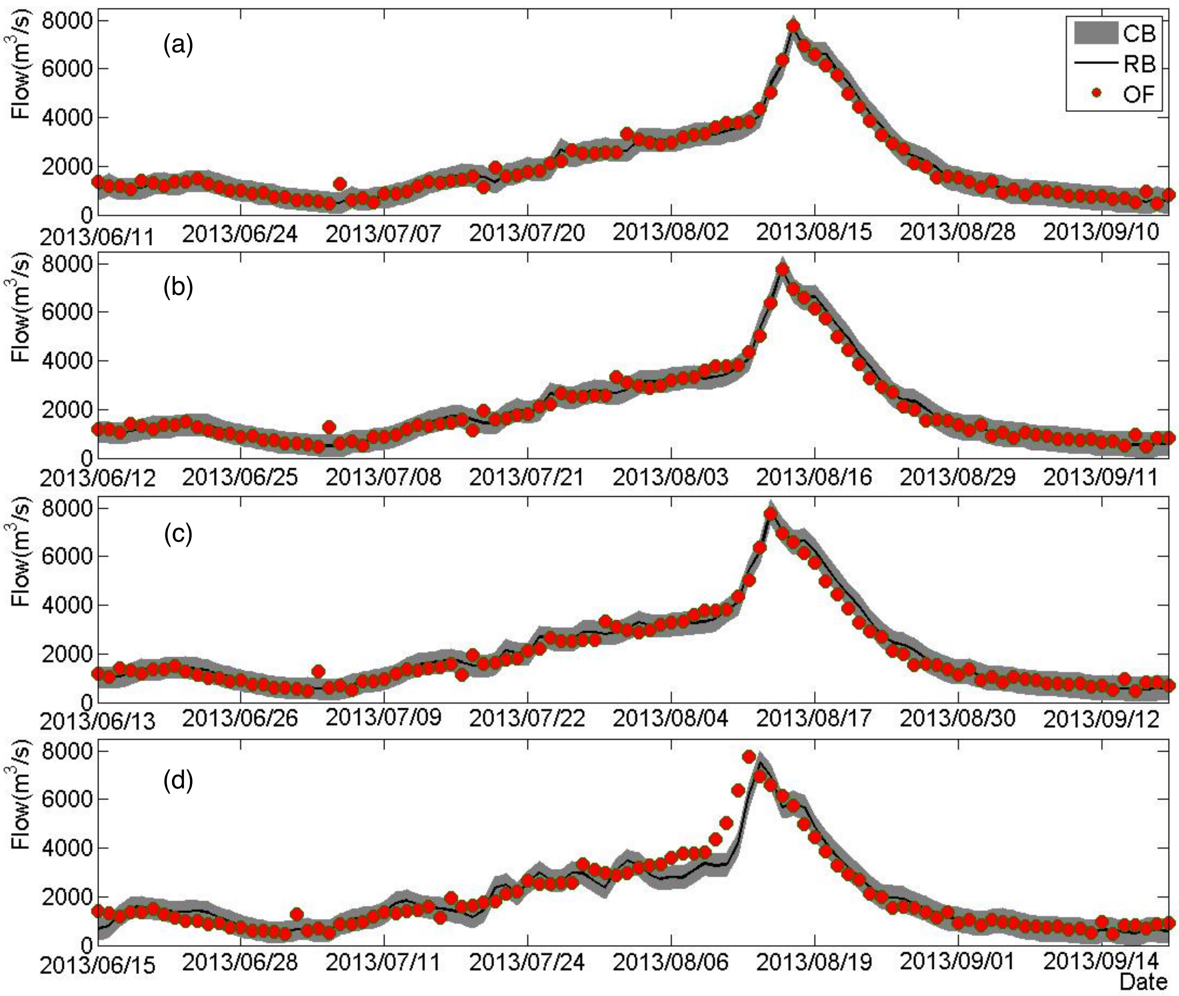

Figure 10. The 1-day, 2-day, 3-day and 5-day forecasting result with Kalman Filter data assimilation for Ayanqian in 2013 flood event, illustrated in (a), (b), (c) and (d), respectively. The meaning of each variable is as defined in Figure 5.

5 Discussion and Conclusions

The main aim of this paper has been to develop flood forecasting models for the Nen River Basin above the Nierji Reservoir using the adaptive DBM methodology so as to maximize the lead time and accuracy of forecasts as an input to flood management in the whole Songhua River. This paper represents the first application of the DBM methodology to flood forecasting in China with Kalman filter data assimilation, coupled to a stochastic optimisation approach to finding weights and nonlinearity identification. It has produced accurate forecasting results with lead times up to 5 days ahead with the associated uncertainties updated on a daily basis.

The main findings from this study can be summarized as follows:

(1) In this data-sparse basin, the stochastic optimisation of rain gauge weights produces better results than either Thiessen polygons or simply averaging.

(2) The adaptive version of the Kalman filter can, to some extent, handle data and model uncertainties and improve the accuracy of the forecasting model significantly; the coefficients of determination have increased by up to nearly 10% as shown in Table 4.

(3) The multiple input models are generally more accurate than the single input models because the multiple input models all include an upstream discharge input which does not have the degree of uncertainty of the rainfall-runoff process, even though the observed discharges themselves may be subject to significant uncertainties, particularly for the flood peaks.

(4) The models identified by the DBM methodology provide a suitable basis for forecasting in the study basin. The coefficients of determination for the final 1-day, 2-day, 3-day, 4-day and 5-day ahead forecasting at the Nierji Reservoir are 98%, 97%, 96%, 96% and 93%, respectively, and can therefore provide a useful forecast for the operation of downstream reservoirs under flood conditions.

One of the limitations of the DBM methodology is that it will provide results only where data series are available for calibration (except for cases where simple scaling can be used for nearby sites, as with the Menlu sub-basin in this study). However, a particular advantage of the approach is that the model representations are easily updated as new data are made available. In doing so, we note that models can be applied to predicted water levels, rather than discharges, making the installation of new forecasting points as required rather inexpensive (see for example Leedal et al., 2013). Thus new data sources are easily added to the forecasting system, and models are easily evaluated and recalibrated after each major event provides new information about the system response.

Acknowledgements

The work was supported by National Key Project (Grant No. 2016YFC0400903) and is partly funded by the National Natural Science Foundation of China (Grant Nos.51320105010, 51779030). We are grateful for the comments from referees which led to a greatly improved paper.

References

Alfieri L., Smith P.J., Thielen-del Pozo J., and Beven K.J., 2011, A staggered approach to flash flood forecasting – case study in the Cevennes Region, *Adv. Geosci.* 29, 13-20.

Beven, K J, 2012, *Rainfall-Runoff Modelling: The Primer*, 2nd edition, Wiley-Blackwell: Chichester.

Beven, K J, Leedal, D T, Smith, P J and Young, PC, 2011, Identification and representation of state dependent non-linearities in flood forecasting using the DBM methodology, in L. Wang, and H. Garnier,(Eds.), System Identification, Environmetric Modelling and Control, Berlin. Springer-Verlag, 341-366.

Beven, K. J. and M. J. Kirkby, 1979, A physically based, variable contributing area model of basin hydrology / Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant. *Hydrological Sciences Bulletin*, 24.

Brath, A., Montanari, A. and Toth, E., 2002. Neural networks and non-parametric methods for improving real-time flood forecasting through conceptual hydrological models. *Hydrology and Earth System Sciences Discussions*, *6*(4), pp.627-639.

Chang, F.J., Chiang, Y.M. and Chang, L.C., 2007. Multi-step-ahead neural networks for flood forecasting. *Hydrological sciences journal*, *52*(1), pp.114-130.

Chen, P., J. D. Wang, and S.L. Liang. 2012. A Data-Based Mechanistic Approach to Time-Series LAI Modelling and Estimation. *Journal of Remote Sensing,* 16: 505-519.

Cheng, C. and M. Zhao, et al., 2006, Using genetic algorithm and TOPSIS for Xinanjiang model calibration with a single procedure. *Journal of Hydrology*, 316 (1-4), pp.129-140.

Chau, K.W., Wu, C.L. and Li, Y.S., 2005. Comparison of several flood forecasting models in Yangtze River. *Journal of Hydrologic Engineering*, 10(6), pp.485-491.

Cloke, H.L. and Pappenberger, F., 2009. Ensemble flood forecasting: A review. *Journal of Hydrology*, 375(3-4) , pp.613-626.

Dawson, C.W., Harpham, C., Wilby, R.L. and Chen, Y., 2002. Evaluation of artificial neural network techniques for flow forecasting in the River Yangtze, China. *Hydrology and Earth System Sciences Discussions*, 6(4), pp.619-626.

Franz, K.J., Hartmann, H.C., Sorooshian, S. and Bales, R., 2003. Verification of National Weather Service ensemble streamflow predictions for water supply forecasting in the Colorado River basin. *Journal of Hydrometeorology*, *4*(6), pp.1105-1118.

Griffith, DA, 1993, Statistics for spatial data, *Geographical analysis*, 25(3): 271-275

Guo, L., Wang, J., Xiao, Z. et al. 2014, Data-based mechanistic modelling and validation for leaf area index estimation using multi-angular remote-sensing observation time series. *International Journal of Remote Sensing*, 35: 4655-4672.

Guo, S. and J. Zhang, et al., 2009, Flood forecasting system of Hanjiang Basin based on meteorological model(VIC). *Advance in Science and Technology of Water Resources(China)* , 29 (03) , pp.1-5.

Han, D., Kwong, T. and Li, S., 2007. Uncertainties in real-time flood forecasting with neural networks. *Hydrological Processes*, *21*(2), pp.223-228.

Herr, H D and Krzysztofowicz, R., 2010, Bayesian ensemble forecast of river stages and ensemble size requirements, *Journal of Hydrology,* 387: 151-164.

Hutchinson M F., 1998, Interpolation of rainfall with thin plate smoothing splines-Part1: Two-dimensional smoothing of data with short range correlation, *Journal of Geographic Information and Decision Analysis*, 2: 139-151.

Krzysztofowicz, R., 2002, Bayesian system for probabilistic river stage forecasting, *Journal of Hydrology*, 268: 16-40.

Lambert, A. O., 1972, Catchment models based on ISO functions, *J. Institute Water Engineering,* 26: 413-422.

Leedal, D T, Weerts, A, Smith, P. J. and Beven, K J, 2013, Application of Data Based Mechanistic modelling for flood forecasting at multiple locations in the Eden Catchment in the National Flood Forecasting System (England and Wales), *Hydrol. Earth Syst*. Sci., 17: 177-185.

Lees, M., P.C. Young, S. Ferguson, K.J. Beven and J. Burns (1994), An adaptive flood warning scheme for the River Nith at Dumfries, in W. R. White and J. Watts (Eds.), *River Flood Hydraulics*, 65-75, Wiley, Chichester.

Liu, J., Chen, X., Zhang, J. and Flury, M., 2009. Coupling the Xinanjiang model to a kinematic flow model based on digital drainage networks for flood forecasting. *Hydrological processes*, *23*(9), pp.1337-1348.

Li, K. and B. Shen, et al., 2015, Application and improvement of TOPMODEL for rainfall runoff modeling in semi-humid region. *Journal of Hydraulic Engineering(China)*, (12) , pp.1453-1459.

Li, X.Y., Chau, K.W., Cheng, C.T. and Li, Y.S., 2006. A Web-based flood forecasting system for Shuangpai region. *Advances in Engineering Software*, 37(3), pp.146-158.

Li, J., L. Zhang, X. Shi, and Y. Chen, 2016: Response of long-term water availability to more extreme climate in a rain-fed basin, *International Journal of Climatology*, 10.1002/joc.4910.

Li, Z., Z. Hao, X. Shi*, S. J. Déry, J. Li, S. Chen, and K. Li, 2016: An agricultural drought index to incorporate the irrigation process and reservoir operations: A case study in the Tarim River Basin, *Global and Planetary Change*, 143, 10-20.

Liu, T. and Fan-long M., et al., 2012, Research on Short Term Flood Forecast for Nierji Reservoir Based on BP Neural Network. *Heilongjiang Science and Technology of Water Conservancy (China)*, 10(40), pp.13-16.

Liu, Y. and Z. Li, et al., 2016, TOPKAPI-based Flood Simulation in Semi-humid and Semi-arid Regions. *Water Power(China)*, (01), pp.18-22.

Liu, Y. and Z. Liu, et al., 2016, Runoff Simulation for the Upper Reaches of Heihe River Basin Based on TOPKAPI Model. *Water Power(China)*, (12), pp.20-23+118.

Liu, Z., 2004, Application of GIS-based distributed hydrological model to flood forecasting. *Journal of Hydraulic Engineering*, (05), pp.70-75.

Liu, Z., Martina, M.L. and Todini, E., 2005. Flood forecasting using a fully distributed model: application of the TOPKAPI model to the Upper Xixian Catchment. *Hydrology and Earth System Sciences Discussions*, *9*(4), pp.347-364.

Lu, M. and X. Li, 2015, Strategy to automatically calibrate parameters of a hydrological model: a multi-step optimization scheme and its application to the Xinanjiang model. *Hydrological Research Letters*, 9 (4), pp.69-74.

Moradkhani, H, Hsu, K, Gupta, H V, and Sorooshian, S, 2005, Uncertainty assessment of hydrological states and parameters: x sequential data assimilation using particle filters, *Water Resources Research*, 41: W05012; doi: 10.1029/2004WR003604

Oleyiblo, J. O. and L. I. Zhi-Jia, 2010, Application of HEC-HMS for flood forecasting in Misai and Wanan catchments in China. *Water Science and Engineering*, 3 (1), pp.14-22.

Panigrahy, N; Jain, SK: Kumar, V; Bhunya, PK, 2009, Algorithms for computerized estimation of Thiessen Weights, *Journal of Computing in civil engineering*, 23(4): 239_247, Peng, Y. and X. Sun, et al., 2017, A Flood Forecasting Model that Considers the Impact of Hydraulic Projects by the Simulations of the Aggregate Reservoir's Retaining and Discharging. *Water Resources Management*, 31 (3), pp.1031-1045.

Ren-Jun, Z., 1992. The Xinanjiang model applied in China. Journal of hydrology, 135(1-4), pp.371-381.

Reggiani, P and Weerts, A H, 2008, A Bayesian approach to decision making under uncertainty: an application to real time forecasting on the River Rhine, *J. Hydrology*, 356: 56-69.

Romanowicz, R, Young, P C and Beven, K J, 2006, Data assimilation and adaptive forecasting of water levels in the River Severn catchment, UK, *Water Resour. Res.,* 42, W06407, doi:10.1029/2005WR004373

Romanowicz, RJ, Young, PC, Beven, KJ and Pappenberger, F, 2008, A Data Based Mechanistic Approach to Nonlinear Flood Routing and Adaptive Flood Level Forecasting, *Advances in Water Resources*, 31:1048–1056

Schaake, J C, Hamill, T H, Buizza, R and Clark, N, 2007, HEPEX – the hydrological ensemble prediction experiment, Bull. Amer. Meteor. Soc., 88: doi: 10.1175/BMS-88-10-1541

Sene, K, Weerts, A, Beven, K J, Moore, R, Whitlow, C, Laeger, S and Cross, R, 2014, Uncertainty estimation in fluvial flood forecasting applications, Chapter 17 in Beven, K J and Hall, JW (Eds.) Applied Uncertainty Analysis for Flood Risk Management, Imperial College Press / World Scientific: London, 462-498

Smith, P J, Beven, K J, Tych, W, Hughes, D, Coulson, G and Blair, G, 2008, The provision of site specific flood warnings using wireless sensor networks, in *Flood Risk Management: Research and Practice,* edited by P Samuels, S Huntingdon, W Allsop, and J Harrop, CRC Press / Balkema, ISBN 978-0-415-48507-4.

Smith, P. J., K. J. Beven, A. Weerts, and D. Leedal, 2012, Adaptive correction of deterministic models to produce accurate probabilistic forecasts, *Hydrol. Earth. Syst. Sci.*, 16: 2783-2799

Smith, P. J., Beven, K. J., Horsburgh, K., 2013a, Data Based Mechanistic modelling of tidally affected river reaches for flood warning purposes: An example on the River Dee, UK, *Quart. J. Roy. Meteor. Soc.*, *139*(671), 340-349, DOI:10.1002/qj.1926

Smith, P J, L. Panziera and K. J. Beven, 2013b, Forecasting flash floods using Data Based Mechanistic models and NORA radar rainfall forecasts, *Hydrological Sciences Journal, 59*(7), 1403-1417, DOI:10.1080/02626667.2013.842647

Smith, P. J., K.J. Beven, D. Leedal, A.H. Weerts and P. C. Young, 2014, Testing probabilistic adaptive real-time flood forecasting models, *J. Flood Risk Management,*7(3), 265-279, DOI: 10.1111/jfr3.12055

Solomatine, D.P. and Xue, Y., 2004. M5 model trees and neural networks: application to flood forecasting in the upper reach of the Huai River in China. *Journal of Hydrologic Engineering*, 9(6), pp.491-501.

Taylor, C.J., Pedregal, D.J., Young, P.C., Tych, W., 2007. Environmental time series analysis and forecasting with the Captain toolbox. *Environ. Model. Software* 22, (6), 797-814.

Thiessen, A. H., 1911, Precipitation averages for large areas, *Mon.Weather Rev.*, 39(7): 1082–1084.

Wu, C.L. and Chau, K.W., 2006. A flood forecasting neural network model with genetic algorithm. *International journal of environment and pollution*, 28(3-4), pp.261-273.

Wang, B., 1996, Reservoir flood control forecasting method and application.

Wang, M. and Y. Peng, et al., 2012, Application of DHF Model in Flood Forecasting for Shifosi Basin. *South-to-North Water Diversion and Water Science & Technology(China)*, 10 (2), pp.93-97.

Wei, Ding and Chi. Zhang, et al., 2015, An analytical framework for flood water conservation considering forecast uncertainty and acceptable risk. *Water Resources Research*, 51(6), pp.4702-4726.

Xue, X. and K. Zhang, et al., 2016, New Multisite Cascading Calibration Approach for Hydrological Models: Case Study in the Red River Basin Using the VIC Model. *Journal of Hydrologic Engineering*, 21 (2).

Xiao, X. and H. Zhang, et al., 2017, TOPMODEL model for flood prediction applied in Beixi, Jiulong River, China. *Hydraulic Science and Technology(China)*, (2).

Young, P.C. and K.J. Beven, 1994, Data-based mechanistic modelling and the rainfall-flow non-linearity, Environmetrics, V.5, 335-363.

Young, P.C., 2002, Advances in real-time flood forecasting, Philos. T. Roy. Soc. A, 360, 1433–1450.

Young, P.C., 2011, Recursive Estimation and Times Series Analysis, Springer-Verlag: Berlin, Heidelburg

Young, P.C., Romanowicz, R and Beven, K J, 2014, A Data-Based Mechanistic Modelling Approach to Real-Time Flood Forecasting, Chapter 16 in Beven, K J and Hall, JW (Eds.) Applied Uncertainty Analysis for Flood Risk Management, Imperial College Press / World Scientific: London, 407-461

Yao, C., Li, Z.J., Bao, H.J. and Yu, Z.B., 2009. Application of a developed Grid-Xinanjiang model to Chinese watersheds for flood forecasting purpose. *Journal of Hydrologic Engineering*, 14(9), pp.923-934.

Yao, C. and Y. Ji, et al., 2012, Parameter estimation and application of grid-based Xinanjiang model. *Journal of Hohai University (Natural Sciences) (China)*, (01), pp.42-47.

Yao, C. and Z. Li, et al., 2012, A priori parameter estimates for a distributed, grid-based Xinanjiang model using geographically based information. *Journal of Hydrology*, 468-469, pp.47-62.

Yao, C. and K. Zhang, et al., 2014, Improving the flood prediction capability of the Xinanjiang model in ungauged nested catchments by coupling it with the geomorphologic instantaneous unit hydrograph. *Journal of Hydrology*, 517, pp.1035-1048.

Yang Peng, Liu Xiaohu, LIANG Shuxian, 2003, The application of the mathematical model of distance weighted mean to rainfall isoline in Haihe River Basin, *Hydrology (China)*, 2(23):

Zhao, R., 1984, Watershed hydrological simulation, *Water Resources and Electric Power Press*.

Zhao, R., 1992, The Xinanjiang model applied in China. *Journal of Hydrology*, 135, pp.371-381.

Zhi-Jia, L. I. and C. Yao, et al., 2007, Development and application of grid-based Xinanjiang model. *Journal of Hohai University (Natural Sciences) (China)*, (02) , pp.131-134.

Zhou, H., Wang, J., Liang, S. et al., 2017. Extended Data-Based Mechanistic Method for Improving Leaf Area Index Time Series Estimation with Satellite Data. *Remote Sensing*, 9(6), 533;