New Approaches to Hydrological Prediction in Data-sparse Regions (Proc. of Symposium HS.2 at the Joint IAHS & IAH Convention, Hyderabad, India, September 2009). IAHS Publ. 333, 2009

Estimating precipitation for poorly-gauged areas in western China

JUNLIANG JIN[1,2], GUIHUA LU[1,3] & ZHIYONG WU[1,3]

1 *State Key Laboratory of Hydrology, Water Resources and Hydraulic Engineering, Hohai University, Nanjing 210098, China*
jinjunliang@gmail.com

2 *College of Hydrology and Water Resources, Hohai University, Nanjing 210098, China*

3 *Research Institute of Water Problems, Hohai University, Nanjing 210098, China*

Abstract Hydrological simulations in data-sparse areas have large uncertainties. This paper proposes spatial geo-statistical interpolation algorithms based on the hydrological analogy method to estimate the spatial distribution of precipitation for data-sparse areas using Tropical Rainfall Measuring Mission (TRMM) precipitation radar (PR) data and a small number of available recorded rainfall data. Taking the Kaidu River basin in Xinjiang, China, as a case study, using to the relationship between TRMM PR data and sparsely-recorded rainfall, the spatial distribution of precipitation was estimated with the proposed method. A macro-scale land hydrological model, the Variable Infiltration Capacity (VIC) model, was then established over the study basin with the derived data. Hydrological simulation over five data-sparse basins (including Dashankou, Xining, Jiayuguan, Yingluoxia and Qingshizui) indicates that the estimated precipitation from TRMM PR data significantly improved the accuracy of hydrological simulation; the proposed method can therefore be used to estimate the spatial distribution of precipitation for sparsely-gauged areas in western China.

Keywords data-sparse area; TRMM; hydrological simulation; VIC model

INTRODUCTION

Precipitation is the basis of the hydrological cycle. However precipitation gauges are scarce in most areas of western China. How to obtain areal precipitation over large scales that can be used to drive land surface hydrological models is a central issue for hydrologists concerned with this region. With the development of computer and GIS science, many scholars have carried out research in this topic (Wang *et al*., 2003; Zhang & He, 1998; Seiji *et al*., 2004; Zhou *et al*., 2006).

Traditional hydrological methods obtain the areal precipitation of a watershed through raingauge observations. These methods calculate areal precipitation by interpolating point precipitation, using techniques such as the Thiessen polygon method and area weighted method. These methods improve in accuracy with the density of precipitation gauges.

The development of remote sensing technology and satellites for precipitation estimates, along with more accurate tropical rainfall measurements from the Tropical Rainfall Measuring Mission (TRMM) Microwave Imager (TMI) and precipitation radar (PR) instruments, have made it possible to monitor tropical rainfall diurnal patterns and their intensities from satellite information. The TRMM satellite was launched on 27 November 1997 from Japan's Tanegashima launch facility. Fundamentally, the TRMM research programme is dedicated to measuring tropical–subtropical rainfall over a lengthy time period, and by doing so, acquiring the first accurate, representative and consistent ocean climatology of precipitation. TRMM was initially launched into a low-altitude (350 km), non-sun synchronous orbit inclined at 50 degrees to the Earth's equatorial plane, with a nominal mission lifetime of three years, but with expectations for a longer lifetime. Thus it has better ability to measure rainfall than other passive satellite remote sensing. There are many studies carried out utilizing the TRMM PR product (Huffman *et al*., 1995, 2007; Huffman, 1997).

This paper describes a method to create meteorological forcings in data-sparse regions that can be used to drive a model of land surface hydrology. Taking the Kaidu River basin of Xinjiang province as a case study, it brings forward a new method of interpolating precipitation in data-sparse regions on the basis of analysing the correlation of TRMM precipitation and a small number of available recorded observations. Motivation for this study is the creation of a national, multidecade, terrestrial, meteorological forcing data set to drive land surface model simulations of the national water balance. Therefore, the effect of the correction is discussed in terms of the land surface water budget by analysing long-term simulations using the Variable Infiltration Capacity (VIC) land surface model. This highlights a new method of estimating precipitation for hydrological simulation in data-sparse regions.

DATA SOURCES AND METHODOLOGY

Description of the Kaidu River basin

The Kaidu River basin is located at the northern edge of YanQi basin of XinJiang province, China (Figs 1 and 2). It lies approximately between 80°58′E–86°55′E and 41°47′N–43°21′N and derives from the Sa'arming Hills of the north Tianshan Mountains. It is one of the nine headstreams of the Tarim River, and also is the biggest river in Bayinbuluke autonomous prefecture. The upper Kaidu River flows through the small Yourdusi basin eastward then turns to southeast to Uygur Autonomous Region, and along a big canyon to the Dashankou hydrological station, the outlet of this mountain river.

High mountains, canyons and basins, i.e. a complicated topography, exist in the Kaidu River basin. The elevation changes from 4589 m to 1340 m (outlet) and mean basin elevation is 3100 m. The area upstream of Dashankou hydrological gauge is 19 012 km^2. Precipitation and air temperature over the basin are distributed unevenly. Abundant precipitation falls in the mountains and little on the plain.

Fig. 1 Position of Kaidu River basin and meteorological station information (basin area: 19 012 km^2).

Tropical Rainfall Measuring Mission precipitation data

The TRMM satellite was launched in November 1997, and a number of experimental, real-time data sets based on the TRMM products and other satellite sources are currently available (Huffman *et al.*, 1995, 2007; Huffman, 1997), including the 3B42RT product, which is a merger of the 3B40RT and 3B41RT products. The 3B40RT product is a merger of all available Special Sensor Microwave/Imager (SSM/I) and TMI precipitation estimates. The SSM/I data are calibrated to the TMI using separate global land and ocean matched histograms. The 3B41RT product consists of precipitation estimates from geostationary infrared (geo-IR) observations using spatially and temporally varying calibrations by the 3B40RT product. The 3B43 product used in this study describes gridded estimates on a calendar month temporal resolution and a 0.25-degree by 0.25-degree spatial resolution spanning 50°N–S latitudes globally.

National meteorological information centre

In this study, 753 meteorological stations covering most regions of mainland China for the period 1951–2008 were investigated. The data set contains daily precipitation, daily maximum and daily

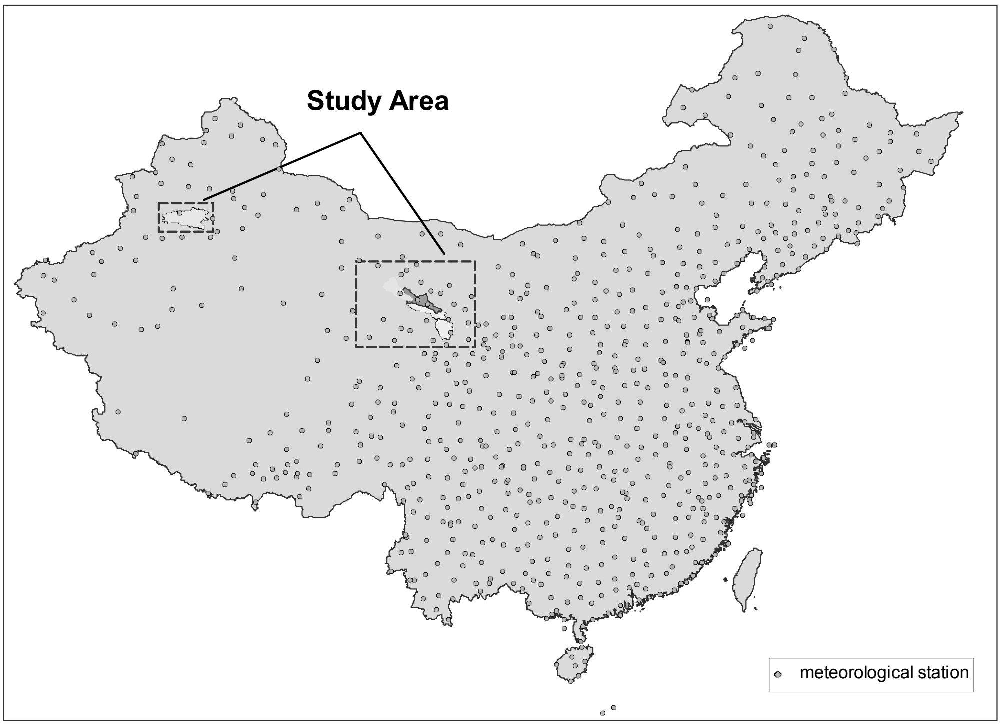

Fig. 2 Location of study area and 753 meteorological stations over China.

minimum air temperature derived from the China Meteorological Data Sharing Service system. There is only one meteorological station (Bayinbuluke station) in the Kaidu River basin.

Hydrology model and observation inputs

The Variable Infiltration Capacity (VIC) model (Liang *et al.*, 1994, 1996) was used to evaluate the method. It was run using combined precipitation from TRMM and gauge observed rainfall over the Kaidu River basin. VIC is a macroscale hydrological model that solves full water and energy balances, and was originally developed by Xu Liang at the University of Washington, USA (Liang *et al.*, 1994, 1996). VIC has been applied, in its various forms, to many USA watersheds including the Columbia River, the Ohio River, the Arkansas-Red rivers, and the Upper Mississippi rivers, as well as being applied globally (Abdulla *et al.*, 1996; Cherkauer & Lettenmaier, 1999; Hamlet & Lettenmaier, 1999; Liang & Xie, 2001). The model was designed both for inclusion in GCMs as a land–atmosphere transfer scheme, and for use as a stand-alone macroscale hydrology model.

The VIC model was set up over the Kaidu River basin at a resolution of 30 km. The model domain consists of 39 computational grid cells. A 30 km × 30 km river network based on a 1-km DEM was developed over the entire Kaidu River basin for the purpose of defining the model's river routing scheme using the method of Lohmann *et al.* (1996, 1998) which takes daily VIC surface and subsurface runoff as input to obtain model simulated streamflows at the basin outlet.

Development of the forcing data set

The development of the forcing data set progressed through a number of steps in terms of the spatial and temporal resolution and the sophistication of the interpolation methods. The following steps describe the method to estimate the spatial distribution of precipitation for a data-sparse area with TRMM data and a small number of available recorded rainfall data.

Step 1 Verify the correlation coefficient (*R*) of precipitation between monthly TRMM and gauge observations To decide if grid cells need to be corrected or not, an *R* statistical test was carried out to determine whether the TRMM data set is statistically similar to the observed data set or not:

$$R = \frac{\sum_{i=1}^{n}(T_i - \bar{T})(G_i - \bar{G})}{\sqrt{\sum_{i=1}^{n}(T_i - \bar{T})^2 \sum_{i=1}^{n}(G_i - \bar{G})^2}} \tag{1}$$

where T_i and G_i are the TRMM and gauge precipitation at time step i, respectively, and $\bar{T}$ and $\bar{G}$ are the mean TRMM and gauge precipitation.

Statistics of the correlation coefficient between monthly rainfall of Bayinbuluke and TRMM grid cells over the basin were calculated. The results show that the R of all 39 grid cells are larger than 0.88, 76% are larger than 0.90. It was assumed that all of the grid cell rainfalls have frequency distributions similar to the Bayinbuluke gauge which is considered representative of the basin. Therefore, based on the observations of the Bayinbuluke gauge, the spatial distribution of rainfall can be estimated.

Table 1 Description of the data sets.

Site	Data set description
Bayinbuluke	1981–1987, daily precipitation, daily Tmax, Tmin 1998–2007, monthly precipitation
TRMM PR	1998–2007, monthly precipitation, global 50°S–50°N, 0.25° × 0.25°
Dashankou	1981–1987, daily streamflow

Step 2 Calculate the correction ratio This step calculates the correction ratio of all grid cells in the basin. The aim of the correction is to force the rainfall of the TRMM to match that of the gauge observations. The correction ratio is calculated based on the algorithm below:

$$x_{i,p} = x_{o,p} \cdot (\overline{x_i} / \overline{x_0}). \quad (i = 1,2,3,\ldots, n) \tag{2}$$

where x_i represents the TRMM precipitation of grid cell which to is be corrected, x_0 represents the TRMM precipitation of the grid cell which has the nearest distance to the observation gauge, $x_{0,p}$ is the gauge observed precipitation of grid cell (In this case, $x_{0,p}$ is the rainfall of Bayinbuluke station). $\overline{x_i} / \overline{x_0}$ denoted as λ.

Commonly, basin climate is described by long-term averages. So in this study, we used the mean monthly precipitation of TRMM for every year to calculate the correction ratio λ:

$$\lambda_{i,j} = \overline{x_{i,j}} / \overline{x_{0,j}}. \quad (i = 1,2,\ldots,n; j = 1,2,\ldots,12) \tag{3}$$

where $\overline{x_{i,j}}$ represents the grid cell i, the j monthly mean precipitation, $\overline{x_{0,j}}$ represents the grid cell in which the observation gauge was located, the j monthly mean precipitation, $\lambda_{i,j}$ represents the grid cell i, and the j monthly correction ratio. This algorithm utilized the distributed correlation over the basin. As for the multi-year average, the average correlation over basins is generally steady or changes in small ranges. It presents much more precipitation on the windward slope than on the leeward slope. It estimates precipitation on the basis of the small amount of available recorded rainfall data by using analogous relations of TRMM precipitation data over the basin.

RESULTS AND ANALYSIS

The meteorological input data (forcing data) for the VIC model include daily precipitation, maximum temperature (Tmax), minimum temperature (Tmin). Tmax and Tmin are calculated on the basis of observed data according to the environmental adiabatic lapse rate of –0.65°C/100 m of elevation ascent.

The VIC model was used to test the effect of the correction method on the two data sets of streamflow simulations. The average relative error, *Er* and the Nash-Sutcliffe coefficient of efficiency (Nash & Sutcliffe, 1970), *NSE*, were chosen as assessment criteria:

Average relative error: $$Er = \left(\sum_{i}^{n} Q_C - \sum_{i}^{n} Q_O\right) / \sum_{i}^{n} Q_O \qquad (4)$$

where Q_c is the simulated discharge, and Q_o is the observed discharge.

Nash-Sutcliffe Coefficient of Efficiency: $$NSE = 1 - \sum_{i}^{n} (Q_{i,c} - Q_{i,o})^2 / \sum_{i}^{n} (Q_{i,o} - \overline{Q_o})^2 \qquad (5)$$

where $Q_{i,c}$ is the simulated discharge, $Q_{i,o}$ is the observed discharge, Q_0 is the mean observed discharge over the entire period, all in m^3 s^{-1}; *n* is the number of daily or monthly precipitation or stream flow pairs in the analysis. The Nash-Sutcliffe efficiency measures the goodness of fit using a ratio of error variance to the variance of the observation where 1 represents a perfect match, and smaller and negative values represents worse matches.

The VIC model was driven by the proposed corrected daily forcing and raw gauged precipitation over the entire five river basins (Dashankou, Xining, Jiayuguan, Yingluoxia and Qingshizui) in western China. VIC has seven user-calibrated hydrological parameters shown in Table 2. We keep the thickness of the first soil moisture layer constant (*d*1 = 0.1 m), and use basin observed hydrographs to calibrate the model because they reflect the integrated basin hydrological response. We use observed daily hydrographs from five catchments with drainage areas varying from 6810 km^2 to 18 730 km^2. We use an auto-optimization procedure base on Rosenbrock (1960) for calibration. The optimization procedure used the *NSE* criterion as the objective function. After parameters were calibrated, the effects of land surface water budget were measured by using the *Er* and *NSE* criteria.

Simulations over Kaidu River basin and the other four sites are shown in Fig. 2: monthly streamflow was accumulated from daily flow. Table 3 indicates that the result using the corrected data set has a significant improvement compared to the raw data set. For Kaidu River basin, the average relative error, *Er* and the Nash-Sutcliffe coefficient of efficiency, *NSE*, were –0.35 and 0.10 respectively, for the raw data set; the corresponding values for the corrected data set were –0.09 and 0.61. The results show an improvement from using the corrected precipitation estimates to drive the hydrological model.

Table 2 The seven VIC user-calibrated hydrological parameters.

Parameter	Units	Description
b_infilt	N/A	variable infiltration curve
Dsmax	mm/day	maximum velocity of baseflow
Ds	fraction	fraction of Dsmax where non-linear baseflow begins
Ws	fraction	fraction of maximum soil moisture where non-linear baseflow occurs
*d*1	m	thickness of the first soil moisture layer
*d*2	m	thickness of the second soil moisture layer
*d*3	m	thickness of the third soil moisture layer

Table 3 Results of the hydrological simulation with raw and corrected precipitation over five basins in western China.

Sites	Dashankou		Xinning		Jiayuguan		Yingluoxia		Qingshizui	
	Raw	Corrected	Raw	Corrected	Raw	Corrected	Raw	Corrected	Raw	Corrected
Er (%)	–35	–8.9	–8.9	–0.50	0.90	2.70	–11	–2.1	–17	–4.9
Daily *NSE*	–0.28	0.44	0.23	0.46	0.07	0.28	0.74	0.78	0.53	0.69
Monthly *NSE*	0.11	0.61	0.53	0.68	0.32	0.41	0.82	0.89	0.76	0.86

Er: Quantitative accuracy coefficient of efficiency; *NSE*: Nash-Sutcliffe efficiency; Raw: represents the simulation using raw precipitation; Corrected: represents the simulation using precipitation interpolated with the proposed method.

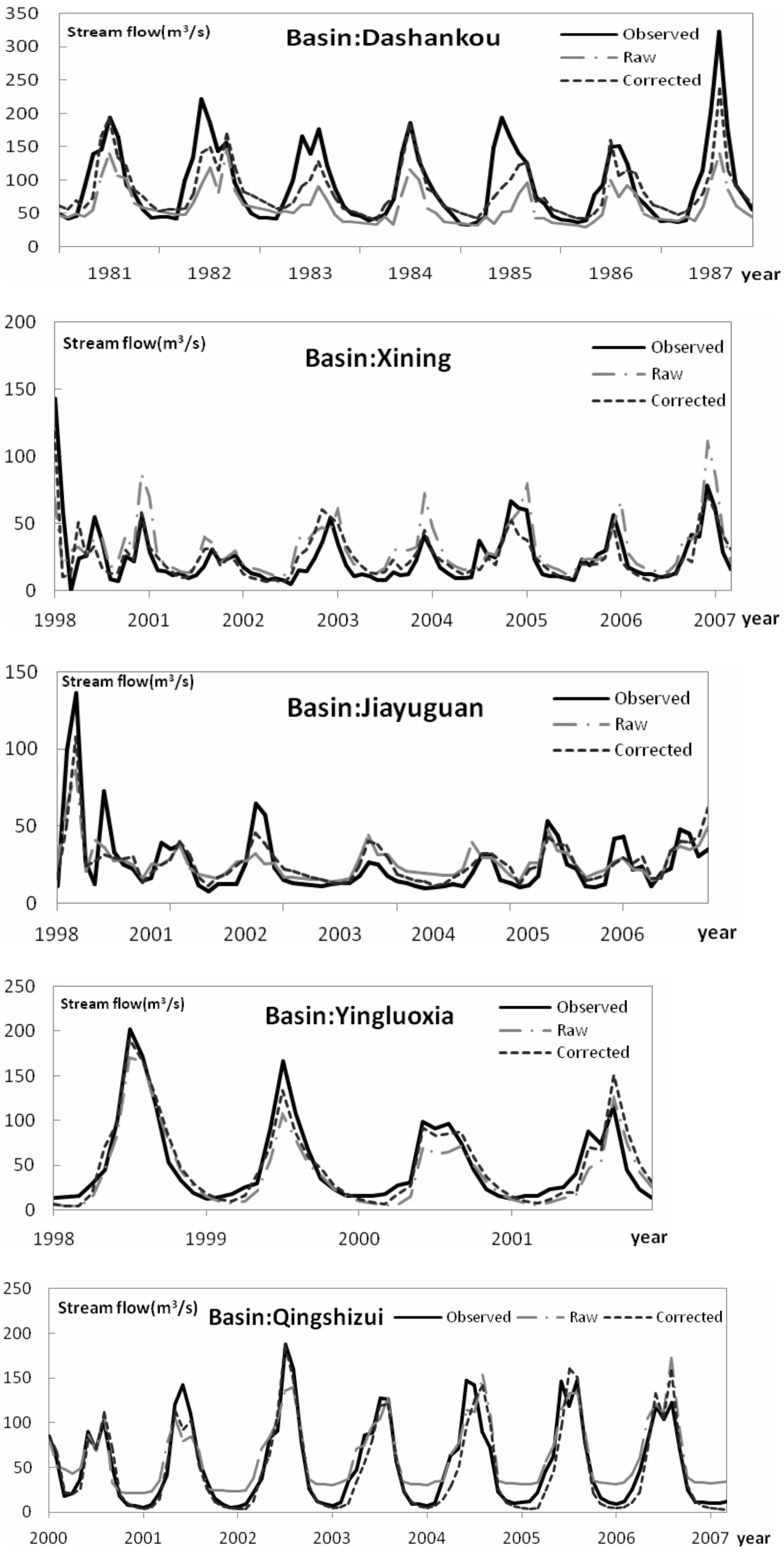

Fig. 3 Average monthly discharge of the five study sites in western China (Dashankou, Xining, Jiayuguan, Yingluoxia and Qingshizui).

SUMMARY AND CONCLUSIONS

Precipitation is the single most uncertain input in hydrological simulations over poorly-gauged areas. In this research, we have improved the forcing data set by using TRMM precipitation data to drive the VIC macro-scale land hydrological model to simulate the streamflow of five data-sparse basins in western China. The developed data set was based on a small number of gauge observations combined with TRMM precipitation, and provided an improvement compared to uncorrected precipitation from gauges alone. The results show that hydrological simulation using the VIC land hydrological model and calibrated parameters over western China is feasible. This method can be used for initial hydrological simulation and water resources evaluation in data-sparse regions.

At the same time, some simulations, such as of the wet season of 1983 and 1987 of the Kaidu River basin were poor. We consider this to be due to the original data set which had only one gauge in the basin. The lack of data resulted in the poor simulations. Improvements in simulation may be possible by acquiring more observed data, but this is the subject of further study.

Acknowledgements This work was supported by the National Key Technology R&D Program (Grants 2006BAC05B02 and 2007BAC03A060301), National “948” project (Grant 200758) and the Special Fund for Public Industry of the Ministry of Water Resources (Grant 200701039).

REFERENCES

Abdulla, F. A., Lettenmaier, D. P., Wood, E. F. & Smith, J. A. (1996) Application of a macroscale hydrologic model to estimate the water balance of the Arkansas-Red river basin. *J. Geophys. Res.* **101**, 7449–7459.

Cherkauer, K. A. & Lettenmaier, D. P. (1999) Hydrologic effects of frozen soils in the Upper Mississippi River basin. *J. Geophys. Res.* **104**, 19599–19610.

Hamlet, A. F. & Lettenmaier, D. P. (1999) Effects of climate change on hydrology and water resources in the Columbia River Basin. *Am. Water Res. Assoc.* **35**, 1597–1623.

Huffman, G. J. (1997) Estimates of root-mean-square random error for finite samples of estimated precipitation. *J. Appl. Met.* 1191–1201.

Huffman, G. J., Adler, R. F., Bolvin, D. T., Gu, E. J. N. & Bowman, K. P. (2007) The TRMM multi-satellite precipitation analysis: quasi-global, multi-year, combined-sensor precipitation estimates at fine scale. *J. Hydromet.* **8**, 38–55.

Huffman, G. J., Adler, R. F., Rudolf, B., Schueider, U. & Keehn, P. R. (1995) Global precipitation estimates based on a technique for combining satellite-based estimates, rain gauge analysis, and NWP model precipitation information. *J. Climate*, **8**, 1284–1295.

Lohmann, D., Nolte-Holube, R. & Raschke, E. (1996) A large-scale horizontal routing model to be coupled to land surface parameterization schemes. *Tellus A* **48**, 708–721.

Lohmann, D., Raschke, E., Nijssen, B. & lettenmaier, D. P. (1998) Regional scale hydrology: I. Formulation of the VIC-2L model coupled to a routing model. *Hydrol. Sci. J.* **43**, 131–141.

Liang, X. & Xie, Z. H. (2001) A new surface runoff parameterization with subgrid-scale soil heterogeneity for land surface models. *Adv. Water Res.* **24**, 1173–1193.

Liang X., Lettenmaier, D. P. & Wood, E. F. (1994) A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *J. Geophys. Res.* **99**, 14415–14428.

Liang X., Lettenmaier, D. P. & Wood, E. F. (1996) surface soil moisture parameterization of the VIC-2L model: Evaluation and modification. *Global Planet. Change* **13**, 195–206.

Nash, J. E. & Sutcliffe, J. V. (1970) River flow forecasting through conceptual models: Part 1, A discussion of principles. *J. Hydrol.* **10**, 282–290.

Rosenbrock, H. H. (1960) An automatic method for finding the greatest or least value of function. *The Computer J.* **3**, 175–184.

Seiji, H., Shogo, M., Masataka, W. & Baohua, X. (2004) Simulation of water runoff using estimated global precipitation data: Taking the upper reaches of the Changjiang River as an Example. *Acta Geographica Sinica* **59**, 125–135.

Wang, J. H., Jiang, D., Wang, H. & Qin, D. Y. (2003) Retrieval of annual precipitation in Yellow River by RS. *Resources Sci.* **25**, 8–13.

Zhou, Z. H., Jia, Y. W., Wang, H., Zhu, H. H., Wang, J. H. & Yan, D. H. (2006) Interpolating precipitation in space and time in large-scale basin based on rain gauges. *J. China Hydrol.* **2006**, 6–11.

Zhang, J. Y. & He, H. (1998) Study on runoff modeling in non-data watershed using geo-referenced data. *Adv. Water Sci.* **9**, 345–350.