A large-scale routing scheme for stream simulation and its application to river basins in China

FENG CHEN[1,2] & ZHENGHUI XIE[1]

1 *Institute of Atmospheric and Physics, Chinese Academy of Sciences, Beijing 100029, China*
zxie@lasg.iap.ac.cn

2 *Graduate School of Chinese Academy of Sciences, Beijing, China*

Abstract This paper presents a routing scheme for large-scale river basin stream simulation, with a grid flow direction scheme and the Muskingum-Cunge method considering lateral flow into the river. The grid flow direction scheme is to identify the grid flow directions by incorporating the GTOPO30 Digital Elevation Model (DEM) data, the multi-branch tree method, and the "stream burn" method. The routing scheme was applied to route runoff at river basins derived from the land surface model VIC-3L with the regionalized parameters at the resolution 0.5°×0.5° in continental China into streamflow at hydrological stations. The results show that the created grid flow directions are reasonable, and the continental simulation results agree well with the observations on a large river basin scale in China.

Key words DEM; flow direction; large-scale routing scheme; Muskingum-Cunge method; Pfafstetter basin numbering system

INTRODUCTION

The land component of climate models represents many processes that control exchanges of energy, momentum and water between soil, vegetation and the atmosphere, and has long been recognized as important for hydrological forecasting such as streamflow prediction and climate change studies (Dickinson *et al.*, 1993; Bonan *et al.*, 1996; Nijssen *et al.*, 1997; Zeng *et al.*, 2002; Dai *et al.*, 2003). Streamflow is derived from runoff with a routing scheme, which can be used in evaluating model efficiency by meteorologists and predicting water resources and disasters by hydrologists (Saunders, 1999). Therefore, a suitable large-scale routing scheme for streamflow simulation by a land surface model is necessary for streamflow simulation.

A routing scheme consists of grid flow directions identification, overland flow concentration and river routing. To identify the flow directions, there has been a strong research effort toward resolving a satisfactory method for determining flow directions in large-scale models (O'Donnel *et al.*, 1999; Fekete *et al.*, 2001; Turcotte *et al.*, 2001; Hao *et al.*, 2002; Olivera *et al.*, 2002; Shaw *et al.*, 2003, 2005; Ye *et al.*, 2005). There have been two main methodologies used in up-scaling hydrological routing information (Shaw *et al.*, 2005). The first method attempts to consider all of the topographic information available for each sub-grid from a more detailed DEM in up-scaling routing patterns. Although this method can produce very good results with flow directions reflecting the actual drainage pattern of the basin, it is very computationally expensive and can require costly software (Shaw *et al.*, 2003). The second approach uses average elevations, determined from a more detailed DEM, for each grid cell to route flow between neighbouring grid cells. This approach is computationally efficient, but does not produce adequate representation of the channel network of the basin (Arora & Boer, 1999; Shaw *et al.*, 2003).

This paper presents a new routing scheme for large-scale river basin stream simulation, with a grid flow direction scheme and the Muskingum-Cunge method considering lateral flow into the river. The flow direction scheme is to identify the flow directions based on the principle of multi-branch tree generation (Weiss, 1997), which is widely used in computer sciences, and the stream burn method (Hutchinson, 1989; Hao, 2002) in which the actual river information is combined with the DEM data in order to reduce the bias between the channel structure of a basin derived from DEM and the actual drainage patterns. Streamflow simulation for river basins in continental China by the routing scheme and the land surface model VIC (Liang *et al.*, 1994, 1996; Liang & Xie, 2001) with the regional parameters at the resolution 0.5° × 0.5° (Xie *et al.*, 2007) is given.

METHODOLOGY

As mentioned above, a grid flow direction identification method and a routing algorithm are the two key components of a routing scheme. We identify the flow directions and extract the drainage network for a region by the multi-branch tree method and the stream burn method. Based on the extracted flow directions and drainage network, the large-scale basin is sub-divided into small sub-basins and numbered following the Pfafstetter basin numbering system (Yang *et al.*, 2000). The daily runoff at a grid produced by a land surface model is convolved with a triangular unit hydrograph as the overland flow concentration, and the river routing is calculated by the extracted channel network and Muskingum-Cunge method considering lateral flow into the river. The schematic diagram of the routing scheme is shown in Fig. 1.

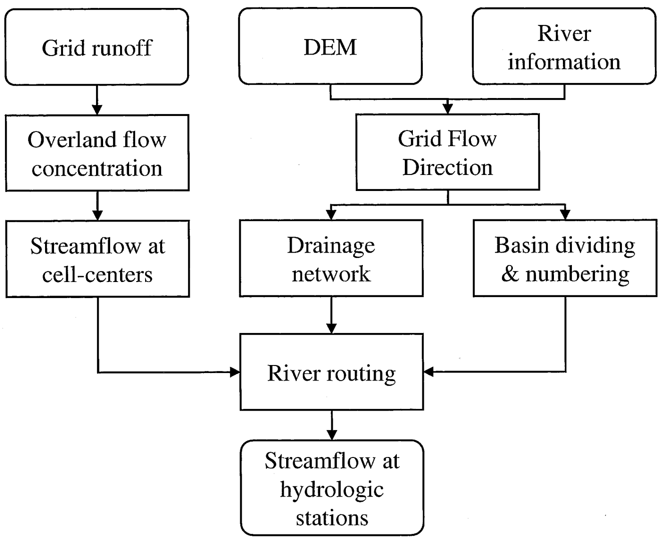

Fig. 1 The schematic diagram of the routing scheme.

Grid flow direction identification

By using the DEM data, many important hydrological characteristics can be gained, such as slope rate, slope direction, drainage network, basin boundary (Saunders, 1999). Since DEMs only describe the cell centre's elevation or the cell average elevation, the channel structure of a basin derived from a DEM may not adequately reflect actual drainage patterns on a large-scale, which is from the discontinuous river net, the sinks in the DEM, and the parallel river net in smooth regions. For each cell in a basin, flow out direction at a cell is assigned to the lowest neighbour cell by examining the elevation of itself and the eight surrounding cells according to the D8 algorithm (O'Callaghan, 1984). It performs well at a sufficiently fine resolution; however, the drainage network extracted by the D8 algorithm is less reliable when the resolution of the DEM becomes coarser. We use the multi-branch tree method and stream burn ideology (Hutchinson, 1989; Hao, 2002) to extract a continuous, no-sink and unparallel drainage network.

The streamflow at the outlet of a basin comes from runoff at grid cells for the basin. Each grid cell in the basin can be considered as a node in the multi-branch tree; the multi-branch tree contains the entire cells in the basin, and the outlet of the basin as the root node. Every father node contains a subset of the cells whose flow is assigned to the father node cell. Once the flow out direction at a cell is identified, the node associated with the cell is added into the multi-branch tree at its father node. The grid flow direction identification for a basin is reduced to construct the multi-branch tree. For the outlet of a basin i.e. the root node of the multi-branch tree for the basin, is called the first layer node. The flow out directions for neighbours of the first layer node are identified by D8 algorithm, and those neighbour nodes whose flow out directions are assigned to the first layer node are added into the tree as the son nodes of the root node. For constructing the (m + 1)th layer nodes of the tree, the flow out directions for the nodes, which are not included in the m layers but at least one of their neighbours are identified by D8 algorithm, and those nodes whose flow out directions are assigned to the mth layer nodes are added into the tree as the son nodes of the mth layer nodes. There may still be cells whose flow directions are not identified after searching all the nodes in the tree, which means that the DEM has sinks. In order to solve this problem, we find the lowest cell among the cells whose neighbour has been in the tree, and assign the flow to its lowest neighbour that has been in the tree. When all the sinks are well treated, the grid flow directions can be identified. For all the nodes in the multi-branch tree connected with the root tree, the drainage network is continuous. Figure 2 shows the drainage network (the resolution of the DEM is 0.1°×0.1°). It is seen from Fig. 2(b) that the river is discontinuous in the circles by comparing the actual drainage network. Figure 2(c) shows the continuous drainage network extracted by using the multi-branch tree method.

However, the extracted drainage network is incorrect in the circle in Fig. 2(c). The main reason for this problem is that there are little differences between the cell elevations. In order to reduce the bias between the drainage network derived from DEM and the actual one, we use the stream burn method (Hutchinson, 1989; Hao, 2002) to include the natural river information into the DEM data. We can get the river DEM based on GIS, and the new DEM by the original DEM minus the river DEM. The drainage network, including the natural river information, is better than the original one (Fig. 2(d)).

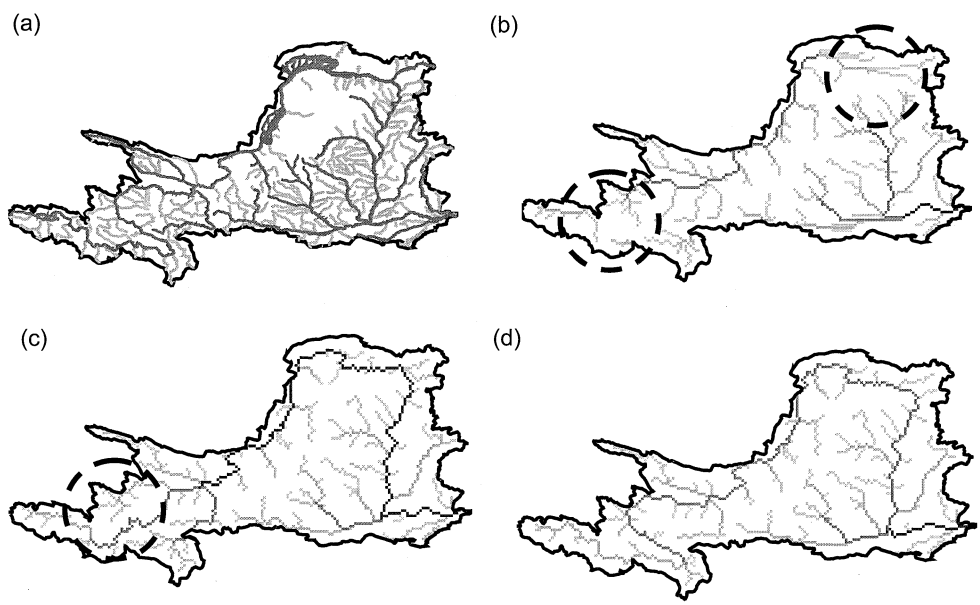

Fig. 2 The drainage network of the Yellow basin (above the Huayuankou station): (a) the actual drainage network; (b) the drainage network extracted from the original DEM by using Arc/Info analytical software; (c) the drainage network extracted from the original DEM by using the multi-branch tree method; (d) the drainage network extracted from the new stream-burned DEM by using the multi-branch tree method.

Based on the multi-branch tree method and the stream burn method, the grid flow directions for all grid cells in the region are identified, the extracted drainage network is continuous, and every cell is connected with the outlet. Otherwise, the water in sinks always choses the lower way to flow out.

River and basin creating

The river and basin created is based on the flow directions and the accumulated flow matrix, which is calculated by the number of cells flowing into the cell (Ye *et al.*, 2005). We suppose that the river channel is created in the cells whose accumulated flow values are bigger than other cells nearby. The whole river (or basin) can be identified by searching the flow direction matrix and the accumulated flow matrix when the outlet cell of a river (or basin) is found.

Basin partition

To represent the hydrological process well, the large basin is sub-divided into several sub-basins by the method which always divides the basin into nine sub-basins, and numbering them with the Pfafstetter basin numbering system (Yang *et al.*, 2000). According to the Pfafstetter basin numbering system, the whole basin can be divided into several sub-basins, and the relationship between any two sub-basins can be easily found from the numbering system. For example, the sub-basin labelled XYZ, where XY is the upper basin-level's serial number, and Z is the serial number of the current basin-level. Figure 3 shows the relationship between the Yellow River basin (above the Huayuankou station) and its sub-basins.

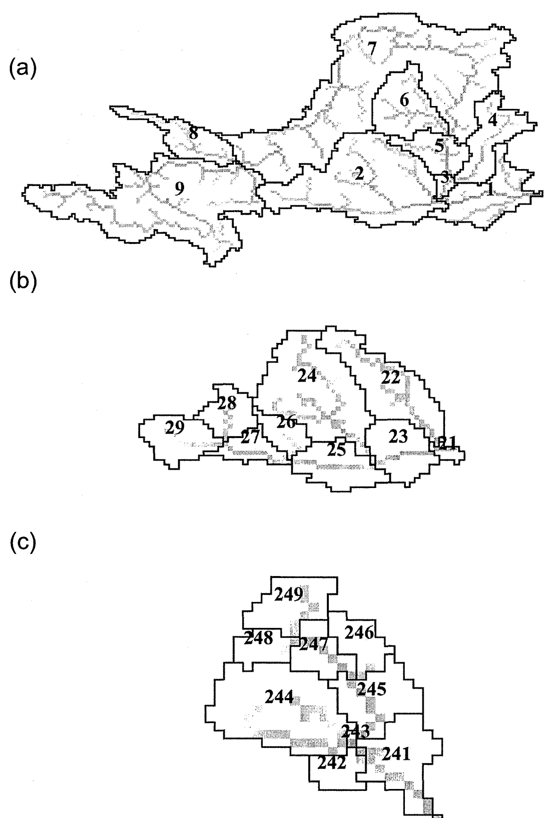

Fig. 3 The relationship between Yellow river basin (above the Huayuankou station) and its sub-basins: (a) the first level coding; (b) the second level coding of the sub-basin 2 in the first level; (c) the third level coding of the sub-basin 24 in the second level.

Overland flow concentration

Although the grid-based land surface model produces only one time series of runoff values for each grid cell, this runoff should not be considered as being produced at a single point, but as being distributed non-uniformly over the area of that grid cell (Nijssen *et al.*, 1997). To account for differences in travel time of runoff produced in different parts of the grid cell, the daily runoff produced by a land surface model is convolved with a triangular unit hydrograph, which simulates routing within the grid cell.

River routing

We suppose that the hydrographs produced for each grid cell in a basin are then summed together and distributed uniformly along the trunk stream and routed to the basin outlet. The Muskingum-Cunge method considering lateral flow into the river (Cunge, 1969; Price *et al.*, 1975; Rui *et al.*, 1995; Rui *et al.*, 2000; Yuan *et al.*, 2005) is applied to flood routing to obtain the outlet's discharge.

APPLICATION

Based on the runoff produced by Xie *et al.* (2007), which presented a methodology for regional parameter estimation of the VIC-3L land surface model and simulated the streamflow in continental China, we chose eight hydraulic stations located in five large basins for verification.

The detailed information for the stations and the verification results can be seen in Table 1. Compared with the actual control area, the relative errors of the simulated control area of the eight basins are –3.1%, 3.9%, 0.9%, 6.5%, 2.5%, 7.5%, 11.0% and 1.5%, respectively. The main reason for the bias is that the resolution is too coarse in our study, and the boundary of basins could not be so regular. Generally, large basins will have a small relative error while the small basins will have a large one.

Table 1 The information of eights stations and the verification results.

River basin	Station	Actual control area (10^4 m^2)	Simulated control area (10^4 m^2)	Area bias[a] (%)	Bias[b] (%)	RRMSE[c] (%)	Ce[d]	Simulation period
Pearl River	Wuzhou	32.7	31.7	–3.1	21.0	2.7	0.82	1/1/1980–31/12/1990
Yangtze River	Yichang	100.6	104.5	3.9	0.8	2.1	0.89	1/1/1980–31/12/1990
	Datong	170.5	169.0	–0.9	6.2	1.4	0.88	1/1/1980–31/12/1990
Huaihe River	Bengbu	12.1	12.9	6.6	41.4	5.9	0.66	1/1/1980–31/12/1990
Yellow River	Tangnaihai	12.2	11.9	2.5	–14.6	3.5	0.81	1/1/1980–31/12/1989
	Guide	13.4	14.4	7.5	0.8	4.3	0.67	1/1/1980–31/12/1989
Songhuajiang River	Jiangqiao	16.3	18.1	11.0	–17.8	5.4	0.74	1/1/1980–31/12/1990
	Haerbin	39.0	39.6	1.5	–8.2	5.1	0.54	1/1/1980–31/12/1990

[a] Area bias – the bias of the control area between simulated and actual: $bias = (A_s - A_o)/A_o \times 100\%$

[b] bias - defined as: $bias = (\overline{Q}_s - \overline{Q}_o)/\overline{Q}_o \times 100\%$

[c] RRMSE – the relative root mean squared error, defined as: $RRMSE = \frac{1}{n}\sqrt{\sum_{i=1}^{n}(Q_{s,i} - Q_{o,i})^2} / \overline{Q}_o \times 100\%$

[d] CE - the Nash-Sutcliffe model efficiency coefficient, defined as:

$CE = (\sum_{i=1}^{n}(Q_{o,i} - \overline{Q}_o)^2 - \sum_{i=1}^{n}(Q_{s,i} - Q_{o,i})^2) / \sum_{i=1}^{n}(Q_{o,i} - \overline{Q}_o)^2 \times 100\%$, with $Q_{s,i}$ and $Q_{o,i}$ the simulated and observed flows in month i.

Figures 4 and 5 show the observed and simulated monthly streamflow time series and the mean monthly hydrographs for the eight large basins, respectively. The routing scheme performed well based on the runoff produced by the VIC-3L model with the transferred parameters, and the simulated streamflows matched the observed values well. Although the relative errors of some stations are not very good, four in eight basins' relative errors are less than 10%, and the smallest is 0.8%. All the relative root mean squared errors are <6%, which shows small amplitude of the difference between the simulated and observed streamflows. Five in eight basins' Nash-Sutcliffe coefficients are larger than 0.7, and four basins' coefficients are larger than 0.8.

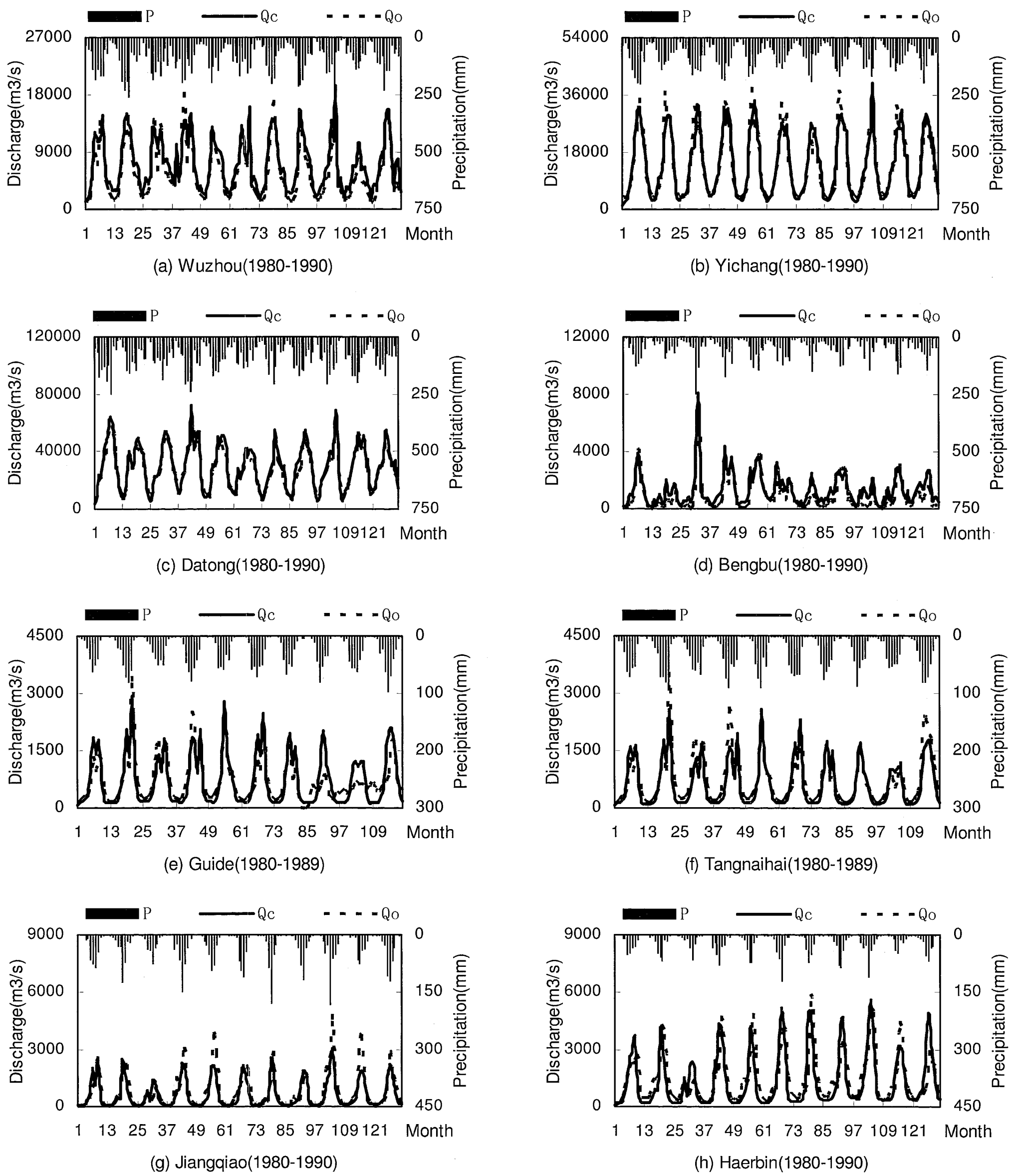

Fig. 4 The observed and the simulated monthly streamflow time series for the eight verification basins.

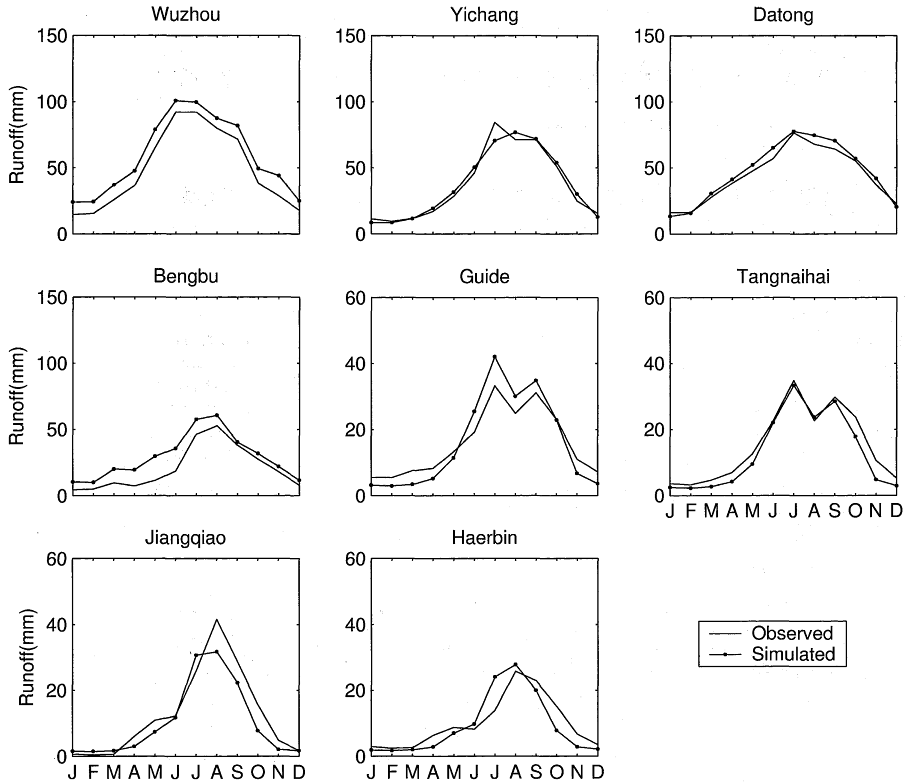

Fig. 5 Mean monthly hydrographs of observed and simulated flows for the eight verification basins.

CONCLUSIONS

In this paper, a routing scheme for large-scale river basin streamflow simulation was presented. Based on the D8 algorithm, the multi-branch tree method and the stream burn method were combined to identify the grid flow directions. The grid cell daily runoffs produced by the VIC-3L model with transferred parameters, were routed to the hydrological stations by this routing scheme, and then compared to the monthly-observed streamflows at the corresponding basins.

The results show that the drainage network extracted from DEM matches up well to the actual one, and the simulated streamflows matched the observed values well. Although the relative errors of some stations are not very good, the bias may not only be introduced by the routing scheme, but also by the model uncertainty and the parameter uncertainty. In general, the routing scheme performs well in a large-scale stream simulation and can be applied to the large-scale routing.

Acknowledgements This work was supported by the National Natural Science Foundation of China under Grant no. 90411007, the Knowledge Innovation Project of Chinese Academy of Sciences under Grant nos. KZCX2-YW-217 and KZCX2-YW-126-2, the National Basic Research Program under the Grant 2005CB321704, and CAS International Partnership Creative Group "The Climate System Model Development and Application Studies".

REFERENCES

Arora, V. K. & Boer, G. J. (1999) A variable velocity flow routing algorithm for GCMs. *J. Geophys. Res.* **104**(D24), 30965–30979.

Bonan, G. B. (1996) A Land Surface Model (LSM version 1.0) for ecological, hydrological, and atmospheric studies: technical description and user's guide. NCAR Technical Note NCAR/TN-417+STR. National Center for Atmospheric Research. Boulder, Colorado, USA.

Cunge, K. A. (1969) On the subject of a flood propagation method (Muskingum Method). *J. Hydraul. Res.* **7**(2), 205–230.

Dai, Y. J., Zeng, X. B., Dickinson, R. E., Baker, I., Bonan, G. B., Bosilovich, M. G., Denning, A. S., Dirmeyer, P. A., Houser, P. R., Niu, G. Y., Oleson, K. W., Schlosser, C. A. & Yang Z. L. (2003) The common land model. *Bull. Am. Met. Soc.* **84**(8), 1013–1023.

Dickinson, R. E., Henderson S. A. & Kennedy P. J. (1993) Biosphere–Atmosphere Transfer Scheme (BATS) version 1e as coupled to the NCAR Community Climate Model. NCAR technical note NCAR/TN-387+STR. National Center for Atmospheric Research, Boulder, Colorado, USA.

Fekete, B. M., Vörösmarty, C. J. & Lammers, R. B. (2001) Scaling gridded river networks for macroscale hydrology: development, analysis, and control of error. *Water Resour. Res.* **37**(7), 1955–1967.

Hao, Z. C. & Li, L. (2002) The method of creating of the net of river based on DEM. *Hydrology* **22**(4), 8–10 (in Chinese).

Hutchinson, M. F. (1989) A new procedure for gridding elevation and stream line data with automatic removal of spurious pits. *J. Hydrol.* **106**, 211–232.

Liang, X. (1994) A two-layer Variable Infiltration Capacity land surface representation for General Circulation Models. *Water Resour. Series* TR140, 208.

Liang, X., Lettenmaier, D. P. & Wood, E. F. (1996) One-dimensional statistical dynamic representation of subgrid spatial variability of precipitation in the two-layer Variable Infiltration Capacity model. *J. Geophys. Res.* **101**(D16), 403–422.

Liang, X. & Xie, Z. H. (2001) A new surface runoff parameterization with subgrid-scale soil heterogeneity for land surface models. *Adv. Water Resour.* **24**(9–10), 1173–1193.

Nijssen, B., Lettenmaier, D. P., Liang X., Wetzel, S. W. & Wood, E. F. (1997) Stream simulation for continental-scale river basins. *Water Resour. Res.* **33**(4), 711–724.

O'Callaghan J. F. & Mark D. M. (1984) The extraction of drainage networks from digital elevation data. *Computer Vision, Graphics, and Image Processing* **28**, 323–344.

O'Donnel, G., Nijssen, B. & Lettenmaier, D. P. (1999) A simple algorithm for generating streamflow networks for grid-based, macroscale hydrological models. *Hydrol. Processes* **13**(8), 1269–1275.

Olivera, F., Lear, M. S., Famiglietti, J. S. & Asante, K. (2002) Extracting low-resolution river networks from high-resolution digital elevation models. *Water Resour. Res.* **38**(11), 1231.

Price, R. K. & Harrison, A. J. M. (1975) *Flood Routing Studies*, 33–34. Natural Environment Research Council, London, UK.

Rui, X. F. (1995) *The Theory for Runoff Producing and Concentration*, 85–87. Water Resources and Electric Power Press Beijing, China (in Chinese).

Rui, X. F. & Wang, L. L. (2000) A research on flood routing method with prediction. *Adv. Water Sci.* **11**(3), 291–295 (in Chinese).

Saunders, W. (1999) Preparation of DEMs for use in environmental modeling analysis. *ESRI User Conference*, 24–30 July.

Shaw, D., Martz, L. W. & Pietroniro, A. (2003) A methodology for preserving channel flow networks and connectivity patterns in large-scale distributed hydrological models. *Hydrol. Processes* **19**(1), 149–168.

Shaw, D., Martza, L. W. & Pietroniro, A. (2005) Flow routing in large-scale models using vector addition. *J. Hydrol.* **307**, 38–47.

Turcotte, R., Fortin, J. P., Rousseau, A. N., Massicotte, S. & Villeneuve, J. P. (2001) Determination of the drainage structure of a watershed using a digital elevation model and a digital river lake network. *J. Hydrol.* **240**(3–4), 225–242.

Weiss, M. A. (1997) *Data Structures and Algorithm Analysis in C.* Addison-Wesley, Menlo Park, California, USA.

Xie, Z. H., Yuan, F., Duan, Q. Y., Liang, M. L., Zheng, J. & Chen, F. (2007) Regional parameter estimation of the VIC Land Surface Model: methodology and application to river basins in China. *J. Hydromet.* **8**(3), 447–468.

Yang, D. W., Musiake, K., Kanae, S. & Oki, T. (2000) Use of the Pfafstetter basin numbering system in hydrological modeling. In: *Proc. 2000 Annual Conference, Japan Society of Hydrology and Water Resources*, 200–2001.

Ye, A. Z., Xia, J., Wang, G. S. & Wang, X. N. (2005) Drainage network extraction and subcatchment delineation based on digital elevation model. *J. Hydrol. Engng* **36**(5), 531–537 (in Chinese)

Yuan, F., Ren, L. L., Yu, Z. B. & Xu, J. (2005) A river flow routing model based on digital drainage network. *J. Hydrodynamics* 17(4), 483–488.

Zeng, X. B., Muhammad, S., Dai, Y. J., Dickinson, R. E. & Myneni, R. (2002) Coupling of the common land model to the NCAR community climate model. *J. Climate* **15**(14), 1832–1854.