Available online at www.sciencedirect.com
SCIENCE DIRECT
Journal of Hydrology 319 (2006) 36–50
Journal of Hydrology
www.elsevier.com/locate/jhydrol

Input data resolution analysis for distributed hydrological modeling

Roshan Shrestha[a,*], Yasuto Tachikawa[b], Kaoru Takara[b]

[a]Department of Urban and Environmental Engineering, Graduate School of Engineering, Kyoto University, Kyoto, Japan
[b]Disaster Prevention Research Institute, Kyoto University, Kyoto, Japan

Received 13 August 2003; revised 22 April 2005; accepted 28 April 2005

Abstract

A distributed hydrological model is often needed to analyze spatially variable hydrologic behavior. Such a model can be difficult to set up, especially for an ungauged basin, as it demands a massive amount of data. Moreover, there is an additional challenge of selecting a proper grid resolution as the grid size selection generally leads to predictive uncertainty and also directly determines the amount of work required. In this study, a distributed macro-scale hydrological model, named as the MaScOD model, is applied with a 10-min spatial resolution to the Huaihe River basin, China, to simulate discharge at Bengbu (132,350 km^2) and at sub-basins at Wangjiaba (29,844 km^2) and at Suiping (2093 km^2). A range of input data resolutions are used, from 10 min to 2.5°, based on an experimental hydro-meteorological input data set abstracted from the GAME re-analysis data and the Hubex-IOP EEWB data. Performance of the model is evaluated by comparing observed discharge against simulated discharge for a range of IC-ratio values (the ratio between the input forcing resolution and the Catchment area). Similar results are obtained for all three catchments, despite their different sizes. It is found that improvement in distributed model performance is more pronounced below the IC-ratio 1:10, whereas the rate of improvement is negligible above the IC-ratio 1:20. The IC-ratio range 1:10–1:20 is found to be the optimum performance range considering the data and resource demands of distributed models. This may provide a preliminary criterion for selecting the scale for distributed hydrological modeling in ungauged basins.

Keywords: Input data resolution; Distributed hydrological modeling; Optimal model performance; IC-ratio; MaScOD model; Ungauged basins

1. Introduction

To fulfill the heavy demand for distributed data is a major challenge in distributed hydrological modeling, despite the superior ability of such models to analyze spatially variable hydrologic behavior and the impacts of natural and human activities on runoff (Refsgaard and Abott, 1996). Ungauged basins present the greatest challenge, as most ungauged basins have basically no hydro-meteorological data other than that from regional or global data sets obtained from reanalysis of a limited number of observations using a General Circulation Model (GCM) or a meso-scale numerical weather prediction model. Regional

* Corresponding author. Address: Fluvial and Marine Disaster Research Division, Disaster Prevention Research Institute, Kyoto University, Uji Campus, Gokasho, Uji, Kyoto 611-0011, Japan. Tel.: +81 774 38 4127; fax: +81 774 38 4130.
E-mail address: roshan@rdp.dpri.kyoto-u.ac.jp (R. Shrestha).

0022-1694/$ - see front matter © 2005 Elsevier Ltd All rights reserved.
doi:10.1016/j.jhydrol.2005.04.025

hydro-meteorological data, such as GAME-IOP 1.25° reanalysis data which covers the Asian Monsoon region (JMA, 2000), and HUBEX-IOP EEWB 5-min data which covers the Huaihe Basin in China (Kozan et al., 2001), can help compensate for the absence of ground-based hydrological instrumentation. However, the limitations of regional datasets may severely limit the accuracy in the hydrological simulation. Obtaining satisfactory results by using regional datasets depends on the resolution of the dataset and catchment scale because current regional datasets are often too coarse (Burlando and Rosso, 2002), they may not be adequate, in the case of small-scale hydrological modeling, to describe the variability in hydrologic process components at the basin scale. To understand the hydrologic predictive uncertainty associated with the gap between the need and the availability of input data for distributed hydrological modeling will provide guidance in setting up an appropriate hydrological modeling framework and on the selection of an appropriate data scale.

There is a choice to make regarding the required resolution of hydro-meteorological input data. The resolution of input data has a direct link with the modeling resolution in distributed hydrological modeling, since the model's resolution is often set equal to or finer than the input data. The grid cell size selection will generally lead to predictive uncertainty and the challenge is to determine a scale above which spatial variability can be neglected, with average characteristics of a given area providing sufficient information for accurate modeling of basin runoff (Sivapalan and Kalma, 1995; Blöschl and Sivapalan, 1995; Molnar and Julien, 2000). Coarser resolution hydro-meteorological datasets, such as outputs from currently available atmospheric models (which may extend to hundreds of kilometers) do not satisfy the need of hydrologists. Also, there are serious scale issues within hydrological analysis (Koren et al., 1999) and within meteorological analysis (Renssen et al., 2001), and these problems are essentially mismatched. The scaling issue assumes an even greater significance when developing regional or global hydrology models (Singh and Woolhiser, 2002) or in continental scale catchment modeling. The discrepancy in scale between meteorological models and hydrologic models will continue until reliable criteria emerge to provide guidance on the optimal scale for investigating hydrological processes.

The scale and resolution issues are raised here with a view to finding if there is a certain preferred or consensus scale of input data at which optimal performance may be feasible in distributed hydrologic modeling. Several researchers have investigated scale and resolution issues for distributed hydrologic modeling. For example, Bathurst's (1986) suggestion to divide the watershed into elements no larger than 1% of the total area was a conclusion from his study on the Wye watershed (10.55 km^2) using the SHE model, to ensure that each grid element was more or less homogeneous. Introducing the concept of a representative elementary area (REA), Wood et al. (1988) found that an REA of approximately 1 km^2 existed for the hydrologic response of the Coweeta watershed. The size of REA was more strongly influenced by basin topography than by rainfall length scales (Woods et al., 1995) and its limited utility are discussed by Fan and Bras (1995). Zhang and Montgomery (1994) proposed a 10 m grid size as a compromise between increasing spatial resolution and data handling requirements by examining the effect of digital elevation model grid size on the portrayal of the land surface and hydrological simulations. Bruneau et al. (1995) suggested an optimum region for modeling with a grid size of 50 m after analyzing the effect of space and time resolutions using TOPMODEL on the Coetdan Experimental watershed (12 km^2), in France. The simple scaling and multi-scaling framework (Gupta et al., 1994), the HRU (Hydrological Response Units) concept (Flügel, 1995), and the basin-scale model equations (Kavvas et al., 1998) provide understanding of scaling effects in distributed hydrologic modeling. Other research focusing on the effects of grid size on model parameters and hydrologic response include Quinn and Beven (1991), Franchini et al. (1996), Saulnier et al. (1997), Sunada et al. (2001), Horritt and Bates (2001), and Shrestha et al. (2002). However, despite these efforts, a suitable resolution for distributed hydrological modeling is still difficult to establish. Selecting a higher resolution in distributed hydrological modeling brings with it heavy tasks of data acquisition, defining the model parameter values, and complex calculations. These tasks increase the cost of modeling. In addition, higher resolution increases the risk of insistent error amplification. On the other hand, selecting a lower resolution greatly reduces the workload but risks losing the advantage of the distributed modeling approach, leading to poor results due to lack of consideration of important spatial features. Therefore, a choice of appropriate scale is prominently needed to attain optimum model performance. We consider that the first step in this process is the selection of an appropriate input data resolution. Knowledge of the effect of forcing input scale is important for both hydrological and meteorological studies. Ability to choose an adequate input resolution at the preliminary investigation stage will result in an appropriate modeling framework, with fewer problems later and higher simulation accuracy.

In this paper, a criterion for selection of an appropriate input data resolution is expressed in terms of the IC-ratio (the ratio between the model input spatial resolution and the area of the catchment; Shrestha et al., 2002), based on the sensitivity of a distributed hydrological model's performance to the scale of an experimental hydro-meteorological input dataset based on GAME-IOP reanalysis data and HUBEX-IOP EEWB data. Applying the MaScOD (Macro-Scale OHyMoS assisted Distributed) hydrological model with a fixed 10-min model resolution, the study is conducted on the Huaihe River Basin, China, by simulating the hydrographs at Bengbu, Wangjiaba and Suiping (having contributing areas of 132,350; 29,844 and 2093 km^2, respectively) for various spatial input data resolutions, from 10 min to 2.5°. In Section 2, the MaScOD macro-scale distributed hydrological model is described. Section 3 deals with the hydro-meteorological input data used in this study, and the discharge simulation results are given in Section 4. Section 5 provides an analysis of the results in terms of the IC-ratio. In Section 6, the selection of an appropriate input data grid resolution (in the context of a proposed IC-ratio Rule) is discussed, and conclusions are presented in Section 7.

2. The MaScOD macro-scale distributed hydrological model

Macro-scale hydrological modeling is often practiced for a large river basins. It is possible to incorporate preliminary tasks involved in the modeling, such as basin partitioning, hydrological process modeling for a sub-basin, linking sub-basin models together to make a total runoff model, and processing channel network linkages to incorporate river flow routing, into an automatic procedure (Tachikawa et al., 2002) with the assistance of an object oriented hydrological modeling system—OHyMoS (Takasao et al., 1996; Ichikawa et al., 2000). The "Macro-Scale OHyMoS assisted distributed hydrological model" (referred to hereafter as the MaScOD model) is developed accordingly, and is briefly presented here.

The MaScOD model subdivides a watershed basin into grid-cells according to a defined grid system that also facilitates the direct input of hydro-meteorological data from a meso-scale atmospheric model. This model considers the exact location and linkage of river segments within grid-cells to give better flow routing, resulting in a better simulation of the discharge hydrograph. An automated procedure rearranges the vector river networks (Fig. 1) by dividing up grid cell frames into separate sub-network elements and re-assigning new identities.

The MaScOD model consists of a MaScOD Element Model (MEM) on every grid cell. The total number of MEMs for the entire catchment depends on the size of the catchment, and the grid cell resolution (which is fixed at 10-min in this study). Each MEM, having a river segment inside the grid cell, contains a runoff process model (RPM), based on the Xinanjiang model (Zhao, 1992), and a flow routing model (FRM) based on the lumped stream kinematic-wave equation (Shiiba et al., 1996). The MEM can have multiple numbers of RPM inside the same grid cell when the grid cell contains multiple river segments. The total number of MEM constituents (the RPM and FRM) is determined on the basis of the river network rearrangement procedure. Each RPM receives input from its own sub-catchment, i.e. from the fraction of the grid cell defined as the proportion of the length of the RPM's river segment to the total length of river segments inside the grid cell. The RPM yields runoff directly from impervious area and from groundwater storage (Fig. 2). This is given by

$$R = Q_i + k_g S_g^2 \tag{1}$$

where R is the runoff; Q_i is the discharge from impervious area; k_g is the yield parameter; and S_g is

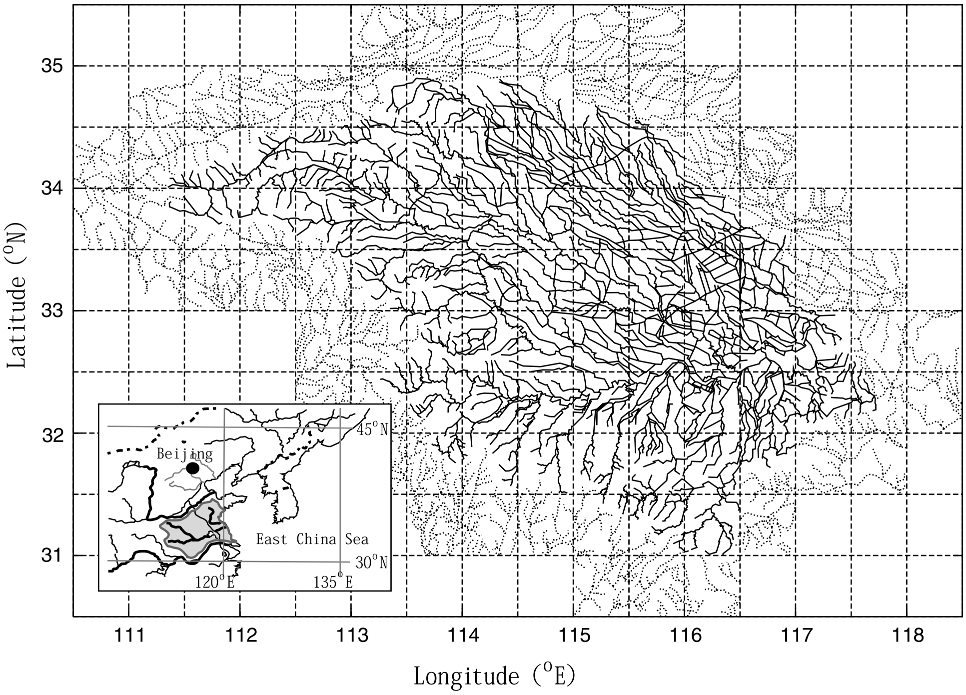

Fig. 1. River networks inside the study region, Huaihe River Basin, China.

the groundwater storage. The Q_i is given by

$$Q_\mathrm{i} = A_\mathrm{i}(P - E) \tag{2}$$

where A_i is the impervious proportion of a basin; P is the precipitation; and E is the evapo-transpiration. The groundwater storage S_g is continuously updated by the discharge from the pervious portion Q_p, which occurs after the soil water storage capacity is exceeded:

$$Q_\mathrm{p} = (P - E)(1 - A_\mathrm{i}) - W_\mathrm{m} + W, \quad \text{when } i_\mathrm{m} \le i_0 + P - E \tag{3}$$

If

$$i_\mathrm{m} \ge i_0 + P - E, \quad \text{then } Q_\mathrm{p} = (P - E)(1 - A_\mathrm{i}) - W_\mathrm{m} + W + W_\mathrm{m}\left(1 - \frac{i_0 + P - E}{i_\mathrm{m}}\right)^{1+b} \tag{4}$$

where W is the current soil water depth that contributes to evapo-transpiration; i_0 is current water depth in the unsaturated zone;, i_m is the maximum soil water depth; and W_m is the maximum storage depth over the basin expressed as

$$W_\mathrm{m} = \frac{i_\mathrm{m}}{1 + b}(1 - A_\mathrm{i}) \tag{5}$$

The soil water depth i is given by

$$i(A) = \begin{cases} i_\mathrm{m}\left\{1 - \left(\frac{A - A_\mathrm{i}}{1 - A_\mathrm{i}}\right)^{1/b}\right\} & \text{when } A_\mathrm{i} \le A \le 1.0 \\ 0 & \text{when } 0 \le A \le A_\mathrm{i} \end{cases} \tag{6}$$

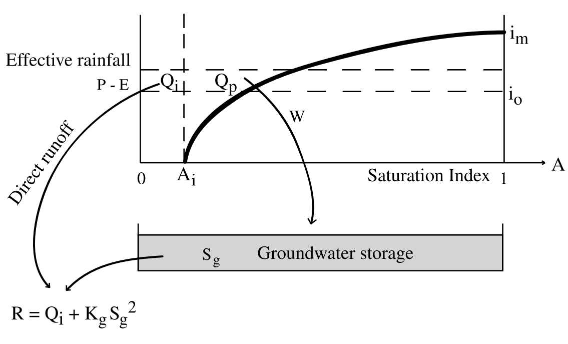

Fig. 2. Schematic of the runoff process model (RPM).

where A is the fraction (in the range 0–1) of a sub-basin that contributes to a particular segment of the river; and b is a model parameter that depends upon the shape of the soil water storage capacity distribution.

The FRM models flow routing along the river network. This is given by

$$Q_j(x,t) = Q_j(0,t) + q_0(t)x \tag{7}$$

where $Q_j(x,t)$ is the jth segment discharge at distance x from an upper end; $Q_j(0,t)$ is the inflow at the upper end of the drainage segment; and $q_0(t)$ is the discharge flux rate in space along a drainage segment, as given by the RPM. The RPM and the FRM are connected through a data-sending port (DSP) and a data-receiving port (DRP) (Fig. 3). The FRM computes discharges from the MEM outlets. To reduce the computational burden, it assumes that the discharge varies linearly along each river segment at each time step instead of computing that for each computational cross-section within the grid-cell (Shiiba et al., 1996).

The functions of the DSP and DRP are to exchange computed data (Fig. 4) within each grid-cell and between the MEMs. Inside the MEM, the DRP feeds grid-cell-mean input data (such as precipitation and evapo-transpiration averaged over each grid-cell) to the RPM, which produces runoff from the grid-cell. The DSP feeds that value into the FRM. The FRM receives information through its DRP from the RPM of the same grid and the adjacent MEMs of upstream grids in order to provide the discharge to the downstream MEMs. The accumulated values are transferred to the downstream MEMs only after calculations for all upstream MEMs are completed.

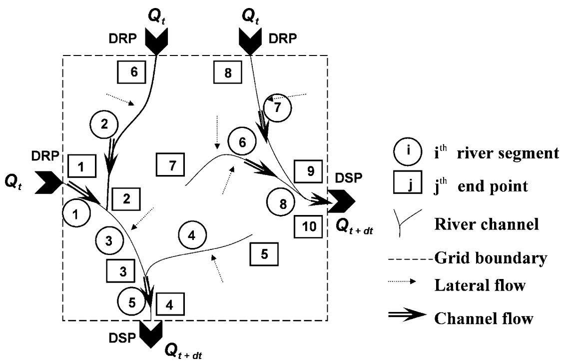

Fig. 3. Schematic of the flow routing model (FRM).

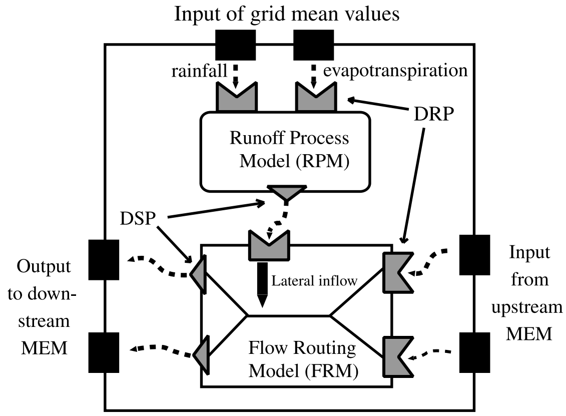

Fig. 4. Structure of the MaScOD element model (MEM).

The flow path of the river network is linked to provide a total runoff simulation model that is a combination of MEMs (Fig. 5).

The RPM in the MaScOD model considers the soil properties and interacts with the FRM in terms of lateral flow. The FRM considers topographic attributes such as channel slope using kinematic wave routing. Piecewise connectivity of the flow network and the calculation sequence from upstream to downstream help reduce bias.

The discharge simulation is conducted in Huaihe River Basin (270,000 km^2). The Huaihe River, one of the China's major rivers, runs for about 1000 km between the Yellow and Yangtze Rivers, through Henan, Anhui, Jiangsu, and Shandong provinces.

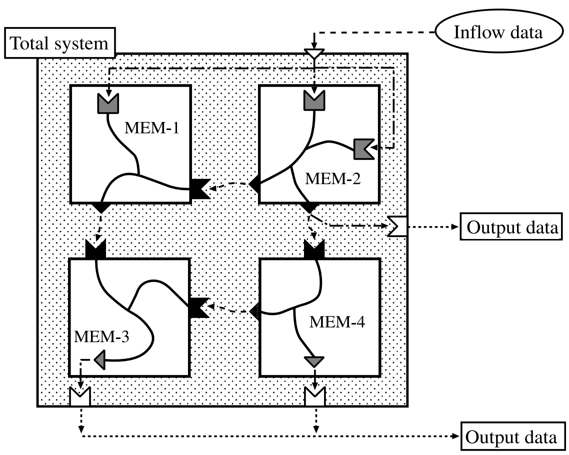

Fig. 5. Schematic of total system.

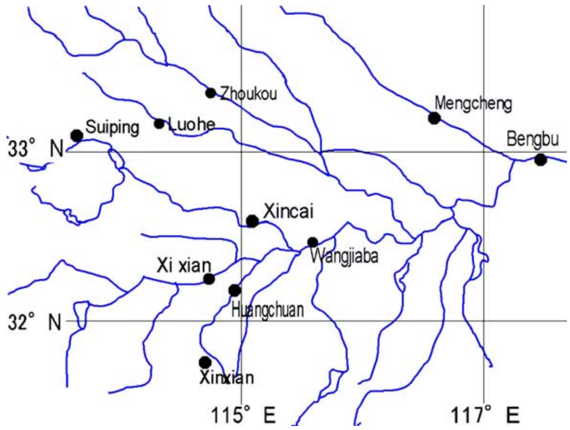

Fig. 6. Locations of discharge observation points.

This region is home to a population of about 165 million and frequently undergoes disastrous flooding. The MaScOD model parameters were calibrated on the basis of field observation records in the Shiguan River Basin (6000 km^2), a sub-basin of the Huaihe River Basin, which is assumed to be representative of the physiographical characteristics of the entire basin. The hydro-meteorological data obtained from 48 rainfall stations, five discharge stations and four pan evaporation stations inside the basin were utilized to calibrate the model. The model parameters calibrated for the Shiguan River basin were subsequently assigned to the entire study area, the basic assumption being that the similar regions should have similar parameter values. Separate parameter sets were calibrated for mountainous and flat regions. Topographic maps were used to identify similar regions before the calibrated parameters were assigned. The details of the calibration process are described by Tachikawa et al. (2001).

In this study, simulation results are compared with the observed discharges at Suiping, Wangjiaba and Bengbu (Fig. 6), the corresponding contributing areas being 2093; 29,844; and 132,350 km^2, respectively.

3. Experimental hydro-meteorological input data

3.1. Input data source

Grid precipitation and grid actual evapo-transpiration data, for the period of May 1st, 1998 to August 31st, 1998 are the forcing input data to the hydrological model. Experimental forcing data were created from two data sets, namely the HUBEX-IOP EEWB dataset and the GAME re-analysis dataset. The HUBEX-IOP EEWB data (abbreviated from the "Huaihe River Basin Experiment—Intensive observation period—Estimation of Energy and Water Budget" termed 'EEWB data' here-after) have 5-min spatial resolution and 1-h temporal resolution (Kozan et al., 2001). The precipitation field of the EEWB data is generated from ground-based observation using a time and space interpolation technique. The evapo-transpiration field is the output of a simple biosphere with urban canopy (SIBUC) model (Tanaka, 1998). In an earlier study, the EEWB data did not produce good results for discharge simulation in the Bengbu basin (Tachikawa et al., 2002), the largest basin of this study, while the results for the smaller sub-basins were good.

The GAME (GEWEX Asia Monsoon Experiment) Re-analysis data (Version 1.1) with 1.25-degree spatial and 6-h time resolutions (termed as "GAME data" here-after) is another data set used in this experiment (JMA 2000). These data were produced using a 4DDA system in a co-operative study of the Japan Meteorological Agency (JMA), and the Earth Observation Research Center (EORC), NASDA. Using the GAME data, the river discharges in the three study basins were successfully simulated by Shrestha et al. (2002), which yielded a better simulation result than the EEWB data used by Tachikawa et al. (2002) for the large basin but did not yield a better simulation result for the smaller two sub-basins.

3.2. Generation of experimental forcing data

Possible errors in the magnitude of the EEWB data, especially in the evapo-transpiration field over the grid-cell system, may be one of the reasons behind the simulation error for the large basin. The coarser resolution of the GAME data was suspected to be the dominant reason for simulation error in the two smaller basins (this is proved true by the results presented in this paper). Hence, an alternate data set is needed to investigate the effect of forcing data resolution, one that overcomes the weaknesses of both of the two available data sets. To acquire an experimental data set with a 10-min spatial resolution, the GAME data are converted according to the spatial pattern of the EEWB data such that the resulting experimental 10-min data set preserves the spatial pattern of the EEWB data and the magnitudes of the precipitation and evapo-transpiration of the GAME data. The reason behind adopting the spatial pattern of the EEWB data is that the EEWB data are basically ground-based data, which take into account both the distance and the direction of each ground observation station. Also, the EEWB data are the finest resolution data among the available distributed data for the study basin. Therefore, these data are assumed to represent the spatial pattern better than any other data. The reason behind adopting the magnitudes of the precipitation and evapo-transpiration of the GAME data is that they have produced a good fit of the observed basin-scale discharge, showing their better accuracy in describing the water balance (Shrestha et al., 2002). The temporal scale (6-h time resolution) of the GAME data is adopted for the experimental data.

Various methods may be adapted to transfer the spatial pattern from one data set to another, in this case from EEWB to GAME data. The spatial pattern of the precipitation and evapo-transpiration of the EEWB data is considered in terms of a fluctuation from its mean value over the corresponding spatial domain of a single grid cell of the GAME data, as shown in Eqs. (8) and (9):

$$P_{4j} = \begin{cases} P_{1j} + P_{2j} - P_{3j} & \text{if } (P_{1j} + P_{2j} - P_{3j}) \geq 0 \\ 0 & \text{otherwise} \end{cases}$$
$$\text{For time } j = 1, 2, 3, \ldots, m \quad (8)$$

$$P_{5j} = P_{4j} \frac{\sum P_{1j}}{\sum P_{4j}} \quad \text{For time } j = 1, 2, 3, \ldots, m \quad (9)$$

Here, the P_{4j} are intermediate newly created experimental 10-min data. The P_{1j} are the GAME 1.25-degree data; the P_{2j} are the EEWB 10 min data; and the P_{3j} are the average of EEWB 10-min data at 1.25-degree resolution, all of them have the same unit of intensity per unit area. Sometimes, the P_{4j} values may appear negative, but are forced by Eq. (8) to become zero. The accumulated value of the experimental data inside the catchment frequently appears to be slightly different from the accumulated value of the original GAME1.25 data due to forcing the negative P_{4j} values to become zero. Hence, in Eq. (9), a scaling factor of the ratio between the new accumulated value and the accumulated value of the original GAME1.25 data is applied to all the P_{4j} data value to ensure that the total accumulated input of P_{5j} values is the same as that of the original GAME1.25 data. Thus the P_{5j} are the final experimental input data (Fig. 7). Fig. 7 also shows a typical spatial pattern of rainfall intensity for the 1.25-degree GAME reanalysis data. It can be seen that the rainfall intensities for the experimental 10-min dataset are higher and more

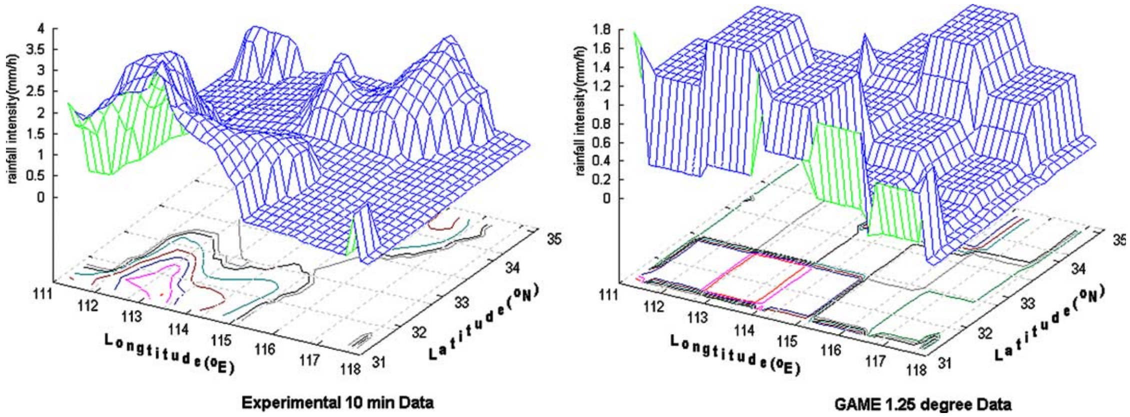

Fig. 7. Typical example of spatial patterns of experimental data (P_{5j}) and GAME1.25 data (P_{1j}).

Table 1
Comparison of performance indices using different forcing data

	Pearson MC coeff.			Nash Sutcliffe coeff.			Index of agreement		
	E	G	Exp	E	G	Exp	E	G	Exp
Suiping	0.727	0.456	0.722	0.516	0.190	0.493	0.828	0.582	0.835
Wangjiaba	0.911	0.761	0.881	0.635	0.465	0.693	0.893	0.802	0.931
Bengbu	0.677	0.729	0.759	−0.031	0.433	0.441	0.692	0.851	0.866
Average	0.772	0.648	0.787	0.373	0.363	0.543	0.804	0.745	0.877

Note: E: EEWB 10-min data; G: GAME 1.25-deg data; Exp: experimental data.

localized. The hydrologic response of a small catchment to these two rainfall fields will be different, as demonstrated later in this paper.

The experimental 10-min data set is used further to create 20-min, and 30-min data sets, etc., by passing a spatially averaging window of (2×2), (3×3), etc., up to 150-min (2.5-degrees), respectively. The total accumulated value of the input data is kept constant.

4. Discharge simulation results

The experimental data and the EEWB data, both of 10-min spatial resolution, and the GAME 1.25-degree were fed into the MaScOD model to simulate discharge in all three-study basins: Suiping (2093 km^2), Wangjiaba (29,844 km^2) and Bengbu (132,350 km^2). Four performance indices (a) the Pearson moment correlation coefficient (PMC), (b) the Nash-Sutcliffe coefficient of efficiency (NSI) (Nash and Sutcliffe, 1970), (c) the Index of agreement (IOA) (Willmott, 1981) and (d) root mean square error (RMSE) were used to evaluate the model performances by comparing the simulated and observed discharges. Eqs (10)–(13) describe the evaluation criteria

$$\mathrm{PMC} = \frac{\sum_{i=1}^{N}(O_\mathrm{i} - \bar{O})(P_\mathrm{i}\bar{P})}{\left[\sum_{i=1}^{N}(O_\mathrm{i} - \bar{O})^2\right]^{0.5}\left[\sum_{i=1}^{N}(P_\mathrm{i} - \bar{P})^2\right]^{0.5}} \tag{10}$$

$$\mathrm{NSI} = 1.0 - \frac{\sum_{i=1}^{N}(O_\mathrm{i} - P_\mathrm{i})^2}{\sum_{i=1}^{N}(O_\mathrm{i} - \bar{O})^2} \tag{11}$$

$$\mathrm{IOA} = 1.0 - \frac{\sum_{i=1}^{N}(O_\mathrm{i} - P_\mathrm{i})^2}{\sum_{i=1}^{N}(|P_\mathrm{i} - \bar{O}| + |O_\mathrm{i} - \bar{O}|)^2} \tag{12}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(O_\mathrm{i} - P_\mathrm{i})^2}{N}} \tag{13}$$

Here, O_i represents observed value at the ith time; P_i represents simulated value at the ith time; N is the total number of observations/simulations; $\bar{O}$ and $\bar{P}$ are the mean values of O_i and P_i, respectively. The best condition is that Eqs. (10)–(12) yield a value of unity, and that the RMSE is zero.

In terms of the IOA values shown in Table 1, the simulation results from the experimental (P_{5j}) data are found to be better than the corresponding results from both the original EEWB and GAME data sets for all three basins (Tachikawa et al., 2002; Shrestha et al., 2002). The NSI and PMC results were also quite favorable. The experimental data are then used as the base-line data and to prepare the input data of various resolutions.

The hydrograph of simulated discharge changes as the input data resolution changes (Fig. 8). Coarser resolution input data, which is just a set of spatially averaged values, yielded different simulation results

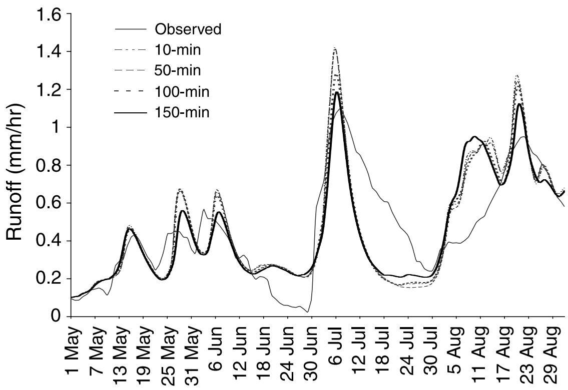

Fig. 8. Simulated hydrographs at Bengbu (132,350 km^2) obtained from various input data resolutions.

than finer resolution input data in all three basins. This is solely due to the effect of spatial variability.

Overlapping simulation results obtained from various input data resolutions, yields a band of simulated hydrographs. Wide bands of simulated hydrographs (Fig. 9) show that there is high sensitivity to input data resolution in all three study catchments. The hydrograph obtained from the finest resolution input data is closest to the observed hydrograph for the smallest catchment (Fig. 9(a)); however, for the larger catchments (Fig. 9(b) and (c)), the hydrographs obtained from the finest resolution input data are not the best ones.

Fig. 10 shows accumulated discharge for the three study catchments. The bandwidth of accumulated discharge diverges more in the case of the smaller basin than for the larger basins, which might be due to a lower capacity of the smaller basin to store soil-water within the catchment compared to that of the larger basins. Despite the presence of significant

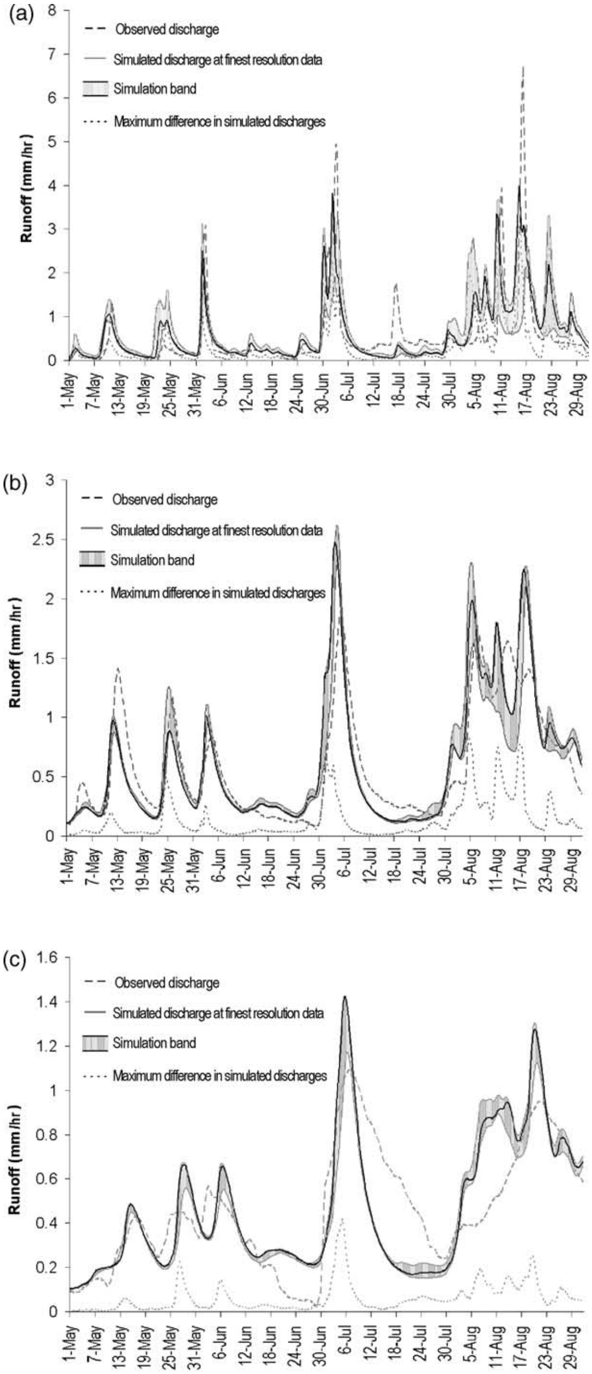

Fig. 9. The simulation band for hydrographs at (a) Suiping (2093 km^2), (b) Wangjiaba (29,844 km^2), and (c) Bengbu (132,350 km^2) produced by overlapping the hydrographs obtained from different resolutions of input data yields. It can be seen that the finest resolution data is not necessarily the best.

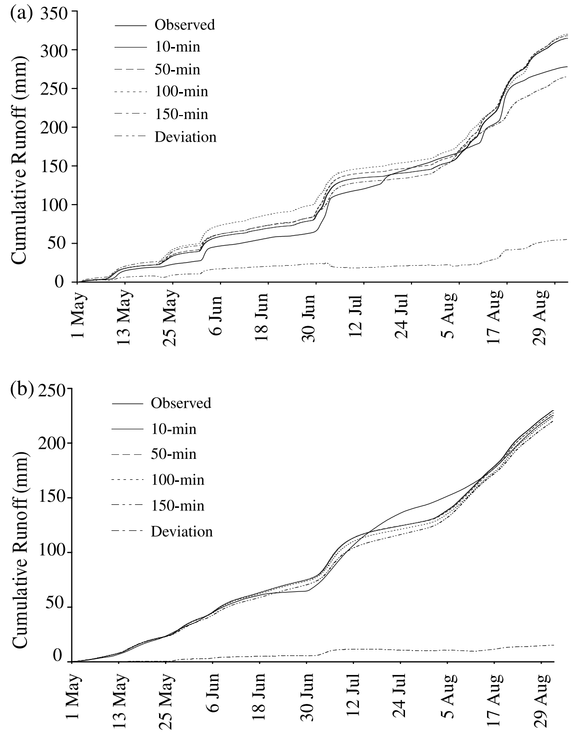

Fig. 10. Cumulative runoff for various input data resolutions (a) at Suiping (2093 km^2 (b) at Bengbu (132,350 km^2); the smaller catchment has the larger difference in accumulated discharge.

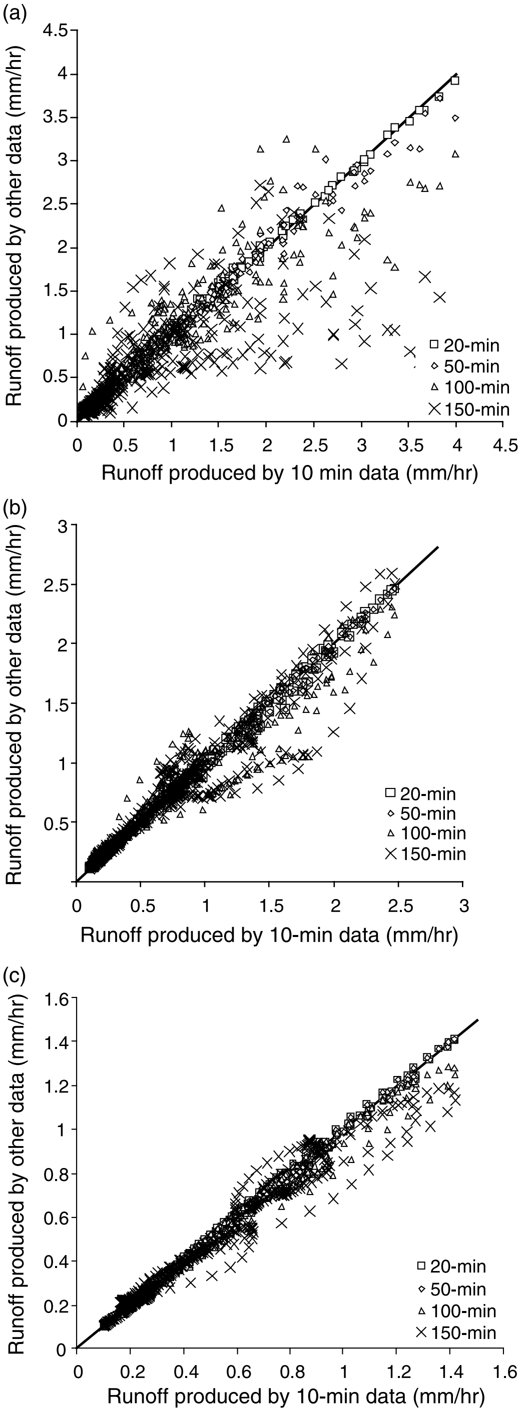

Fig. 11. Deviation of simulated discharge due to change in resolution of input data (a) at Suiping (2,093 km^2), (b) at Wangjiaba (29,844 km^2), (c) at Bengbu (132,350 km^2).

mismatches between the simulated and observed hydrograph (Fig. 8), the accumulated values of the simulated and observed discharge at Bengbu (Fig. 10(b)) match well. This indicates that the model has preserved the basin outflow volumes. The mismatch in hydrographs and good match in accumulated discharge could have occurred due to changes in the runoff processes induced by human activities, e.g. irrigating large paddy fields or flow regulation at the hydraulic control structures and reservoirs that exist within the study basins.

The effects of forcing input data resolution on the distributed hydrological modeling are shown in Fig. 11. In which, the discharges simulated by various resolution input data are compared with the discharge simulation by 10-min resolution. Using coarse resolution input data can give very different runoff to that obtained using fine resolution input data, especially for the smallest catchment. The hydrograph peaks are particularly affected. This shows that resolution issues need more careful consideration in high flow simulation than in low flow simulation.

5. Model performance in terms of the IC-ratio

The components of flow processes, being functions of the detailed geometry of flow pathways in different catchments, are difficult to compare (Beven, 2002). Some indices, for example average slope, flow lengths, watershed relief, etc. represent various catchment features but they may not be suitable for testing against forcing data resolutions and scale issues over different catchments since these values are highly variable from one catchment to another. A wide range of responses may be obtained when using these indices to investigate scale issues. Size of catchment is probably the only absolute statistic providing consistent information and a sound basis for investigating suitable resolution for hydrologic modeling. The ratio of the input forcing resolution to catchment size, called the IC-ratio (Shrestha et al., 2002), may therefore prove to be a useful index for investigating the effects of input data resolution on discharge.

The finest spatial resolution adopted in this experiment is 10-min (approximately 13.5 km in each direction at 33°N, or 175 km^2). The input data for the Suiping basin (2093 km^2) varies from 10-min to 2.5-degree resolution. Therefore, the IC-ratio value ranges from about 1:12 to 1:0.05. Model performance is found to be consistently better at 1:12 than near 1:0.05. The convention of the ratio is to keep the numerator equal to one, such that a higher denominator value (called a higher IC-ratio) corresponds to finer resolution and a lower denominator value (called a lower IC-ratio) corresponds to a coarser resolution of input for a given catchment. The IC-ratio values for the Wangjiaba basin (29,844 km^2) and the Bengbu basin (132,350 km^2) vary from 1:168 to 1:0.75 and 1:745 to 1:3.3, respectively, within the framework of this experiment.

Concise information on the changes in the model performance in response to the altered resolution of forcing data can be represented, as in Fig. 12, using the IC-ratio index. This facilitates the comparison of the responses at various scales of catchment size. In this figure, the Pearson moment correlation coefficient (the PMC of Eq. (3)) between the observed and simulated hydrographs is used as the performance index. The figure shows that the smallest basin (Suiping) has higher sensitivity in response to the forcing data resolution, as the steepness of the performance versus IC-ratio curve is highest for that case. A similar sensitivity is displayed in Fig. 11. In the small basin, the simulation results deteriorate faster as the input resolution becomes coarser. Model performance improvement with finer resolution input

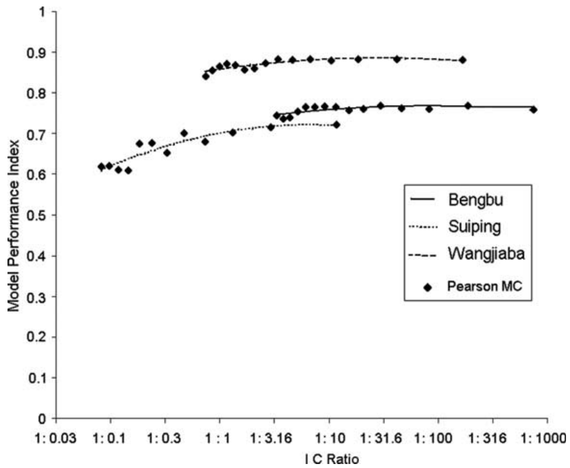

Fig. 12. Model performances versus IC-ratio for the three catchments.

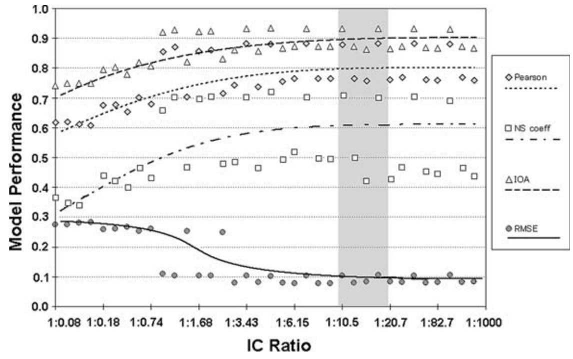

Fig. 13. Model performance versus IC-ratio for four indices.

data has a tendency to level off. At the higher IC-ratio values, the improvement rates of the model performances are not very significant.

The four model performance indices are evaluated and plotted against the IC-ratio in Fig. 13. Since the IC-ratio is a dimensionless number, the performance indices in Fig. 13 are not separated for different basins but are plotted as an overall trend line for each index. The Pearson moment correlation coefficient, the Nash Sutcliffe coefficients and the index of agreement all follow similar trend lines, which indicate a faster rate of performance improvement in the lower IC-ratio range. As the IC-ratio increases, the rate of improvement in performance gradually reduces and the value of model performance indicator levels off to a constant value. In contrast, the trend of the RMSE followed a converging path toward constant values at both ends of the IC-ratio range. The results obtained in this experiment are satisfactory while the IC-ratio remains above 1:10. All the performance indices are found to level off above the IC-ratio 1:20 and the performance improvement beyond that are likewise un-attractive. Poorer performance index values below the IC-ratio 1:10 suggest that if there are less than 10 input grid cells over the basin, the distributed model does not yield good results. This is more critical in the case of smaller basins.

6. Selection of input grid resolution

Although most modelers are well aware that an appropriate resolution of input data may differ from an appropriate resolution of model disaggregation, it is generally preferred, for convenience in the simulation exercise, to keep the input data grid resolution the same as that of the grid resolution of the model. In numerical modeling, a finer-resolution model can accept a coarser-resolution input data by splitting up the input data in proportion to the grid size of the model. However, a coarser-resolution model cannot be easily devised to take a finer-resolution input data without losing the data properties. This limitation becomes a dictating factor in selecting the modeling framework while the needed input resolution is finer than the resolution opted for the distributed hydrological modeling.

A finer-resolution input data is preferred for its better description of spatial variability. It may be impractical effort to include every details of input field in catchment-scale modeling, especially in the case of large and scarce-information catchments. This leads to the option of looking for a homogenous area, which may still preserve the representative characteristics of the larger heterogeneous input data field. However, this introduces a confusion relating to the scaling problem because it is not easy to predict the largest input resolution that is capable of preserving the average hydrologic response characteristics of the large regional heterogeneity of the input data field.

The results obtained in this study indicate that, while the response to input resolution change is not same in all basins, they are however quite similar that the size of the largest homogenous region of input data field depends upon the catchments size. Looking toward a higher degree of input disaggregation, there is initially a tendency to yield better results always, but too much disaggregation fails to improve the performance much. With respect to a further increase in resolution, little improvement can be achieved once it crosses the resolution corresponding to the IC-ratio 1:20. For example, the model performance is almost the same at ratios 1:20 and 1:200, but the later case needs to bear the cost of having to acquire high-resolution data. The slight improvement in results obtained by using higher resolution data is usually such as to discourage choosing a high-resolution data.

The characteristic relationship between model performance and scale is represented qualitatively by the performance curve in Fig. 14. The relationship between the cost and scale in Fig. 14 is difficult to determine mainly because the conditions of one basin

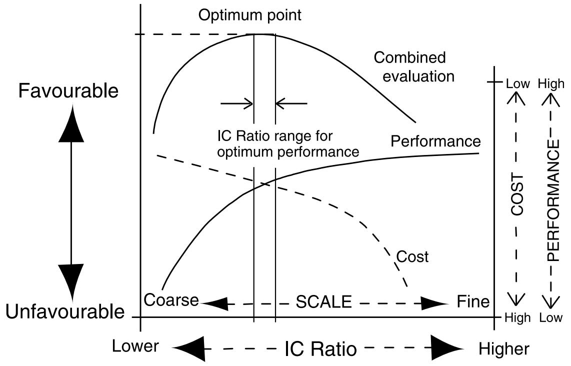

Fig. 14. Trade-off between model performance and model cost with respect to scale in terms of the IC-ratio.

differ from those of another. However, the cost is likely to increase geometrically upon for an arithmetic increase in resolution. The cost curve shown in Fig. 14 indicates the higher cost associated with finer resolution scales (or higher IC-ratio). Even though a quantitative cost analysis is not conducted in this study, the optimum performance range can be appreciated simply by observing the performance versus IC-ratio curve in Fig. 13, bearing in mind the cost considerations shown in Fig. 14

In an un-gauged basin, where a hydrologist needs to start work from a very low base-line, a lot of unknowns are expected to be encountered in the process of setting up the hydrological model. The needs for spatially distributed data with sufficient accuracy and resolution scale, as demanded by a distributed hydrologic model, hinders the modeler right from the start of the exercise. Being able to select an appropriate input data resolution, as a function of the catchment area in terms of IC-ratio, may help designing the modeling work. A useful range of IC-ratio values that provides a satisfactory simulation result is noticed to be within 1:10 to 1:20. Larger catchments may be modeled successfully at coarser scale by representing its larger heterogeneous input data field as homogenous units but smaller catchments need finer scales, which the IC-ratio tends to quantify. Model sensitivity to the discretization scales of input data exists up to an upper limit of the IC-ratio with constant parameter values for the selected hydrological/routing model structure. This distinction is likely to be very useful in setting up a model framework for an ungauged basin.

The model resolution was kept constant at 10-min for this study, in order to consider the effects of input data resolution in isolation from other modeling issues. While a resolution of 10-min is required to model the smallest catchment considered here, the effort required to set up a distributed model at this resolution may not be justified for larger catchments, particularly if an input data resolution coarser than 10-min is selected based on the proposed IC-ratio rule. If this is the case, and provided that the river network and hydrological processes can be adequately modeled at the coarser resolution, the hydrological model's spatial resolution could be chosen to match the resolution of the hydro-meteorological input data.

7. Conclusions

The performance of a macro-scale distributed hydrological model depends upon the quality of the model, the selected model parameters, and the quality of the input data. The quality of the input data set is associated with accuracy and resolution issues. The ratio of input data resolution to the catchment area, called the IC-ratio (Shrestha et al., 2002), can serve as a useful index to evaluate the suitability of data and choose a model resolution for distributed hydrological modeling.

In this study, three catchments, ranging from small (2093 km^2) to large (132,350 km^2) size, were used to investigate the effects of data resolution on the discharge simulation results. The two smaller catchments were found to be more sensitive to data resolution than the larger one, indicating that resolution issues need more attention for smaller basins, and that model performance depends upon the input data resolution. However, demanding very high-resolution data is not a sensible or practical solution in response to this. The minimum amount of input data, the minimum number of parameters, and the minimum computational load which produce reasonable simulation results, make the distributed hydrological model easier to apply and more cost-effective. We find that an IC-ratio of 1:10–1:20 gives overall optimum performance of a macro-scale distributed hydrological model, bearing in mind the effort required to process high resolution data for an ungauged basin. An IC ratio in this range gives a sensible balance of accuracy, cost, time, and complexity.

Our recommendation of an IC ratio in the range of 1:10–1:20 implies that the distributed modeler should obtain hydro-meteorological input data at a scale such that about 10–20 pixels are sufficient to cover the catchment. We certainly would not recommend an IC ratio of less than 1:6, i.e. using less than six input data pixels to model a catchment. This has implications for the minimum size of ungauged catchment that can be successfully modeled by a distributed model. The finest spatial resolution of hydro-meteorological input data that was obtained for this study was 10-min data, giving a pixel covering approximately 175 km^2 at 33°N. This required a substantial amount of processing and cross-checking of the quality of the data, based on both the 1.25° GAME-IOP reanalysis dataset that covers the Asian Monsoon region (JMA 2000), and the poorer-quality 5-min HUBEX-IOP EEWB dataset that covers the Huaihe River Basin in China (Kozan et al., 2001). Thus good quality regional hydro-meteorological datasets at resolutions of 10-min or finer are difficult to obtain. This, along with our recommended range for the IC-ratio, implies that the use of hydro-meteorological datasets to model ungauged catchments may be difficult for catchments much smaller than 1700 km^2. The ease with which appropriate data can be obtained improves as the catchment size increases. However, future improvements in resolution and accuracy of the outputs of the atmospheric models used for regional data reanalysis will improve this situation.

In our effort to establish a preliminary data resolution criterion (i.e. the proposed IC-ratio Rule) for modeling an ungauged basin, which would indicate at least an approximation of the data requirements, and to highlight the necessity of considering catchment area for setting an objective grid resolution of re-analyzed data at the regional scale, we focused in this study on the trade-off between data density and the efforts needed in hydrological modeling. We appreciate that our case for the IC-ratio Rule could be strengthened and refined by considering more sub-basins of the Huaihe River basin and also that the applicability of the rule could usefully be tested using a gauged sub-basin as a control. It is our intention to address these considerations in a later phase of this research.

Acknowledgements

The contribution of the HUBEX-IOP EEWB data by Dr K. Tanaka, DPRI, Kyoto University and of the GAME Re-analysis data by the Japan Meteorological Agency is sincerely acknowledged. We are grateful to Prof. M. Shiiba and to Dr Y. Ichikawa for providing us with the OHyMoS library source codes and to Dr Timothy Harrold for providing helpful comments. We also acknowledge two anonymous reviewers, whose suggestions have improved the paper.

References

Bathurst, J.C., 1986. Physically based distributed modeling of upland catchment using the Systeme Hydrologique Europeen. Journal of Hydrology 87, 79–123.

Beven, K., 2002. Towards an alternative blueprint for a physically based digitally simulated hydrologic response modeling system. Hydrological Processes 16, 189–206.

Blöschl, G., Sivapalan, M., 1995. Scale issues in hydrological modeling: a review. Hydrological Processes 9, 251–290.

Bruneau, P., Gascuel-Odoux, C., Robin, P., Merot, P., Beven, K., 1995. Sensitivity to space and time resolution of a hydrological model using digital elevation data. Hydrological Processes 9, 69–81.

Burlando, P., Rosso, R., 2002. Effects of transient climate change on basin hydrology. Hydrological Processes 16, 1151–1175.

Fan, Y., Bras, R.L., 1995. On the concept of a representative elementary area in catchment runoff. Hydrological Processes 9, 821–832.

Flügel, W.A., 1995. Delineating hydrological response units by geographical information system analyses for regional hydrological modeling using PRMS/MMS in the drainage basin of the river BRÖL, Germany. In: Kalma, J.D., Sivapalan, M. (Eds.), Scale Issues in Hydrological Modelling. Wiley, England, pp. 181–194.

Franchini, M., Wendling, J., Obled, Ch., Todini, E., 1996. Physical interpretation and sensitivity analysis of the TOPMODEL. Journal of Hydrology 175, 293–338.

Gupta, V.K., Mesa, O.J., Dawdy, D.R., 1994. Multiscaling theory of flood peaks: regional quantile analysis. Water Resource Research 30, 3405–3421.

Horritt, M.S., Bates, P.D., 2001. Effects of spatial resolution on a raster based model of flood flow. Journal of Hydrology 253, 239–249.

Ichikawa, Y., Shiiba, M., Tachikawa, Y., Takara, K., 2000. Object oriented hydrological modeling system. The Fourth International Conference on HydroInformatics CD-ROM. Cedar Rapids, Iowa.

Japan Meteorological Agency (JMA), 2000. GAME-IOP Reanalysis Data (Version 1.1), downloading site ftp://hydro.iis.u-tokyo.ac.jp/GAME/GAIN/GAME-REANL/

Kavvas, M.L., Chen, Z.Q., Tan, L., Soong, S.-T., Terakawa, A., Yoshitani, J., Fukami, K., 1998. A regional-scale land surface parameterization based on areally-averaged hydrological conservation equations. Hydrological Sciences Journal 43 (4), 611–631.

Koren, V.I., Finnerty, B.D., Schaake, J.C., Smith, M.B., Seo, D.-J., Duan, Q.-Y., 1999. Scale dependencies of hydrologic models to spatial variability of precipitation. Journal of Hydrology 217, 285–302.

Kozan, O., Tanaka, K., Ikebuchi, S., 2001. Estimation of energy and water budget in the Huaihe River Basin, China. GAME Publication 28, 32–35.

Molnar, D.K., Julien, P.Y., 2000. Grid-size effects on surface runoff modeling. Journal of Hydrologic Engineering, ASCE January, 8–16.

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models. Part 1. A discussion of principles. Journal of Hydrology 10, 282–290.

Quinn, P.F., Beven, K.J., 1991. The prediction of hillslope flow paths for distributed hydrological modeling using digital terrain models. Hydrological Processes 5, 59–79.

Refsgaard, J.C., Abott, M.B., 1996. The role of distributed hydrological modeling in water resources management. In: Refsgaard, J.C., Abott, M.B. (Eds.), Distributed Hydrologic Modeling. Kluwer, Dordrecht, The Netherlands.

Renssen, H., Isarin, R.F.B., Jacob, D., Podzun, R., Vandenberghe, J., 2001. Simulation of the Younger Dryas climate in Europe using a regional climate model nested in an AGCM: preliminary results. Global and Planetary Change 30, 41–57.

Saulnier, G.M., Obled, C., Beven, K., 1997. Analytical compensation between DTM grid resolution and effective values of saturated hydraulic conductivity within the TOPMODEL framework. Hydrological Processes 11 (9), 1331–1346.

Shiiba, M., Ichikawa, Y., Ikebuchi, S., Tachikawa, Y., Takasao, T., 1996. Scale-up of a runoff model using GIS and an object-oriented hydrological modeling system. In: Kovar, K., Nachtnebel, H.P. (Eds.), Proceedings of HydroGIS 96: Application of Geographic Information System in Hydrology and Water Resources Management, vol. 235. IAHS, pp. 63–70.

Shrestha, R., Tachikawa, Y., Takara, K., 2002. Effect of forcing data resolution in river discharge simulation. Annual Journal of Hydraulic Engineering, JSCE 46, 139–144.

Singh, V.P., Woolhiser, D.A., 2002. Mathematical modeling of watershed hydrology. Journal of Hydrologic Engineering, ASCE 7 (4), 270–292.

Sivapalan, M., Kalma, J.D., 1995. Scale problems in hydrology: the contributions of the Robertson workshop. In: Kalma, J.D., Sivapalan, M. (Eds.), Scale Issues in Hydrological Modeling. Wiley, England, pp. 1–8.

Sunada, K., Aoki, K., Fujimura, T., 2001. Effects of scale of catchment element on runoff response of the distributed model (in Japanese). Annual Journal of Hydraulic Engineering, JSCE 45, 145–150.

Tachikawa, Y., Kawakami, T., Ichikawa, Y., Shiiba, M., Takara, K., 2001. Application of a macro grid based hydrological model to the Huaihe River Basin in China. Annual Journal of Hydraulic Engineering, JSCE 45, 127–132.

Tachikawa, Y., Takara, K., Tanaka, K., Suishu, T., Ichikawa, Y., Shiiba, M., 2002. Simulation of River Discharge in the Huaihe River Basin in China. Journal of Japan Society on Hydrology and Water Resource 15 (2), 139–151.

Takasao, T., Shiiba, M., Ichikawa, Y., 1996. A runoff simulation with structural hydrological modeling system. Journal of Hydroscience and Hydraulic Engineering, JSCE 14 (2), 47–55.

Tanaka, K., 1998. Activities in Huaihe River Basin (HUBEX) study area. Asia-Pacific Network for Global Change Research, Report of 1997–1998, 69–91.

Willmott, C.J., 1981. On the validation of models. Physical Geography 2, 184–194.

Wood, E.F., Sivapalan, M., Beven, K., Band, L., 1988. Effects of spatial variability and scale with implications to hydrologic modeling. Journal of Hydrology 102, 29–47.

Woods, R.A., Sivapalan, M., Duncan, M., 1995. Investigating the representative elementary area concept: an approach based on field data. Hydrological Process 6 (3/4), 291–312. ED 830.

Zhang, W., Montgomery, D.R., 1994. Digital elevation model grid size, landscape representation, and hydrologic simulations. Water Resources Research 30 (4), 1019–1028.

Zhao, R.J., 1992. The Xinanjiang model applied in China. Journal of Hydrology 135, 371–381.