Predictions in Ungauged Basins: PUB Kick-off (Proceedings of the PUB Kick-off meeting held in Brasilia, 20–22 November 2002). IAHS Publ. 309, 2007.

Selection of scale for distributed hydrological modelling in ungauged basins

ROSHAN SHRESTHA, YASUTO TACHIKAWA & KAORU TAKARA
Disaster Prevention Research Institute, Kyoto University, Gokasho, Uji, Kyoto 611-0011, Japan
roshan@flood.dpri.kyoto-u.ac.jp

Abstract A macro scale grid based distributed hydrological model is developed using an object oriented hydrological modelling system (OHyMoS) and applied to the Huaihe River Basin (132 350 km^2) and its sub-basins, Wangjiaba (29 844 km^2) and Suiping (2093 km^2), with different resolution data (10 min to 2.5 degree) obtained from the GAME re-analysis data and Hubex-IOP EEWB data. The model performances are evaluated by comparing simulated discharge against observed discharge. Using the concept of IC Ratio (a ratio between the input grid area and the catchment area), the performance behaviour is observed. The study has found that the distributed model performs well with finer resolution data, and good performance is obtained with an IC Ratio 1:10 or higher, which sets a preliminary criterion for selecting the scale for distributed hydrological modelling in ungauged basins.

Key words distributed hydrological model; GAME re-analysis product; IC ratio; input data resolution

INTRODUCTION

It is broadly accepted that hydrological behaviours are predictable through some means of modelling. The development of hydrological modelling techniques is still progressing in a bid to achieve better accuracy and understanding. The basic principle elements of hydrological modelling procedures generally include data collection, model development, model calibration/validation, and model application. Experience has shown that data collection, one of the most essential and initial elements, is the most time consuming, expensive and heterogeneous part of the job. Further, large parts of the world do not have systematic installation and measurement of important hydrological information over the time period needed, such that they would enable computation of hydrological variables to the accuracy acceptable for practical applications.

On the other hand, the necessity of hydrological modelling is rapidly growing all over the world, since it provides information of prime importance for strategic planning of water use, disaster prevention, sustainable water management and many other applications. In addition, hydrological modelling often becomes a part of multidisciplinary studies. It demands an ability to analyse series of spatially variable hydrological behaviours and impacts of natural and human activities such as massive land-use changes, global warming effects, heavy rainfalls and droughts due to climatic fluctuations (Refsgaard & Abbott, 1996). In an effort to meet such demands, one must adopt the distributed approach to hydrological modelling, which again needs huge volumes of data. Thus hydrologists are having trouble meeting the demand for accurate and efficient modelling because of the limited availability of data.

A good strategy to deal with the problem could be using the best available data in the modelling, leaving the option open to improve it further, and immediate ratification of results until the availability of better data. For most basins, at least some basic hydrological information can be obtained through various sources, for example, General Circulation Models (GCM), remote sensing observation, global data archives and digital elevation models, etc., which can be applied plausibly. The absence of ground-based hydrological instrumentation in a basin no longer means that it is an absolutely ungauged basin. All basins have some data; however, their applicability may be questioned in relation to accuracy and proper scale issues. Usually GCM outputs are considered too coarse and not adequate for hydrological modelling (Burlando, 2002) to describe the variability in the components of hydrological processes at the basin scale. Digital elevation models are sometimes found to be erroneous, and remote sensing data may occasionally contain illusions. The basins having several types of hydrological information may still be referred to as ungauged basins according to the IAHS PUB definition, as the accuracy of available data may not be acceptable for practical application.

The modelling procedure for an ungauged basin begins with uncertainty and confusion. Moreover, to practice distributed hydrological modelling in an ungauged basin, there is an additional challenge to determine the scale at which spatial variability can be neglected, with the average characteristics of a given area providing sufficient information for accurate modelling of basin runoff generation (Molnar, 2000). The ability to choose an adequate scale at the preliminary investigation stage will definitely reduce the amount of painful exercises in the modelling task, and may produce results of higher accuracy.

This paper will illustrate the sensitivity of a distributed hydrological model's performance due to changing the scale of GCM-based input data. Applying a macroscale distributed hydrological model, which uses the object oriented hydrological modelling system (OHyMoS) (Takasao *et al*., 1996), the study was conducted for three basins in China, namely the Huaihe River (132 350 km^2), Wangjiaba (29 844 km^2) and Suiping (2093 km^2) basins.

BACKGROUND

Exciting improvements in the resolution and accuracy of GCM products have reduced the gap between the needs and availability of input data for distributed hydrological models. The available grid resolution of GCM data should resemble the needs of the hydrological model's spatial resolution. Shrestha *et al*. (2002a) demonstrated that the results of distributed hydrological models are better when the IC Ratio (the ratio between input data resolution and the catchment's area) is more than 1:10, suggesting that a marginal IC Ratio could serve as a criterion to finding a suitable resolution.

Macroscale hydrological modelling is often practiced for a large river basin. An object-oriented hydrological modelling system, OHyMoS (Takasao *et al*., 1996; Ichikawa *et al*., 2000) has eased many tasks associated with the modelling procedure for water movement in a grid-based analysis. For example, basin partitioning, hydrological process modelling for a sub-basin, linking sub-basin models together to make a total runoff model, processing channel network linkages to incorporate river flow routing, etc., involve a lot of substantial tasks, which have now been incorporated into an automatic procedure (Tachikawa *et al*., 2000). In this system, a basin is sub-divided into grid boxes, according to the grid system of a meso-scale atmospheric model outputs.

The model consists of a runoff process model, based on the Xinanjiang model (Zhao, 1992), and a lumped stream kinematic-wave flow routing model, combined for each grid cell to constitute an element model. These element models are interconnected by data-sending ports and data-receiving ports defined as per the river network arrangement to construct the total simulation process. The model parameters are calibrated on the basis of field observation records for the Shigan River basin, a sub-basin of the Huiahe River basin, which is assumed to include the representative geomorphology of the entire basin.

EXPERIMENTAL FORCING DATA

Using the grid precipitation and grid evapotranspiration data obtained from the HUBEX-IOP EEWB data (termed EEWB data here after), and 5-min spatial resolution data created at DPRI, Kyoto University (Tanaka *et al*., 1998), discharges are successfully simulated (Tachikawa *et al*., 2000a). Using 2-dimensional forecast data of the GAME Re-analysis (Version 1.1) with 1.25 degree resolution (termed GAME data hereafter), produced by the Japan Meteorological Agency using the 4 DDA-system including the GAME special observation off-line radiosonde data set, the discharges were successfully simulated again and it was observed that the simulation result with GAME data is better than EEWB data for large basins and the opposite is true for smaller basins (Shrestha *et al*., 2002a). Errors in the EEWB data are argued as the main reasons behind poorer simulation of the large basin. On the other hand, the coarser resolution of the GAME data is argued to be the reason behind poorer simulation of the smaller basin.

Merging the spatial pattern of the EEWB data at 10-min spatial resolution and mapping the magnitudes of GAME data into the same resolution, a test data set was created for experimentation. The EEWB data set considers both the distance and direction of each ground observation station to produce distributed data for precipitation fields. It uses the Simple Biosphere with Urban Canopy (SIBUC) model and the data collected during an intense observation period to obtain the evapotranspiration data field. Because the EEWB data is the product of the HUBEX IOP, it offers the most accurate patterns. Also, EEWB data is the finest resolution data among the distributed data available for the experimental basin until now. Therefore this data is chosen to describe the spatial pattern in the test data. On the other hand, GAME data is coarse resolution data produced by a numerical model. This data set has produced a better match of the simulated discharges for larger basins, which would hardly occur if the larger scale averages of the data field had contained a larger bias.

An arithmetic approach is used in the process of generating the experimental data of 10-min spatial resolution. This approach simply adds or reduces the differences in each pixel from the GAME 1.25-degree data. The differences are calculated for each pixel by comparing the EEWB 10-min data and its average over the domain covered up by GAME 1.25 degree data. Thus the new data set is given by:

$$P_4 = P_1 + P_2 - P_3 \quad (1)$$

where P_4 is the new 10-min resolution data, P_1 is GAME 1.25-degree, P_2 is EEWB 10-min, and P_3 is the average of EEWB data over the coarse resolution.

The pixel values of P_4 may sometimes be negative, which requires forcing into zero. This may cause the accumulated value of the new data to appear slightly higher than the original GAME data, which requires normalization to assure no-gains. We use a multiplication factor for this purpose, which is given by taking a ratio between the accumulated values of new and original data. After multiplying the new data and the factor, it maintains the same total input as that of the original data. The new experimental 10-min data is then used further to create coarser resolution data such as 20-min, 30-min and so on by passing an averaging window of (2 × 2), (3 × 3) and so on up to 150-min (2.5-degrees).

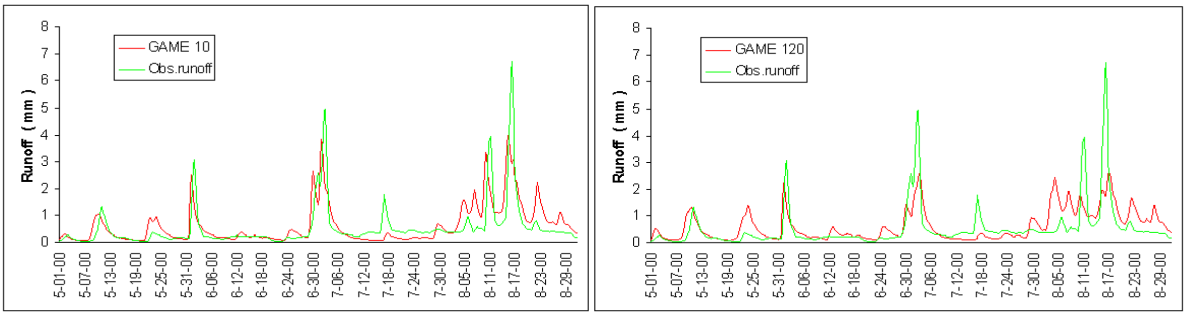
RESULTS AND DISCUSSION

Comparing the simulated discharge against the observed discharge, it is found that the experimental data produces better results than any of the original data, either from GAME or EEWB, for all three study basins. We use this case as the benchmark and proceed for checking the performances of the model at different scales of catchment and at different input resolution. We are, however, limited to conduct more comparisons at different scales of catchments because of the lack of gauged station discharge at additional points within our model set-up domain. We check the performance at different input resolution for the study basins. The model performances are evaluated by using several indexes to compare the hydrographs: Pearson's Moment Correlation Coefficients (PMC coeff.), the Nash-Sutcliffe coefficient (NSI coeff.), Index of Agreement (IOA), Covariance and Root mean square.

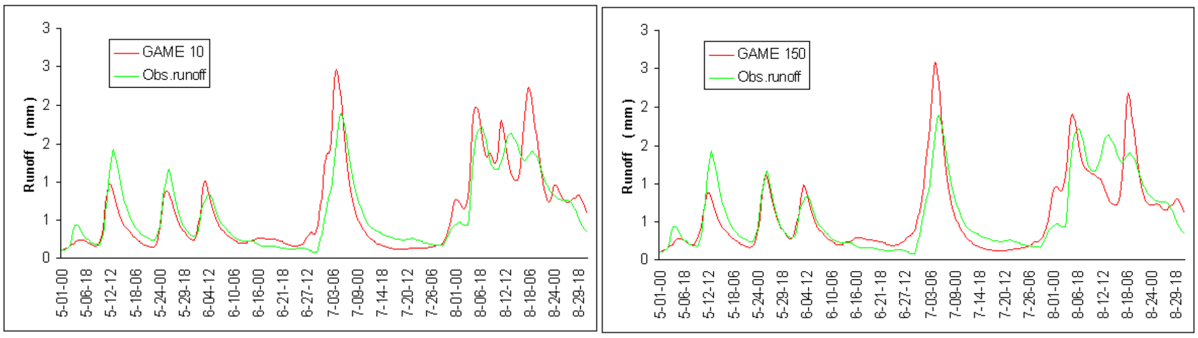
Distinct changes in the hydrographs are clearly visible for different spatial resolution data input. These changes indicate the influence of the spatial resolution and a clear indication of worse simulation results at coarser resolution (see Fig. 1). A point worth noting is that the coarser resolution data applied here for simulation is nothing more than the spatially-averaged value from the same finer resolution data that has kept the constant accumulation and same data pattern over the time series, warranting a just smoothened spatial variability of the data while the input water budget has remained constant.

IC-Ratio – a comparison index

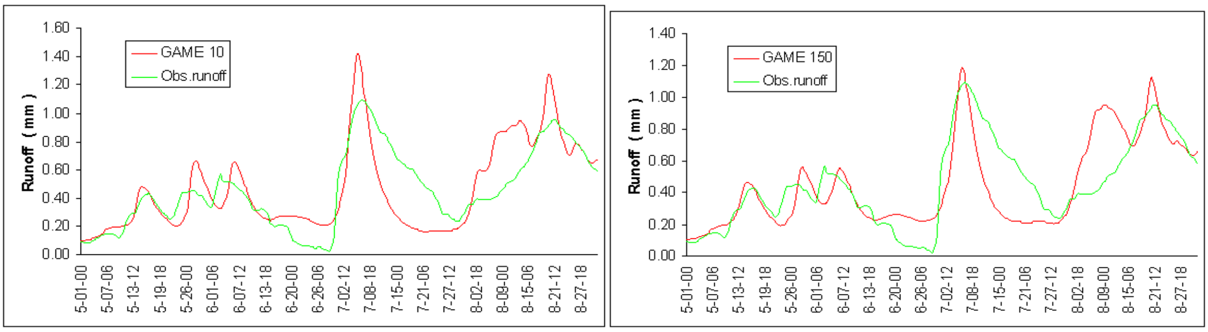
It is obvious to expect different model performances from the data having different orders of heterogeneity as a consequence of differences in forcing resolution. In addition to that, it is expected to obtain variable model performances where the study basins are different. The constitutive relations for different flow processes that are a function of the detailed geometry of flow pathways in different catchments are difficult to compare (Beven, 2002). Some indexes, for example average slope, flow lengths, watershed relief, etc., exist to represent the catchments features but they are not proper indices to describe the effect of forcing data resolutions and scale issues. The concept



(a) Comparison of hydrograph and water budget at Suiping with 10 minute and 2.0 degree data.



(b) Comparison of hydrograph and water budget at Wangjiaba with 10 minute and 2.5 degree data.



(c) Comparison of hydrograph and water budget at Bengbu with 10 minute and 2.5 degree data.

Fig. 1 The hydrographs with routing model: simulated *vs* observed.

of a ratio between the input grid resolution and the catchment's sizes, the IC-Ratio (Shrestha *et al*., 2002b), is a more useful and convenient index for investigating this issue.

Recall that the finest resolution in this experiment is 10-min spatial distance (~13.5 km). The input data for Suiping basin (2093 km^2) varies from 10-min to 2-degree resolution. Therefore the IC-Ratio value ranges from 1:11.77 to 1:0.08. Model performances are consistently better at 1:11.77 (called higher IC-Ratio values here) than near the ranges of 1:0.08 (lower IC-Ratio). A higher IC-Ratio corresponds to finer resolution and a lower IC-Ratio corresponds to a coarser resolution of input, when the catchment's size remains constant. The IC-Ratio values for Wangjiaba basin (29 844 km^2) and Bengbu basin (13 2350 km^2) vary from 1:167.8 to 1:0.75 and 1:744.5 to 1:3.3, respectively, within the framework of this experiment.

The trend of change in model performance from the graphs presented (see Fig. 2) clearly shows that the smaller basin has higher sensitivity in its response to the forcing

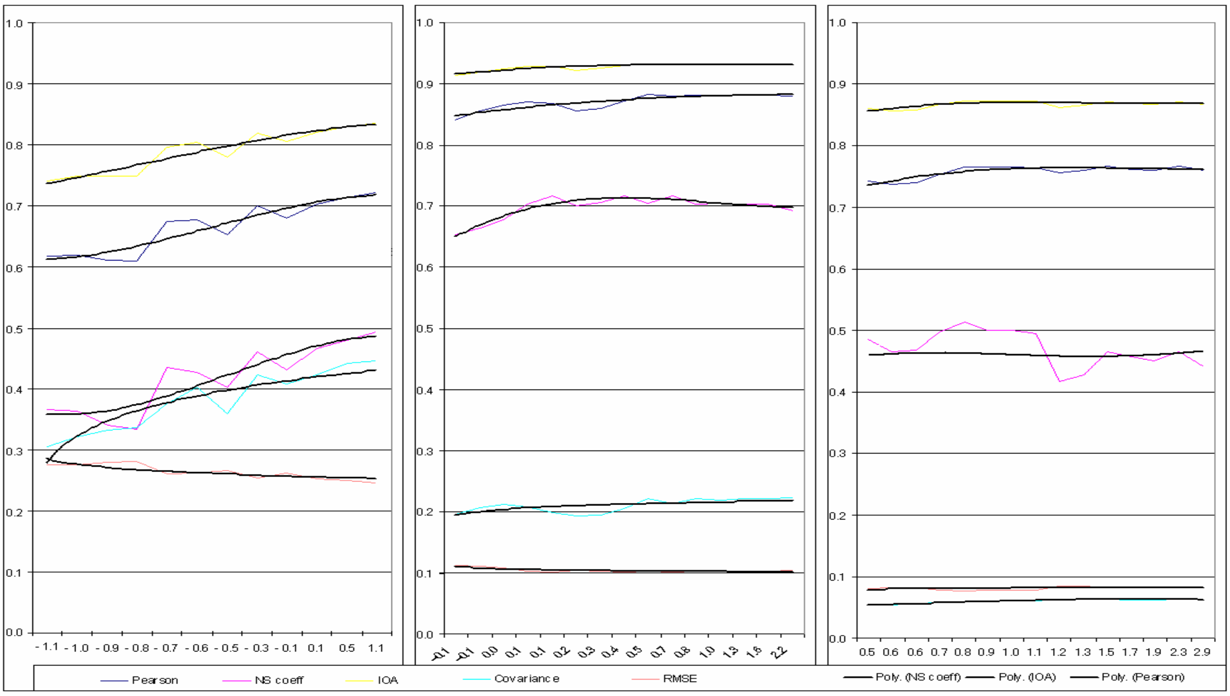


Fig. 2 Model performances of Suiping, Wangjiaba, and Bengbu at different IC Ratios. X-axis is IC Ratio, Y-axis is model performance index.

data resolutions. This indicates that the resolution issues for smaller basins need more attention than for larger basins, which seems quite natural because, in larger basins a larger portion of the heterogeneous field may be represented as a homogeneous segment than in smaller basins.

Improvement of the model performance due to higher resolution of forcing data has a tendency to saturate at some point. The rate of improvement is higher when the IC Ratio values are lower. As the IC Ratio value grows high, the rate of improvement of the model performance is not as significant as it is at the lower IC Ratio. Looking at the trend of RMSE values, the errors are found to stay very stable after the IC Ratio reaches 1:20 (see Fig. 3). Similar behaviour is displayed from the covariance between the observed and simulated discharges. This shows that the rate of improvement in model performance starts to saturate beyond an IC Ratio of 1:20 with respect to the increases in spatial resolution; the response appears to follow a trend of an S-shaped curve with its point of inflection near to 1:10 (Shrestha *et al*., 2002b).

SELECTION OF SCALE FOR UNGAUGED BASIN

Knowing the useful range of IC Ratio values helps perform simulations at a satisfactory level. It is of note that the simulation result with an IC Ratio lower than 1:10 has displayed the possibility of significant improvement if higher resolution input is provided, but this is untrue when the IC Ratio values are higher, for example 1:100 or more. We have seen that the results are always slightly better at the higher IC Ratio side but, at the cost of high-resolution data and calculation loads. Little improvement

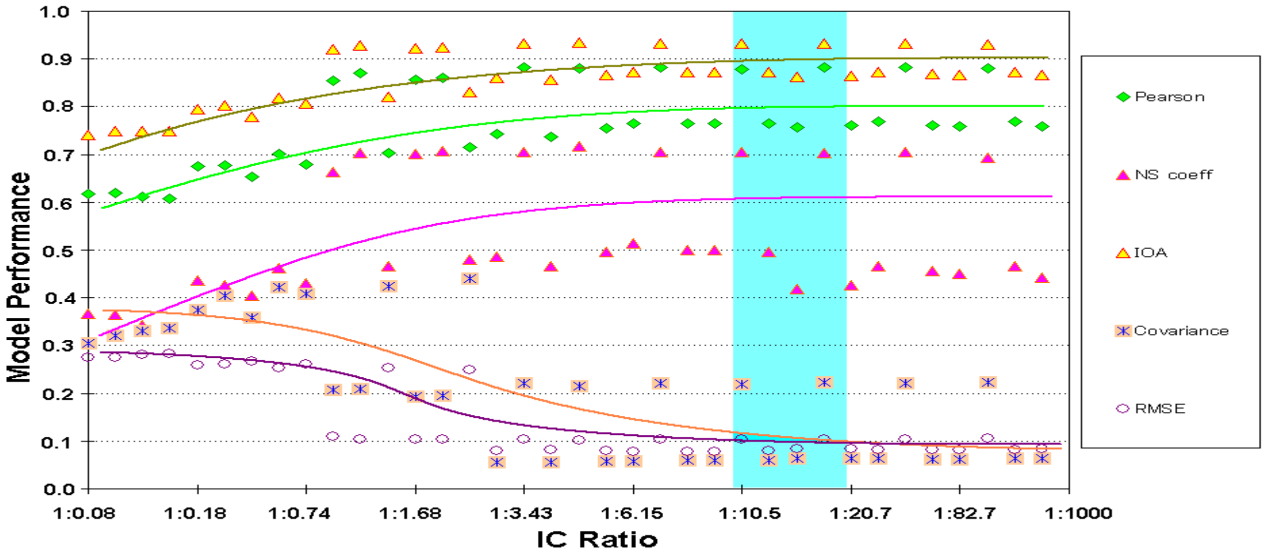


Fig. 3 Model performance *vs* IC Ratio.

in results using higher resolution input and modelling is likely to discourage the choice of high resolution data that maintains the IC Ratio value at 1:100 or 1:200. In fact, the performance improvement beyond IC Ratio 1:20 are no more attractive.

In an ungauged basin, where hydrologists often need to begin work from scratch, a lot of unknown challenges are faced before there is any success in model setting or analysis. The need for spatially-distributed data at the required scale with a sufficient level of accuracy, as demanded by a distributed hydrological model, definitely hinders the modeller from the start. Most likely, the data will either be unavailable or inadequate in ungauged basins. In both cases, it would compel a hydrologist to rely on external sources of information such as GCM data, remotely sensed data and DEMs.

Choosing to go for a higher resolution of distributed hydrological modelling would require substantial tasks in collecting finer scale details. There would be an additional task load in defining the model parameters, acquiring data, and performing computations. In addition, it includes a risk of persistent error amplification. On the other hand, selection of lower resolution reduces the workload considerably, but it has a risk of poor results and losing the advantages of the distributed modelling approach. Distributed modelling rarely yields a nice result using coarse and rough information. In both cases, the modelling effort may end up in failure. Therefore a trade-off is needed between these two sides of problems.

The results obtained in this experiment have demonstrated that an IC Ratio range of 1:10 to 1:20 produces satisfactory results. This means a target of 10 to 20 units of discrete block would not be a bad choice. If there are less than 10–20 units of distributization in the basin, the distributed model may not yield good results, which is more critical in smaller basins. Increasing the distributization has the tendency to always yield better results, but too much distributization, may not improve the results much. The model performance tends to level off when the numbers of blocks are increased beyond 20. This concept is likely to be very useful for setting up model frameworks and selecting a proper scale for distributed hydrological modelling in ungauged basins.

CONCLUSIONS

Better model performance is always the motto of a hydrologist. However, this depends greatly upon the quality of modelling, which refers to the quality of techniques employed, model parameters, and input data. The input data's quality is associated with accuracy and resolution issues. Among the catchments studied, the data resolution effects in the discharge simulation results are found to be more sensitive in the 2000 km^2 catchment range compared to other catchments. The catchments of the order of 132 350 km^2 did not require very high resolution input data to produce fairly good results. The sensitivity of input data resolution across various scales of catchment is assessed via the ratio of the input data resolution and the catchment area termed the IC-Ratio. It is argued that the IC-Ratio can serve as an index to evaluate the suitability of data resolution and thus guide for the selection of a proper scale for distributed hydrological modelling in ungauged basins.

Performance of distributed hydrological modelling depends very much on the forcing data resolution, which plays a significant controlling role in the selection of scale for modelling purposes. Demanding very high-resolution data is, however, not always a solution in view of the efforts required for processing high resolution data, particularly for an ungauged basin. Obtaining the best simulation results from the minimum data, minimum parameter and minimum calculation load is the primary goal for ungauged basins. The results of this study suggest that the simulation results are most favourable when the marginal IC Ratio value is approximately 1:10–1:20 as it trades off best with data resolution and model performance for a wide range of catchment scales. Thus this could be a good choice as a preliminary selection criterion of scale for distributed hydrological modelling in ungauged basins.

REFERENCES

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

Burlando, P. & Rosso, R. (2002) Effects of transient climate change on basin hydrology. *Hydrol. Processes* **16**, 1151–1175.

Ichikawa, Y., Shiiba, M., Tachikawa, Y. & Takara, K. (2000) Object-oriented hydrological modeling system. *HydroInformatics CD-ROM*, Cedar Rapids, Iowa, USA.

Kozan, O. & Ikebuchi, S. (2001) Estimation of energy and water budget in the Huaihe River Basin, China. GAME Publ. no. 28, 32–35. ftp://hydro.iis.u-tokyo.ac.jp/GAME/GAIN/GAME-REANL/.

Molnar, D. K. & Julien, P. Y. (2000) Grid-size effects on surface runoff modeling. *J. Hydrologic Engng* **5**(1), 8–16

Refsgaard, J. C. & Abbott, M. B. (1996) The role of distributed hydrological modeling in water resources management. In: *Distributed Hydrologic Modeling*, Chap. 1. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Shrestha, R., Tachikawa, Y. & Takara. K. (2002a) Effect of forcing data resolution in river discharge simulation. *Ann. J. Hydraul. Engng JSCE* **46**, 139–144.

Shrestha, R., Tachikawa, Y. & Takara. K. (2002b) Concept of IC Ratio in river discharge simulation using distributed hydrological model. In: *Proc. of Annual Conference of Civil Engineers 2002* (JSCE, Sapporo September 2002).

Tachikawa, Y., Takara, K., Ichikawa, Y. & Shiiba, M. (2000a) Test simulation of a macro scale distributed hydrological model for the Huaihe River Basin in China. In: *Proc. International GAME/HUBEX Workshop* (Sapporo), 139–146.

Tachikawa, Y., Takara, K., Ichikawa, Y. & Shiiba, M. (2000b) Development of a macro scale distributed hydrological model using an object-oriented hydrological modeling system. In: *HydroInformatics CD-ROM*. Cedar Rapids, Iowa, USA.

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. *Ann. J. Hydraul. Engng 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. *J. Japan Soc. Hydrol. & Water Resource* **15**(2), 139–151.

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

Zhao, R. J. (1992) The Xinanjiang model applied in China. *J. Hydrol.* **135**, 371–381.