Annual Journal of Hydraulic Engineering, JSCE, Vol.51, 2007, February

AN IMPROVED FRAMEWORK FOR THE PARAMETERS REGIONALISATION OF HYDROLOGICAL MODEL

Satish BASTOLA [1], Kuniyoshi TAKEUCHI [2] and Hiroshi ISHIDAIRA [3]

[1]Student member, Graduate Student, Department of Civil & Environmental Engineering, University of Yamanashi (4-3-11, Takeda, Kofu, Yamanashi, 400-8511, Japan)
[2]Member of JSCE, Dr of Eng., Professor, Interdisciplinary Graduate School of Medicine and Engineering, University of Yamanashi (Takeda 4-3-11, Kofu, Yamanashi, 400-8511, Japan)
[3]Member of JSCE, Dr of Eng., Associate Professor, Interdisciplinary Graduate School of Medicine and Engineering, University of Yamanashi (Takeda 4-3-11, Kofu, Yamanashi, 400-8511, Japan)

Regionalisation of hydrological model parameters is a simple approach to model ungauged basins. However, uncertainties in model parameters and catchment attributes hinder the regionalisation. In this context, this study proposes a methodology for modeling ungauged basins by pairing the regional model with the posterior distribution of parameters. The performance of regional models are evaluated by comparing the loss in performance and quantifying the uncertainty induced on the result of regionalisation. The study revealed the reduction in inconsistency among various regional models and improvement in performance while the results of regionalisation are constrained by regionalised ranges of parameters. Furthermore, the non-parametetric bootstrap methodology used to quantify of the uncertainty in regional models reveals that the indirect calibration method induced more uncertainty on the result of regionalisation compared to the conventional regionalisation schemes investigated in this study.

Key Words*: *Optimization, Rainfall-runoff models, Regionalisation schemes, Uncertainty

1. INTRODUCTION

The reliable prediction of streamflow at sites without adequate data quality and quantity remains a largely unsolved problem[1)]. To address such problems, the Prediction in Ungauged Basins (PUB) was initiated by International Association of Hydrological Sciences (IAHS) widening the scope of such kind of studies[2)]. Among various approaches to model ungauged basins, regionalisation is considered as one of the most popular approach. The regionalisation is identification of functional relationship between the optimal model parameter set and the catchment attributes (CAs) [1), 3), 4)]. Most of the regionalisation schemes follow a two-step approach: (a) estimation of parameters of hydrological models (MPs) at various gauged site, followed by (b) identification of relationship between MPs and CAs. Conceptual hydrological models are popularly used for regionalisation, and parameters of which are determined by calibrating model against observed data, but the calibrated MPs are not unique due to the inability of calibration procedures to uniquely identify a single best parameter set, errors associated with the system input and output, and model structural errors. All these source of uncertainty results in a scattered relationship between MPs and CAs, thus making regionalisation attempt week. Calibration of hydrological models (here after referred as local model interchangeably) and the construction of regional models both introduce uncertainties[1), 4), 5)] that are inevitably propagated into the model prediction.

Wide ranges of regional model structure have been implemented in the past to model ungauged basins[1)], which however, met with limited success so far mainly due to the poor identifiablility of parameters of local models. Within this context, this paper aims to develop the framework for regionalisation by: (a) making model parsimonious (b) implementing robust methods to calibrate model parameters (c) complementing the result of regionalisation with the information obtained from the posterior distribution of MPs and (d) evaluating the performance of various regional model structures by considering both the loss in performance (decrease in the model performance when the parameters obtained from regionalisation schemes are used instead of locally calibrated parameters), and the magnitude of uncertainty induced by regional models in model prediction.

2. METHODOLOGY

The proposed methodology for the development of an improved framework for regionalisation is as follows: (1) select the number of gauged basins with flow records extending over several years, and corresponding CAs that are readily available and relevant with the structure of hydrological model (2) identify both the best probable MPs and local posterior distribution of MPs for each basin (3) evaluate the performance of various regionalisation schemes using, Multiple linear regression (MLR), Artificial neural network (ANN), multiple polynomial regressions (MPR), Partial least square regression (PLSR) and indirect calibration methodology (4) regionalize the ranges of MPs (calculated from identified posterior distribution of MPs) with CAs, assuming that the basins with similar physical characteristics and data aspect will have similar ranges of parameters (4) the result of regionalisation is constrained by the ranges of parameters and (5) quantify the uncertainty induced by the parameters of regional model in the result of regionalisation.

(1) Hydrological model

TOPMODEL which is a variable contributing area physically-conceived semi-distributed hydrological model[6)] is selected in this study. TOPMODEL can be applied more accurately to catchments where the assumptions of the model are justified viz. primarily wet catchments that have shallow, homogeneous soils. For this study, the modified version of TOPMODEL was used. The modified version uses soil topographic index [7)] which provide more flexibility and capability to deal with heterogeneity of the catchment. In addition, the maximum root zone storage parameter was directly calculated from root zone depth and soil properties[6)] instead of calibration. This will increase the identifiablility of the parameter while reducing the number of parameter [1)]. Only three parameters: lateral saturated transmissivity, T_o (m^2/h), time constant, T_d (h/m), and decay parameter, m (m) [6)] were calibrated in this study.

(2) Study area

The study area consists of 26 basins located in different geographic and climate zones (Table 1). The basins located in various climates and geographic region were selected to incorporate wide range of basin attributes which are outlined in Table 2. Other data includes: The 90m DEM from Shuttle Radar Topography Mission (SRTM), soil data from Food and Agriculture Organization (FAO), and land use data from International Geosphere-Biosphere Program (IGBP). Most of study basins are humid with a wetness index greater than 1.

Table 1 Description of study basins and calibrated model parameters

Country	Catchment ID	Area km^2	Calibrated model parameters		
			m	T_o	T_d
[1]Australia	145018	81	0.23	2.5	1.07
	[6]204016	104	--	--	--
	204017	82	0.06	4.95	1.23
	218001	93	0.09	1.04	4.99
	302200	448	0.24	1.03	1.08
[2]Nepal	330	1980	0.13	6.08	1.00
	795	1148	0.03	5.64	7.81
	390	554	0.07	3.21	1.22
	[6]395.5	683	--	--	--
[3]Japan	Arakawa(Yorii)	927	0.02	4.00	9.09
	[6]Ukaibashi	487	--	--	--
	Torinkyo	1095	0.04	4.00	1.22
[4]UK	23006	331	0.02	3.40	0.10
	27034	510	0.02	4.80	5.48
	[6]27035	282	--	--	--
	62001	893	0.02	6.21	0.30
	66011	344	0.01	4.25	8.90
[5]France	J3024010	43	0.08	4.86	2.31
	J4124420	32	0.14	5.49	1.10
	[6]K0744010	181	--	--	--
	J4712010	142	0.04	6.23	1.45
	H2001020	98	0.04	4.20	3.46
	Y5615030	297	0.05	4.82	2.39
	K0753210	371	0.04	4.48	3.10
	K0813020	193	0.04	4.20	1.28
	V3517010	25	0.04	8.72	2.28

[1]catchments located in eastern Australia, and data were obtained from http://www.stars.net.au/tdwg/?datasets, [2]catchments located in Middle mountain physiographic region of Nepal and data were obtained from Department of Hydrology and Meteorology (DHM), Nepal, [3]catchmetns located in Japan, and data were obtained from Ministry of Land, Infrastructure and Transport (MLIT), Japan, [4]catchments located in UK, and data were obtained from http://www.ceh.ac.uk/data/nrfa/index.html, [5] located in France, and data obtained from Model Parameter Estimation Experiment (MOPEX)-France,[6]catchments used for the validation of regionalisation schemes.

Table 2 Catchment attributes selected for regionalisation

Indices	Catchment attributes (CAs)
Physiographic	Area, drainage density, average basin slope, basin shape factor, average topographic index, shape and scale parameter of gamma distribution fitted to the distribution of topographic index, average saturated hydraulic conductivity (basin scale), average maximum root zone depth calculated using soil and land cover map[6)](basin scale)
Climatic	Annual average rainfall, variance of monthly rainfall, annual average potential evapotranspiration (PET), wetness index (calculated as the ratio of average annual rainfall to average annual PET)

(3) Model Calibration

a) Multiobjective Optimization

In this study, multiobjective shuffled complex evolutionary metropolis (MOSCEM-UA) is used to calibrate the parameters of the local model as MOSCEM-UA allows the simultaneous estimation of best parameters along with the underlying posterior distribution of parameters (assuming the initial Gaussian assumptions made for hydrologic parameters) for hydrological models having only few parameters and that requires less model evaluation time[8]. Earlier work on automatic calibration of hydrological models suggests that single objective functions are not adequate to reproduce different aspects of the hydrograph, which led to the calibration being treated as a multiobjective problem. In this context, evolutionary algorithms (EA's) have been recognized as possibly well suited to multiobjective optimization since they can search for multiple solutions in parallel. Among many EA's developed, the non-dominating sorting genetic algorithm (NSGAII) developed by Dev *et al.* [9] and MOSCEM-UA developed by Vrugt *et al.* [10] are two popular multiobjective EA's based on the Pareto domination approach.

b) Objective Function

The identification of a best parameter set is necessary for meaningful prediction of flow and parameter regionalisation. This can be better realized by using multiple objective criteria during calibration, and the objective functions used in this study are as follows:

1) The Nash Sutcliffe efficiency (NSE), it assumes that errors are normally distributed and homoscedastic.
2) NSE for the transformed flow is used to consider the heteroscedastic variance in flow. The flow is transformed explicitly before evaluating the objective function[11] by using, $z=[(y+1)^{\lambda}-1]/\lambda$ where λ =0.3, z is the transformed flow, and y is the observed flow
3) NSE for low flow and NSE for peak flow [12].

(4) Regionalisation schemes

The regionalisation schemes can be broadly classified as conventional approach and improved variants of conventional approach[1]. The conventional approach, here referred to as "direct calibration method", at first, calibrates the parameters of hydrological model at all site independently, and then attempts to identify the functional relationship between MPs and CAs (Fig 1(a)). The identified functional relationship between MPs and CAs are later used for the prediction of model parameters at ungauged basin from the measurable CAs. These approaches despite being straight forward are hindered by weak

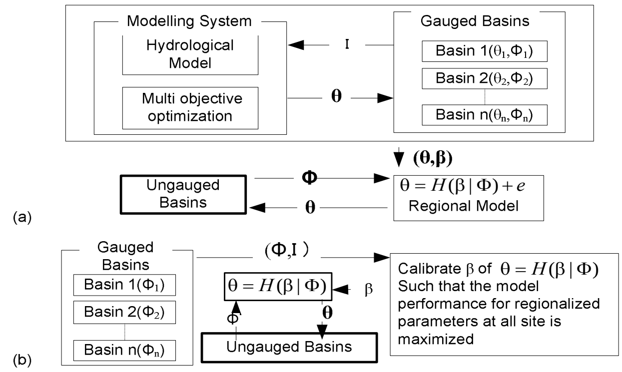

Fig.1 Regionalisation schemes: a) Direct calibration method, and b) Indirect calibration method, where θ is the vector of model parameters, β is the vector of regional parameter, Φ is the vector of catchment attributes, H (.) is a functional relationship between θ and β, I are input time-series, and e is the error term.

identifiablility of MPs. Regionalisation schemes used in this study which fall under direct calibration methods are, MLR (Equation 1), MPR (Equation 2), PLSR and ANN.

$$\theta_j = a_j + \sum_{i=1}^{m} \Phi_i \beta_{i,j} \tag{1}$$

$$\theta_j = a_j + \sum_{i=1}^{m} (\Phi_i \beta 1_{i,j} + \Phi_i^{2} \beta 2_{i,j}) \tag{2}$$

where θ_j is j^{th} model parameter, Φ_i is the i^{th} CAs, $\beta_{i,j}$, $\beta 1_{i,j}$, $\beta 2_{i,j}$ are the i^{th} regional parameter of regression equation for j^{th} model parameter, m is the number of CAs (Table 2), and a_j is the constant. Model structure based on MLR and MPR were used in this study due to their simplicity. On the other hand, regionalisation scheme based on ANN were also used, as it is more flexible modeling structure that can easily account for nonlinearities and interaction effect. The ANN used in this study consists of single input, output and hidden layer. In addition, the partial least square regression[13] (PLSR), a method for constructing prognostic models when the factors are many and highly collinear, is also used in this study. Similar to principal components regression (PCR), PLSR produce factor scores as linear combinations of the original predictor variables, so that there is no correlation between the factors score variables used in the predictive regression model. Contrary to PCR, PLSR produces the weight matrix reflecting the covariance structure between the predictor and tries to extract those latent factors that account for most of the variance.

Indirect calibration method is an improved variant of conventional approach which attempts to calibrate the model at all sites simultaneously, while concurrently attempting to achieve the best possible regional relationship between MPs and CAs (Fig 1(b)). In the indirect calibration method the approximate functional relationship between model parameter and CAs was assumed first. Then the

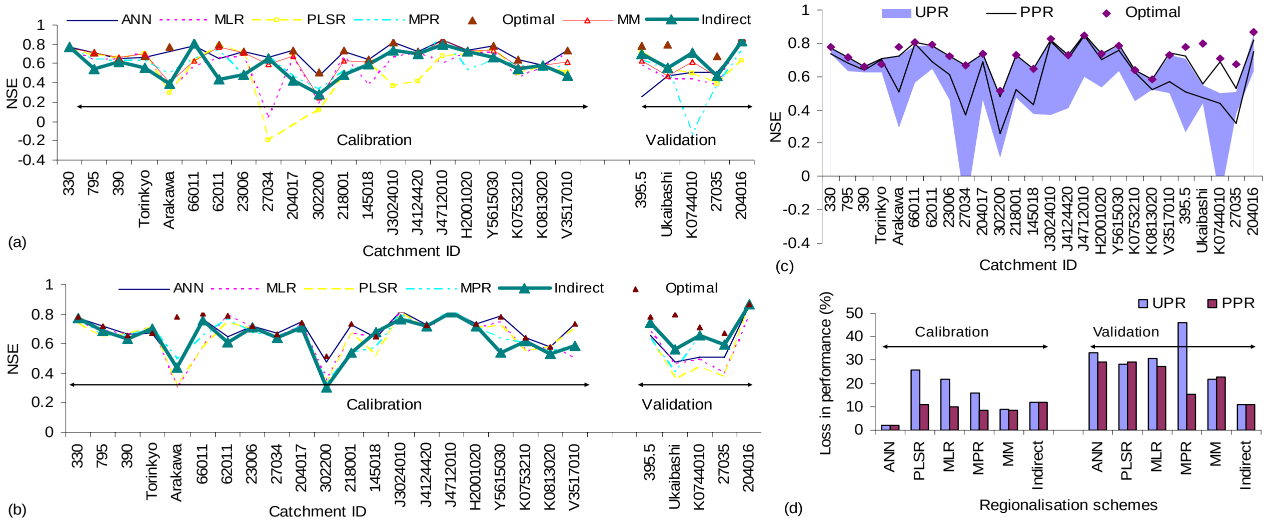

Fig.2 Performance of regionalisation schemes: (a) Unpaired with prior ranges of parameter (UPR), (b) Paired with prior ranges of parameters (PPR), (c) Ensemble performance (NSE) with and with out pairing regional models with prior ranges, and (d) Comparison of loss in model performance for regionalisation schemes.

parameters of such functional relationships were obtained using multiobjective calibration, which can be stated as:

$$\text{Maximized } F(\theta) = \{f_1^*(\theta), f_2^*(\theta), .., f_q^*(\theta)\} \quad (3)$$

where $f_q^*(\theta) = (\sum_{i=1}^{n} f_{i,q}(\theta))/n$ is the average value of the q^{th} objective function(sec 2(3)b), n is the number of basins, q is the number of objective function, $f_{i,q}(\theta)$ is the value of q^{th} objective function for i^{th} basin. The multiobjective method (NSGAII was used to calibrate the regional parameters in the indirect method) results in Pareto optimal set of functional relationships, so the compromised regional parameter set was used to identify the single best functional relationship to estimate MPs for ungauged basins.

In calibration phase, 21 study basins were used whereas 5 basins (see table 1) were used for the appraisal of the performance of regionalisation schemes. For the evaluation purpose, two metrics were used: (1) loss of model performance and (2) uncertainty induced by the regional parameters on model prediction.

(5) Uncertainty in regional models

To assess the uncertainty due to weighting parameter on regionalisation result, the non-parametric bootstrap method was used. In this method, the sub sample of size 19 basins is repeatedly used for regionalisation. Each sub sample is a random sample with replacement from the full sample of size 21. This will lead to the realization of numerous sets of model parameter for each basin and subsequently results in an ensemble of simulated hydrograph. This combination leads to the multiple realizations of functional relationships for all schemes (ANN, MLR, MPR and PLSR) which results in a set of regionalized MPs for each basin. These regionalized model parameters results in the ensemble of simulated hydrograph which is here used as the measure to quantify the uncertainty inherent in regionalisation schemes and compare various schemes. In addition, the indirect calibration also results in a multiple set of regional parameters for similar performance resulting in uncertainty in simulated flow. Similar to direct calibration method, the ensemble of the simulated flow resulted by the use parameter that lie in Pareto optimal front were used here to enumerate the uncertainty in the indirect calibration method.

3. RESULTS AND DISCUSSION

The parameters of modified TOPMODEL were calibrated using 3 years of daily hydro-meteorological data for all selected catchments using MOSCEM-UA (see Table 1). Only a small number of CAs was correlated with MPs at the 10% significance level. Parameter *m* is found to be correlated with wetness index (-0.48), average annual rainfall (AAR) (-0.32), and average maximum root zone depth (ASR) (0.28). Parameter *To* is found to be correlated with, ASR (-0.36), mean elevation (0.28) and basin area (0.41). Similarly, *Td* is found to be correlated with AAR (0.56), variance of monthly rainfall (0.77), drainage density (-0.5) and average basin slope (0.48). The performance of regionalisation schemes measured with respect to NSE over both calibration and validation are shown in Fig. 2(a). Regionalized values of MPs obtained from ANN, MLR, MPR and PLSR resulted in high inconsistency in the model performance (NSE). The ensemble performance for regionalisation schemes are shown as the shaded region in Fig. 2(c). Considering the result over the

Table 3 Standard error estimates for regionalized model parameters for regionalisation schemes paired with prior ranges (PPR) and schemes unpaired with prior ranges (UPR).

Regionalisation schemes	Standard error for the parameters estimated from regionalisation schemes											
	Calibration of regionalisation schemes						Validation of regionalisation schemes					
	m		To		Td		m		To		Td	
	PPR	UPR	PPR	UPR	PPR	UPR	PPR	UPR	PPR	UPR	PPR	UPR
ANN	0.004	0.003	0.33	0.09	1.08	1.081	0.069	0.078	2.95	2.897	3.955	4.017
MLR	0.031	0.042	1.21	1.658	2.047	2.234	0.044	0.047	1.632	1.361	4.174	5.629
MPR	0.025	0.033	0.619	0.89	2.045	2.18	0.028	0.027	3.02	3.762	3.989	4.222
PLSR	0.03	0.04	0.961	1.247	2.03	2.182	0.046	0.054	2.456	2.44	4.02	5.124
Indirect method	0.024	0.024	1.235	1.235	4.175	4.175	0.026	0.026	1.401	1.401	5.677	5.677
MM	0.022	0.027	0.759	0.826	1.598	1.65	0.019	0.017	1.204	1.205	1.895	2.214

calibration period, PLSR and MLR scantily captured the underlying relationship between MPs and CAs resulting in high loss in performance. Similarly, the regionalized parameters obtained from ANN resulted in higher loss in model performance in basins considered for validation, though it closely followed the calibrated model performance on basins which were included for the calibration of regional model. Performances of MPR for both calibration and validation phase are apparently better. The marginal improvement in performance is observed (see standard error estimate for regionalize MPs in Table 3, and performance loss (%) in Fig 2(d)) with the averaged values of the regionalized parameters obtained from various structures, referred to as model mean (MM). Over calibration and validation, the performance of indirect calibration with respect to NSE was only marginally worse than the results obtained from calibrated model parameters (Indirect in Fig. 2(a)), however provided unbalanced model performance (improved performance in some basins and poor in others). This unbalanced model performance could be mainly due to the use of the average objective function as objective criteria to calibrate the parameters of regional models.

(1) Pairing of the regional model with prior ranges of parameters

In order to minimize the inconsistency observed among regionalisation schemes, and to improve their predictive performance, pairing the result of regionalisation with the information from posterior distributions of parameters is investigated. To estimate the posterior distribution of parameters, MOSCEM-UA was used for the calibration of MPs, which estimates best MPs along with the posterior distribution of parameters. Using the estimated posterior distribution of MPs, the ranges of MPs were estimated for each basin. In order to estimate the ranges of MPs for ungauged basins from easily measurable CAs, the statistical relationship between ranges of MPs and CAs was determined using ANN. The performance of ANN in simulating the ranges of parameters was found to be efficient during

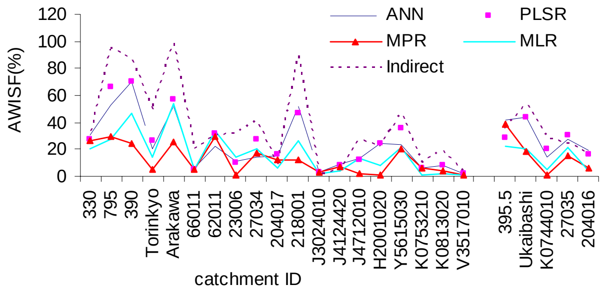

Fig. 3 Effect of parameter uncertainty of regionalisation on model prediction

calibration and sensible during validation. The regionalized ranges (here after referred as prior ranges interchangeably) of parameters were then used to inflict constrains on the result of regionalisation. Fig. 2 (b) shows the performance of regionalisation schemes constrained by the prior ranges of parameters. The shaded zone in Fig. 2 (c) is the ensemble of model performance (NSE) obtained from regionalized MPs obtained from various regionalisation schemes unconstrained from prior ranges of MPs (UPR). The region within continuous lines in Fig.2(c) is the ensemble model performance (NSE) when the regionalized MPs were paired with the prior range of parameters (PPR). The ranges of ensemble performance for constrained regionalisation are narrowed compared to the scheme unpaired with prior ranges of MPs. This narrowing in the performance range implies the reduction in the inconsistency in model performances among various regional model structures. In order to compare the performance of regionalisation schemes with and without prior ranges of MPs objectively, the loss in model performance and standard error in the estimate of regionalized value of MPs were used. The loss of model performance shown in Fig. 2(d), and standard error estimate (Table 2) of the regionalized MPs obtained from regionalisation schemes with and without the use of prior ranges of MPs reveals the apparent improvement in regionalisation when prior ranges of MPs were used.

To enumerate the uncertainty, the non-parametric bootstrap method as explained earlier was used for ANN, MLR, MPR and PLSR. However, for indirect calibration method, Pareto optimal regional parameter sets were used to quantify the uncertainties. In both approaches, the average width of the interval of simulated flow (AWISF) expressed in terms of percentage (%) in all basins (calculated as the ratio of average width of ensemble hydrograph to average regionalized flow) was used as a measure to quantify the uncertainty in model prediction. For all schemes, high value of AWISF is observed for two Australian basins (national ID 145018 and 302200). These two basins have appreciably lower runoff coefficients and wetness index compared to other basins, so they were removed from further discussion. As the variation in the AWISF among schemes can be used as a measure to compare various schemes, so the value of AWISF were obtained for all regionalisation schemes which is shown in Fig. 3. The values of AWISF were significantly higher for indirect method with an average of 38% compared to MPR (12%). For validation, MLR (15%) and MPR (16%) resulted minimal uncertainty in flow compared to PLSR (28%), ANN (29%) and indirect method (32%). Indirect calibration resulted in lesser performance loss, but induced higher uncertainty in the simulated flow compared to direct calibration method. In addition, ANN induced larger uncertainty in flow compared to MLR and MPR, which could be due to a larger number of free parameters.

4. CONCLUSION

The following conclusions are drawn from this study:

- Averaging the regionalized model parameters obtained from various schemes (MLR, ANN, MPR and PLSR) along with indirect calibration methodology demonstrated the prospect of their application in modeling ungauged basin.
- Constraining the result of regionalisation schemes by prior ranges of MPs apparently reduce the inconsistency among regional model and improved the predictive capability.
- The effect of uncertainty in the parameters of regional models on simulated flow varied in an average from 12 to 38% among regionalisation schemes considered.
- Indirect calibration resulted in lesser loss in model performances, but the uncertainty induced in simulated flow was highest compared to direct calibration method, and among direct calibration methods, ANN resulted in the highest uncertainty in model prediction compared to regression based schemes.
- The uncertainty induced by regionalisation schemes on simulated flow along with loss in model performance is indispensable for the comparison of various regionalisation schemes.

ACKNOWLEDGMENT: The authors express special thanks to MEXT and COE program at University of Yamanashi for providing financial support.

REFERENCES

1) Wagener, T., Wheater, H.S. and Gupta, H.V.: Rainfall-Runoff Modelling in Gauged and Ungauged Catchments, Imperial College Press, London, 300, 2004.

2) Sivapalan, M., Takeuchi, K., Franks, S.W. Gupta, V.K., Karambiri, H., Lakshim, V., Liang, X., McDonnell, J.J., Mendiondo, E.M., Connell, O., Oki, T., Pomeroy, J.W., Schertzer, D., Uhlenbrook, S. and Zehe, E.: IAHS Decade on Predictions in Ungauged Basins (PUB), 2003-2012: Shaping an exciting future for the hydrological sciences, *Hydrol. Sci. J.*, Vol. 48 (6), pp. 857–880, 2003.

3) Merz, R. and Blöschl, G.: Regionalisation of catchment model parameters, *J. Hydrol*, Vol. 287, pp.95-123, 2004.

4) Heuvelmans, G., Muys, B. and Feyen, J.: Regionalisation of the parameters of a hydrological model: Comparison of linear regression models with artificial nets, *J. Hydrol*, Vol. 319, pp. 245-265, 2006.

5) Kuczera, G. and Parent, E.: Monte Carlo assessment of parameter uncertainty in conceptual catchments models: the metropolis algorithm. *J. Hydrol*, Vol.211, pp.69-85, 1998.

6) Beven, K.J., Lamb, R., Quinn, P., Romanowicz, R. and Freer, J.: TOPMODEL, *In V.P. Singh (Ed.) Computer models of watershed hydrology, Water Resource, Publication, Colorado*, pp. 627-668, 1995.

7) Sivapalan, M, Beven, K.J and Wood, E F.: On Hydrological Similarity: 2. A Scaled Model of Storm Runoff Production, *Water Resour. Res*, Vol. 23(12), pp. 2266-2278, 1987.

8) Tang, Y., Reed, P. and Wagner, T.: How effective and efficient are multiobjective evolutionary algorithms at hydrologic model calibration?, *Hydrology and Earth System Sciences Discussions*, Vol. 2, pp. 2465-2520, 2005.

9) Dev, K., Agrawal. S, Pratap, A. and Meyarivan, T.: A Fast and elitist multi-objective Genetic Algorithm: NSGA-II, *IEEE transactions on Evolutionary Computation*, Vol. 6, pp. 182-197, 2002.

10) Vrugt, J. A., Gupta, H. V., Bastidas, L. A., Bouten, W. and Sorooshian, S.: Effective and efficient algorithm for multiobjective optimization of hydrologic models, *Water Resour. Res.*, Vol. *39*(8), doi: 10.1029/2002WR001746, 2003.

11) Misirli, F., Gupta, H. V., Sorooshian, S. and Thiemann, M.: Bayesian recursive estimation of parameter and output uncertainty for watershed models, *In Q.Duan et al., (Ed.) Calibration of Watershed Models, Water Sci. Appl. Ser., Washington, D. C.,* Vol. 6, pp. 113–124, 2003.

12) Madsen, H.: Automatic calibration of conceptual rainfall-runoff model using multiple objectives, *J. Hydrol*, Vol. 235, pp. 276-288, 2000.

13) Randall, D. Tobias: An Introduction to Partial Least Squares Regression, TS-509, SAS Institute Inc., Cary, N.C. 1997.

(Received September 30, 2006)