Quantification and Reduction of Predictive Uncertainty for Sustainable Water Resources Management (Proceedings of Symposium HS2004 at IUGG2007, Perugia, July 2007). IAHS Publ. 313, 2007.

Regionalization of parameters of hydrological models: inclusion of model parameter uncertainty

SATISH BASTOLA, HIROSHI ISHIDAIRA & KUNIYOSHI TAKEUCHI

Department of Civil & Environmental Engineering, University of Yamanashi, 4-3-11, Takeda, Kofu, Yamanashi, 400-8511, Japan

satish@ccn.yamanashi.ac.jp

Abstract Regionalization of hydrological model parameters is a simple approach to model ungauged basins, but the uncertainties in model parameters and regionalization schemes hinder such an approach. To address the effect of model parameter uncertainties on regionalization, the vectors of model parameters generated from the posterior distribution of the parameters were regionalized and combined with the model parameters estimated from a non-parametric bootstrap method. In this study, 26 catchments from different regions in the world have been used. The study reveals that the effect of uncertainities in model parameters are significant, and the effect of uncertainities in regionalization propagated through generalized regression schemes were higher compared to univariate and correlated regression based schemes. Finally, the proposed methodology was validated by comparing the ensemble of simulated flow resulting from regionalized vectors of model parameters with the one produced by the model parameters generated from the posterior distribution of parameters.

Key words optimization; rainfall–runoff models; regional models; regionalization; uncertainty

INTRODUCTION

Conceptual hydrological models (CRR) are a popular approach to the modelling of flow at gauged sites and are often applied for the temporal as well as spatial extrapolation of hydrological data. Though some parameters of CRR have a physical basis, they are difficult to measure in the field since they represent effective values at the catchment scale. Consequently, calibration of the model against observed data is essential to identify the parameters of the model. Due to the limited availability of data for model calibration, extrapolation of hydrological data to other basins with no extant observations is one of the most fundamental challenges for hydrologists (Wagener *et al.*, 2004). To address such challenges, the International Association of Hydrological Sciences (IAHS) initiated the Prediction in Ungauged Basins (PUB) programme (Sivapalan *et al.*, 2003). The identification of statistical relationships between model parameters (MPs) and catchment attributes (CAs), generally referred to as conventional schemes or regionalization, is a popular approach to modelling ungauged basins (Seibert, 1999; Merz & Bloschl, 2004; Wagener *et al.*, 2004, Griet *et al.*, 2006). For regionalization, the model structure that provides better model performances and improves the identifiability of parameters is desired, but the trade-off that exists among model complexities, model performance and identifiablility of MPs pose a problem in selecting the most suitable model structure. However, the use of a parsimonious model structure with better-identified parameters at the cost of model performance provides an opportunity for sensible regionalization (Wagener *et al*., 2004). To circumvent the problem of poor identifiablility of MPs, improved variants of the conventional scheme have been discussed in the past. These improved variants either incorporate the information of identifiablility of parameters while constructing the regional models (e.g. weighted regression, Wagener *et al*., 2004), or attempt to improve the identifiablility of parameters, (e.g. sequential regionalization, Lamb, 2000; regional calibration, Fernandez *et al*., 2000). The discussion of the uncertainties and problems of sequential regionalization and the weighted regression method for regionalization can be found in Wagener & Wheater (2006). Apart from the issue of selection of model structure, the proper selection of catchment attributes is also equally important for regionalization. Landscape attributes are generally used as CAs, because it is easier to relate MPs of parsimonious model to landscape attributes. But if the geographical scope of the study is widened, additional attributes related to climate are essential. Due to a lack of objective criteria for the selection of the structure of hydrological models, the inability to identify unique values of MPs, and the subjectivity involved in the selection of CAs and regional models, considerable uncertainties are induced in the process of regionalization that inevitably propagates to model prediction. It is, therefore, important to assess these uncertainties and propagate their effect into the model prediction. Within this context, this paper proposes a methodology to propagate the uncertainties in the regionalization of the parameters of continuous hydrological models (daily time scale) through a modelling system.

METHODOLOGY

The statistical approach for identifying the functional relationship between the MPs and CAs, generally referred to as conventional schemes and commonly expressed as $\theta = H(\beta\,|\,\Phi) + e$, where θ are the MPs, β are the regional parameters, Φ are the CAs, and e is the error term, and $H(.)$ is the regional model structure that relates θ with Φ, are popular for regionalization. These schemes, at first, calibrate the MPs at all basins independently, and then attempt to identify the functional relationship between MPs and CAs. Regionalization schemes selected in this study are: multiple linear regressions (MLR), multiple polynomial regressions (MPR), artificial neural network (ANN) with three layer feed-forward network, and partial least square regression (PLSR) (Randall, 1997), which is a method for constructing prognostic models when factors are many and collinear. The use of regional models (e.g. MLR, MPR, ANN and PLSR) for regionalization is straight forward, but is often hindered by the non-uniqueness in the calibrated MPs arising from 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.

Uncertainty in the regionalization schemes

In this study, the uncertainties associated with regionalization are categorized as: (a) the uncertainty in regionalization schemes, and (b) uncertainty in the calibrated parameters of hydrological models. To deal with the uncertainties in regionalization, Lamb & Kay (2004) used a standard uncertainty around the regional parameter estimate and Griet *et al.* (2006) used a non-parametric bootstrap method to quantify the effect of uncertainty in regionalization schemes. However, these studies did not explicitly take into account the effect of the uncertainties in MPs while attempting regionalization. Similarly, Merz & Bloschl (2004) addressed the issue of uncertainties in MPs through a comparison of MPs for two sub periods and Wagener *et al.* (2004) explicitly incorporated the identifiablility of MPs during regionalization. In this study, the non-parametric bootstrap method was used to assess the uncertainty in regionalization schemes in which sub samples of size *m* (*m* is the number of basins) were randomly sampled from *n* basins, and were repeatedly used for the determination of the parameters of regional models. This leads to the estimation of multiple values of MPs and subsequently estimates ensembles of simulated flow for the target basin. In addition, the parameters of hydrological models are also uncertain (Beven & Binley, 1992; Kuzera & Parent, 1998), which induces considerable uncertainties in regionalization. So in this study, in order to incorporate the effect of hydrological model parameter uncertainty, regionalization schemes were formulated by using the vectors of MPs instead of individual values of MPs, so that the effect of the uncertainty in MPs can be propagated to model prediction via regionalization schemes. As the propagation of the uncertainty in MPs via regionalization using vectors of MPs will inherently be affected by the uncertainties in the scheme itself, the MPs estimated from the non-parametric bootstrap approach were combined with the MPs estimated from the regionalization of vectors of parameters. The outline of methodology adopted in this study (Fig. 1) is as follows:

1. Sample *m* basins randomly with replacement from n number of basins repetitively and identify the regional relationship between the calibrated MPs and CAs of *m* selected basins. This will lead to the estimation of sets of MPs for each target basin (referred to as M1).
2. Regionalize the vector of MPs sampled from the posterior distribution of parameters using various regionalization schemes (referred to as M2 in Fig. 1). If the posterior distribution of parameters follows the multivariate normal distribution, the exponent in the multivariate normal distribution is a chi-square variate with *p* degrees of freedom so that the MPs lying in a pre-specified confidence region can be sampled by selecting the parameter set for which $(\theta-\mu)^T\Sigma^{-1}(\theta-\mu)<\chi^2_{\alpha-1,p}$ is satisfied, where Σ is a covariance matrix evaluated at μ (posterior mean), θ is the vector of MPs and α is the significance level.
3. Combine the regionalized value of MPs obtained from Step 1 and Step 2 for the target basin (referred to as M3 in Fig. 1) to quantify the effect of both the uncertainties in MPs and regionalization schemes.
4. Calculate the average value of each set of regionalized parameters (referred to as MM) obtained from various regionalization schemes used in this study (ANN, MLR, MPR and PLSR) as $MM_{j,k}=(\hat{\theta}^k_{j,1}+..\hat{\theta}^k_{j,i}...+\hat{\theta}^k_{j,l})/l$, where $MM_{j,k}$ is the *k*th set of *j*th model parameters, $\hat{\theta}^k_{j,i}$ is the *k*th regionalized value of *j*th MPs obtained from *i*th regional model structure and *l* is the number of the regional model structure.

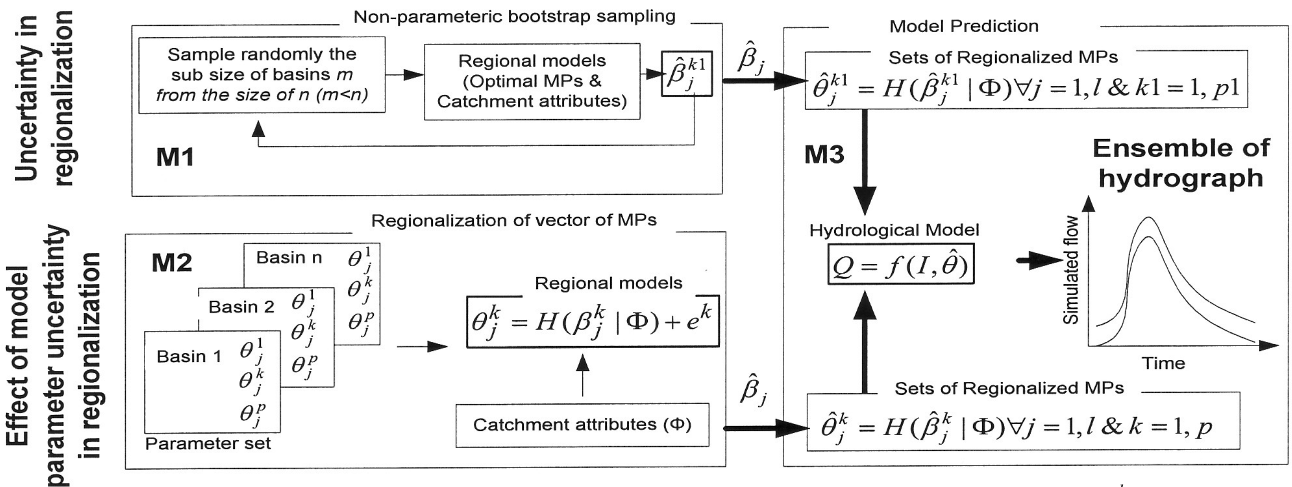

Fig. 1 The outline of the methodology adopted for regionalization, where Q^k is the simulated flow for kth set of regionalized model parameters, I_i is the input, θ^k_j is the k^{th} set of jth model parameter, $\hat{\theta}$ is the estimated MPs, $\hat{\beta}_j$ is the vector of jth estimated regional parameters, $p1$ and p are the sets of regional parameters for nonparametric bootstrap sampling and regionalization of vectors of MPs, n is the number basins, m is the size of the sub sample of basins and l is the number of model parameters (MPs).

CASE STUDY: Regionalization of hydrological model parameters

TOPMODEL (Beven *et al.*, 1995), which is a variable contributing area physically-conceived semi-distributed hydrological model, is selected in this study. TOPMODEL can be applied more accurately to catchments where the assumptions of the model are justified; primarily wet catchments that have shallow, homogeneous soils. The original TOPMODEL with the modification of the soil topographic index (Sivapalan *et al.*, 1987), that provides more flexibility and capability to deal with heterogeneity of the catchment, was used. In addition, the maximum root zone storage parameter was directly calculated from observed root zone depth and soil properties instead of being calibrated (Beven *et al.*, 1995). The calibrated parameters are lateral transmissivity (*To*) (m^2 h^{-1}), time constant (*Td*) (h m^{-1}), and decay parameter (m) (m). Though the parameters of TOPMODEL are physically interpretable, they are values effective at the catchment scale. So, the parameters of TOPMODEL have to be determined by calibration against measured streamflow data. As a multi-objective approach facilitates in retrieving more information from the observed data, and can also provide insight into parameter uncertainty and limitation of model structure (Gupta *et al.*, 1998), the multi-objective shuffled complex evolutionary metropolis (MOSCEM-UA) developed by Vrugt *et al.* (2003) was used. For multi-objective calibration of MPs, the following objective functions were used: (a) the Nash-Sutcliffe efficiency (NSE); (b) the NSE for the transformed flow to consider the heteroscedastic variance in flow, here the flow is transformed explicitly by using, $z = [(y + 1)^{\lambda} - 1]/\lambda$, where $\lambda = 0.3$, z is the transformed flow, and y is the observed flow before evaluating the objective function; (c) the NSE for low flow; and (d) the NSE for peak flow. As different objective criteria focus on different aspect of the hydrograph, the MPs identified by different objective criteria are different and subsequently the regionalized flow corresponding to these

MPs will also be different (Seibert 1999; Wagener & Wheater, 2006). The MOSCEM-UA used in this study is a Pareto-based approach which allows simultaneous identification of the posterior distribution of parameters (provided Gaussian assumptions are valid) and the Pareto optimal parameter set. Instead of individual best MPs corresponding to each objective function mentioned above, MPs lying in the Pareto optimal front were all used for regionalization.

Description of basins and data

The study area consists of 26 basins located in different geographic and climate zones (Table 1). The number of CAs based on landscape and climate attributes were selected, and are as follows: (a) area; (b) drainage density; (c) average basin slope; (d) basin shape factor; (e) mean elevation of a catchment; (f) average topographic index; (g) average hydraulic conductivity at basin scale, calculated from soil texture; (h) average maximum root zone depth, calculated at basin scale using soil and land cover map; (i) average annual rainfall; (j) the SD of monthly precipitation; and (k) wetness index calculated as ratio of mean annual precipitation to mean annual potential evapotranspiration. Most of the basins selected in the study have a wetness index greater than 1, which implies that the basins selected are humid. Other data used in this study includes: the 90-m DEM from the Shuttle Radar Topography Mission (SRTM), soil data from the Food and Agriculture Organization (FAO), and land-use data from the International Geosphere–Biosphere Program (IGBP). The FAO and IGBP data were used to

Table 1 Description of basins and calibrated model parameters.

Country	Catchment Id	Area (km^2)	Calibrated model parameters			Country	Catchment Id	Area (km^2)	Calibrated model parameters		
			m	*To*	*Td*				*m*	*To*	*Td*
Australia[a]	145018	81	0.23	2.5	1.07	UK[d]	27034	510	0.02	4.80	5.48
	[f]204016	104	--	--	--		[f]27035	282	--	--	--
	204017	82	0.06	4.95	1.23		62001	893	0.02	6.21	0.30
	218001	93	0.09	1.04	4.99		66011	344	0.01	4.25	8.90
	302200	448	0.24	1.03	1.08	France[e]	J3024010	43	0.08	4.86	2.31
Nepal[b]	330	1980	0.13	6.08	1.00		J4124420	32	0.14	5.49	1.10
	795	1148	0.03	5.64	7.81		K0744010[f]	181	--	--	--
	390	554	0.07	3.21	1.22		J4712010	142	0.04	6.23	1.45
	[f]339.5	683	--	--	--		H2001020	98	0.04	4.20	3.46
Japan[c]	Arakawa (Yorii)	927	0.02	4.00	9.09		Y5615030	297	0.05	4.82	2.39
	[f]Ukaibashi	487	--	--	--		K0753210	371	0.04	4.48	3.10
	Torinkyo	1095	0.04	4.00	1.22		K0813020	193	0.04	4.20	1.28
UK[d]	23006	331	0.02	3.40	0.10		V3517010	25	0.04	8.72	2.28

[a]Catchments located in eastern Australia, and data were obtained from http://www.stars.net.au/tdwg/?datasets; [b]catchments located in Middle mountain physiographic region of Nepal and data were obtained from Department of Hydrology and Meteorology (DHM), Nepal; [c]catchments located in Japan, and data were obtained from Ministry of Land, Infrastructure and Transport (MLIT), Japan; [d]catchments located in UK, and data were obtained from http://www.ceh.ac.uk/data/nrfa/index.html; [e]catchments located in France, and data obtained from Model Parameter Estimation Experiment (MOPEX)-France; [f]catchments used for the validation of regionalization schemes.

determine the maximum root zone storage parameter of TOPMODEL which otherwise would have to be calibrated, and in addition, the basin average saturated hydraulic conductivity and maximum root zone depth were also calculated and subsequently used as a catchment attributes for regionalization. Among various sources of global data sets which can be used for regional modelling are: the digital soil map of the world and derived soil properties, a new global land cover classification for the year 2000 (GLC2000) produced by the European Commission Joint Research Centre (http://www-gvm.jrc.it/glc2000), and the Global 30 Arc Second Elevation data set at USGS, Global Land One-KM Base Elevation 30-sec. DEM, Hydro1K DEM.

RESULTS AND DISCUSSION

The parameters of the modified TOPMODEL were calibrated using three years of daily hydrometeorological data for all selected basins using MOSCEM-UA (Table 1). Only a few CAs were correlated with MPs at the 10% significance level. Model parameter *m* is found to be correlated with wetness index ($r = -0.48$), 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). In this study, at first, the performance of conventional schemes paired with the information obtained from the posterior distribution of parameters (as ranges of parameters) was investigated for regionalization. The ranges of MPs obtained from posterior distributions of parameters were mapped against CAs using ANN assuming that basins with similar CAs and data aspect will have similar ranges of MPs. The performance of ANN in simulating the ranges of MPs was efficient during calibration and is shown in Fig. 2(a). The loss of model performance in % due to the use of regionalized MPs (Fig. 2(b)), and standard error estimate (Table 2) of the regionalized MPs apparently reveals the improvement in regionalization when prior ranges of MPs are paired with regionalization, and in addition, marginal improvement in regionalization was observed with the average values of the regionalized MPs obtained from various structures, referred to as model mean (MM).

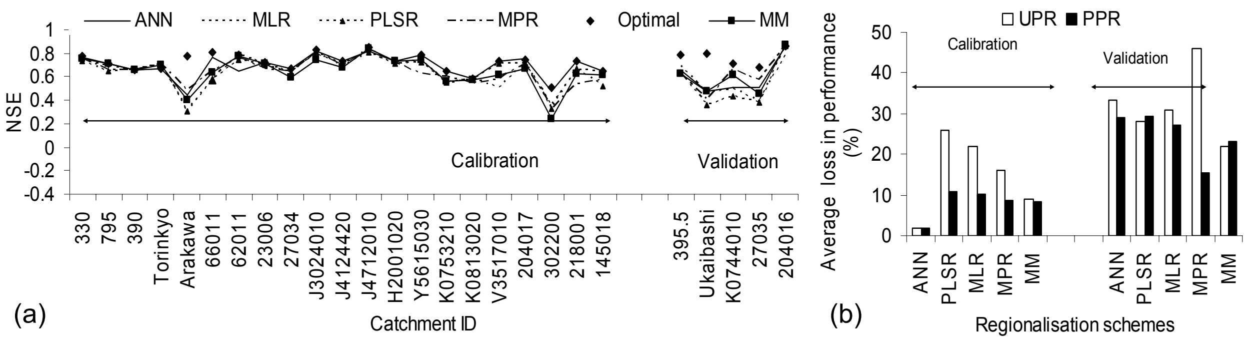

Fig. 2 Performance of regionalization schemes. (a) Performance of various schemes paired with predicted ranges of model parameters. (b) Comparison of the spatial loss in performances for schemes paired with prior ranges (PPR) and the conventional schemes unpaired with prior ranges (UPR).

Table 2 Standard error estimates for regionalized model parameters for regionalization schemes paired with prior ranges (PPR) and schemes unpaired with prior ranges (UPR).

Regionalization schemes	Standard error for the parameters estimated from regionalization schemes											
	Calibration of regional model						Validation of regional model					
	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
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

In this case study, to address the issue of the uncertainties in MPs, the set of 100 MPs generated from the posterior distribution of MPs identified by MOSCEM-UA, were used for regionalization. The posterior distribution of MPs identified for all basins revealed the variation of posterior SD of the MPs among basins, and in addition, the average width of the interval of simulated hydrograph (AWISF) expressed in % calculated from the vectors of MPs sampled from the posterior distribution (within the 90% confidence region) were significant. So, it is essential to propagate the effect of uncertainties in MPs. In this context, regionalization of vectors of MPs is a sensible idea to incorporate the effect of the uncertainties in MPs. Propagating the uncertainty in MPs via regionalization using vectors of MPs will inherently be affected by the uncertainties in the regionalization schemes, so the set of MPs estimated through the regionalization of 38 bootstrap samples of 19 basins with their corresponding CAs and optimal MPs (Table 1) were combined with the MPs estimated from M2. The use of M1 quantifies the uncertainties in conventional regionalization schemes, but it neglects the effect of the uncertainties in MPs. The combined set of MPs results in the ensemble of simulated flow for the target basin. The values of AWISF calculated from the ensemble of flows were subsequently used for the quantification of uncertainties in regionalization. A high value of AWISF was observed for one Australian basin (Catchment ID 302200) which had appreciably lower runoff coefficients and wetness index compared to the other selected basins, so it was removed from further discussion. The value of AWISF estimated by M1, M2 and M3 for all basins considered both for calibration and validation of regional models, are shown in Fig. 3 (a, b and c), and the average (of selected basins) values of AWISF are shown in Fig. 3(d). From Fig. 3, it is apparent that the magnitude of AWISF obtained for all basins varies among the regional model structures used for the propagation of uncertainties. The value of AWISF obtained from various regionalization schemes were significantly correlated at 5% for all methods (e.g. M1, M2 and M3). In addition, Fig. 3 apparently shows that the uncertainties in flow propagated by ANN was higher compared to MLR, MPR and PLSR, which could be due to the larger number of free parameters in ANN. Additionally, the uncertainties in regionalization evaluated by using M1 was high compared to M2, which is more likely due to the unequal and sparse numbers of basins selected across each region (Table 1). Comparison of the ensemble of simulated flow and model performances obtained from regionalized vectors of MPs with the same generated from the posterior distribution for the basin when presumed ungauged, is fundamental for the evaluation of the proposed methodology. In this study, the

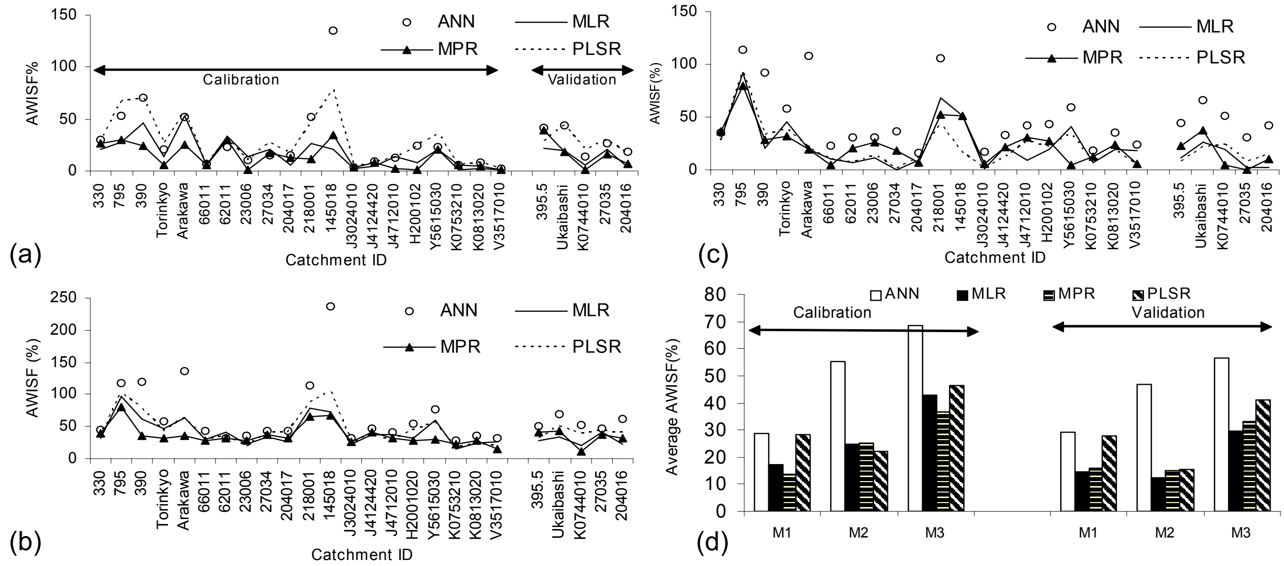

Fig. 3 The comparison of average width of the interval of simulated flow(AWISF): (a) effect of regional model parameter uncertainty (M1), (b) effect of model parameter uncertainty(M2), (c) effect of combined uncertainty, and (d) average AWISF obtained for various regionalization schemes and various methods (e.g. M1, M2 and M3).

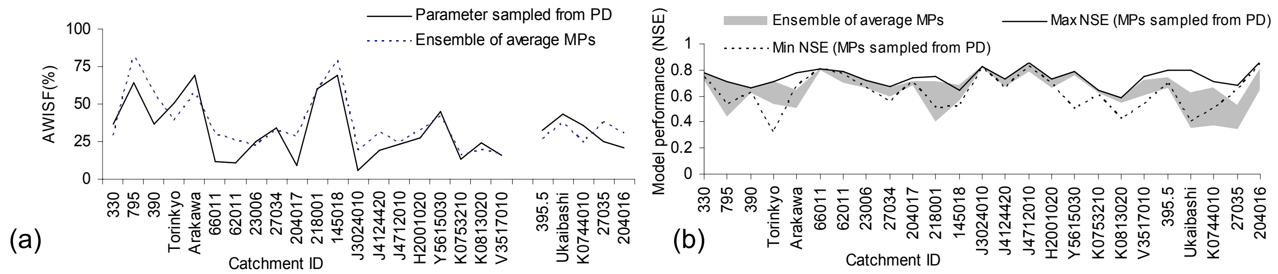

Fig. 4 Model output for model parameters (MPs) estimated from the regionalization of vectors of MPs and MPs sampled from the posterior distribution: (a) average width of the interval of simulated flow, and (b) range of model performance.

regionalized MPs obtained from MM were used for evaluation, as the average value of regionalized MPs obtained from various schemes led to better regionalization (see Fig. 1(b) and Table 2). The value of AWISF and model performances for all basins resulting from the sets of MPs sampled from the posterior distribution of parameters (referred to as parameters sampled from PD) and MM (referred to as ensemble of average MPs) are significantly similar (Fig. 4), and AWISF were also found to be significantly correlated (5%) to the posterior SD of *m* and *Td* (TOPMODEL parameters). Although in this case study, the nonparametric bootstrap methodology was used to quantify the uncertainty in regionalization by using only the optimal value of MPs, the application of the same to the each vector of MPs would be more robust approach to propagate the uncertainties from modelling system to model prediction.

CONCLUSION

It is demonstrated that pairing of the regionalization schemes with posterior distributions of parameters supplements the conventional approach with additional information and improves regionalization. Despite using a parsimonious model and posing calibration as a multi-objective problem, the effects of model parameter uncertainty on the simulated flow were significant. As conventional regionalization ignores the effect of uncertainties in MPs and regionalization schemes, sets of MPs estimated from the regionalization of vectors of MPs were combined with the sets of MPs estimated from the non-parametric bootstrap approach. The ensemble of simulated hydrographs and model performances obtained from both the proposed methodology and from the parameters sampled from the posterior distribution of MPs being significantly similar reveals the prospect of regionalizing the vectors of MPs to incorporate the effect of uncertainty in MPs while simulating flow for ungauged basins. The effect of model parameter uncertainties in regionalization were high enough not to be neglected even for a parsimonious model structure like the one used in this case study. The effect of uncertainties in MPs and regionalization schemes on simulated flow expressed as the average AWISF of all basins was 37% for basins considered for calibration and was 31% for validation, which closely followed the results obtained from the parameters sampled from posterior distribution (32% for calibration and 31% for validation). Though the ensemble of model prediction for ungauged basins encompassed the average observed flow and explained much of the variability of the observed flow, the ensemble of simulated flow could not encompass all the values of the observed time series, which is more likely due to the fact that uncertainties in the model parameters and regionalization can not account for all the uncertainties in the simulation.

REFERENCES

Beven, K. J., Lamb, R., Quinn, P., Romanowicz, R. & Freer, J. (1995) TOPMODEL. In: *Computer Models of Watershed Hydrology* (ed. by V. P. Singh), 627–668. Water Resources Publications, Highlands Ranch, Colorado, USA.

Beven, K. J. & Binley, A. M. (1992) The future of distributed models: model calibration and uncertainty prediction. *Hydrol. Processes* **6**, 279–298.

Fernandez, W., Vogel, R. M. & Sankarasubramanian, A.(2000) Regional calibration of a watershed model. *Hydrol. Sci. J.* **45**(5), 689–707.

Griet, H., Bart, M. & Jan, F. (2006) Regionalisation of the parameters of a hydrological model: comparison of linear regression models with artificial nets. *J. Hydrol.* **319**, 245–265.

Gupta, H. V., Sorooshian, S. & Yapo, P. O. (1998) Toward improved calibration of hydrologic models: multiple and noncommensurable measures of information. *Water Resour. Res.* **34**, 751–763.

Kuczera, G. & Parent, E. (1998) Monte Carlo assessment of parameter uncertainty in conceptual catchments models: the metropolis algorithm. *J. Hydrol.* **211**, 69–85.

Lamb, R. (2000) An approach to the calibration of a conceptual rainfall–runoff model for flood frequency estimation by continuous simulation. *Water Resour. Res.* **35**, 3103–3114.

Lamb, R. & Kay, A. L. (2004) Confidence intervals for a spatially generalized, continuous simulation flood frequency model for Great Britain. *Water Resour. Res.* **40**, W07501, doi:10.1029/2003WR002428.

Merz, R. & Blöschl, G. (2004) Regionalisation of catchment model parameters. *J. Hydrol.* **287**, 95–123.

Seibert, J. (1999) Regionalisation of parameters for conceptual rainfall–runoff model. *Agric. Forest Met.* **98-99**, 279–293.

Sivapalan, M., Beven, K. J. & Wood, E F. (1987) On hydrological similarity: 2. A scaled model of storm runoff production. *Water Resour. Res.* **23**(12), 2266–2278.

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. & Zehe, E. (2003) IAHS Decade on Predictions in Ungauged Basins (PUB), 2003–2012: shaping an exciting future for the hydrological sciences. *Hydrol. Sci. J.* **48**(6), 857–880.

Tobias, R. (1995) An introduction to partial least squares regression. In: *Proceedings of the Twentieth Annual SAS Users Group International Conference* (Cary, NC), 1250–1257. SAS Institute Inc., USA.

Vrugt, J. A., Gupta, H. V., Bastidas, L. A., Bouten, W. & Sorooshian, S. (2003) Effective and efficient algorithm for multiobjective optimization of hydrologic models. *Water Resour. Res*. **39**(8), 1214, doi: 10.1029/2002WR001746.

Wagener, T. & Wheater, H. S. (2006) Parameter estimation and regionalization for continuous rainfall–runoff models including uncertainty. *J. Hydrol.* **320**, 132–154

Wagener, T., Wheater, H. S. & Gupta, H. V. (2004) *Rainfall–Runoff Modelling in Gauged and Ungauged Catchments*. Imperial College Press, London, UK.