Journal of Hydroscience and Hydraulic Engineering
Vol. 27, No.2 November, 2009, 1-15

REGIONALISATION OF JOINT DISTRIBUTION OF MODEL PARAMETERS: PREDICTION ON UNGAUGED BASINS

By

Satish Bastola

Interdisciplinary Graduate School of Medicine and Engineering, University of Yamanashi, Kofu, Yamanashi, Japan

Hiroshi Ishidaira

Interdisciplinary Graduate School of Medicine and Engineering, University of Yamanashi, Kofu, Yamanashi, Japan

and

Kuniyoshi Takeuchi

The International Centre for Water Hazard and Risk Management, Tsukuba, Ibaraki, Japan

SYNOPSIS

In the case study presented in this work, the calibrated parameters of a Conceptual Rainfall Runoff (CRR) model could not be uniquely identified. Moreover, the covariance and variance of model parameters varied among basins. The uncertainty in the calibrated model parameters have been recognized as a concern for successful model parameter regionalisation. Therefore, regionalisation should include assessment of model parameter uncertainty. For this reason, we proposed a regionalisation scheme which, instead of regionalizing the single best value of model parameters, regionalizes their joint probability distribution. A functional relationship between catchment attributes and characteristic parameters of joint distribution are developed. The uncertainty in model prediction is quantified using multivariate stochastic simulation technique, whereby a large number of model parameter vectors are sampled from their estimated joint distribution that lie within 90% confidence region and fed into the CRR model. Regionalisation of joint distribution addresses the effect of model parameter uncertainty in the result of regionalisation by taking into account the covariance between model parameters which are generally ignored in approximate analytic techniques based on Taylor series expansion. In the case study presented in this paper, the predictive uncertainties quantified from the proposed method closely followed the prediction uncertainties quantified from the calibrated joint probability distribution of model parameters.

INTRODUCTION

Conceptual Rainfall Runoff (CRR) models are popular tools for modeling flow at gauged site. Though some parameters of CRR model have a physical basis, they are effective values at the catchments scale and are hardly measurable in the field. Therefore, they are obtained by calibrating the model against observed data. The necessity of calibrating the parameters has led to linking of model parameters of CRR models with catchment attributes in order to model flow at ungauged site. A number of methods for transferring or regionalizing the parameters have been explored in the past (Merz and Blöschl (1); Wagener et al. (2)). Among various methods, the use of multiple regressions to relate model parameters (θ) to measurable Catchments Attributes (CAs) is widely used in modeling ungauged basins (Wagener et al. (2); Heuvelmans et al. (3)). This typically results in a set of multiple regression equations which estimates one model parameter at a time rather than equations that jointly estimate model parameters. Thus any information about covariances and variances of the model parameter is usually lost.

The uncertainty in the calibrated model parameters have been recognized as a concern for successful model parameter regionalisation. Merz and Blöschl (1) addressed the issue of the uncertainties in θ by making a comparison of model parameter for two sub periods. Sequential Regression (SR) addresses this issue by combining model calibration with regionalisation. Though SR relaxes the assumption that the model parameters are independent, and helps in improving the identifiability of model parameter, it is difficult to apply without introducing bias into the regionalized parameters (Wagener et al. (2)). Moreover, regional calibration method (Bastola et al. (4); Parajka et al. (5)) have appeared more recently to minimize the effect of the model parameter uncertainty in the result of regionalisation by skipping the process of model calibration. Regional calibration approach estimates the value of model parameter at all sites simultaneously in an attempt to achieve the best regional relationship, thereby skipping the direct calibration of local model. However, regional calibration method failed to identify the unique regional relationship, thereby inducing considerable uncertainties in model prediction (Bastola et al. (4)). In a similar effort, Wagener et al. (2) used weighted regression to reduce the influence of parameters of a gauged catchment that were poorly identified during calibration, but this does not introduce weights to allow for the model's local performance. Furthermore, the determination of identifiability is dependent on the sample size. In an attempt to include model parameter uncertainty in the result of regionalisation, instead of regionalizing a single best value of model parameters Bastola et al. (4) attempted to regionalize a number of plausible values of model parameters. As the uncertainty in regional parameters can not be avoided, they incorporated non-parametric bootstrap sampling to account for the uncertainty associated with parameters of regional models. Though this approach recognizes the existence of a number of plausible values of model parameters, and addresses the uncertainty in the estimated regional parameters, they, however, assumed that model parameters are independent to each other thereby ignoring the information on parameter covariance. Moreover, enormous numbers of predictive models were required as they rely on the selection of a number of plausible values of model parameters and number of bootstrap samples. All the studies mentioned above attempted to address the issue of model parameter uncertainty but they did not explicitly incorporate the joint probability distribution of model parameters which provides the information on the model parameter uncertainty.

This paper aims at developing an approach for the regionalisation of parameters of CRR models by considering the uncertainty in the calibrated model parameters and their interdependency. The methodology is demonstrated through a case study that includes a number of small to medium size basin (~25 to 1900 km^2) located in different geographic and climatic regions. The motivation for this work lies in the need to address various sources of uncertainty which have drawn much attention in recent years and would otherwise hinder successful regionalisation.

METHODOLOGY

In most CRR models, the model parameters are typically strongly correlated leading to the existence of multiple optima. Thus, such information should be included in deriving a regionalisation procedure. To accomplish this, it is assumed that the joint distribution of model parameters can be approximated by multivariate normal distribution (i.e. the posterior mean and entries of covariance matrix) and they vary among basin depending on their attributes and data aspects. This consequently allows us to seek for a functional relationship between the characteristic parameters of joint probability distribution of model parameters and basin attributes. The methodology is divided into two parts: estimation of joint distribution of model parameters, and regionalisation of joint distribution of parameters.

As the parameters of hydrological model can not be uniquely defined and are uncertain, the goal of model calibration should include the estimation of Posterior Probability Distribution of model parameters (θ) conditioned on observation (D) which is hereafter referred as PPD interchangeably. If one assumes that there is a probability distribution of outcomes based on knowledge, then one can discuss uncertainty. Bayesian statistics describe uncertainty using posterior probability distribution which represents what is known about the parameters given the prior knowledge and data. It combines two sources of information for learning about unknown parameters: the prior probability that represents the prior (subjective) belief about the values of parameters, and the likelihood function that reflects how the uncertainty enters and propagates through the system. Bayesian statistical inferences provide an ideal means of assessing parameter uncertainty. However, such approaches are hindered by difficulties in summarizing and exploring the posterior distribution. Recent advances in Markov Chain Monte Carlo (MCMC) methods have largely overcome these difficulties by providing an alternate means of sampling from a posterior distribution (see Kuczera and Parent (6)). Marshall et al. (7) compared four MCMC sampling algorithms in the context of rainfall-runoff modeling and found that adaptive metropolis algorithm is superior in many respects and can offer a relatively simple basis for assessing parameter uncertainty.

The metropolis algorithm produces a Markov chain sequence of samples that constitute a random walk in the parameter space. Each iteration of the algorithm proceeds by first generating candidate parameter samples using a suitable arbitrary probability distribution referred to as the jump distribution. These candidates are then either accepted or rejected using criterion which ensures the algorithm is sampling from the posterior distribution when the Markov chain has become stationary. Vrugt et al. (8) demonstrated that an MCMC sampler known as the Shuffled Complex Evolution Metropolis algorithm (SCEM-UA) is efficient and well-suited to infer the posterior distribution of hydrologic model parameters. It is an adaptive sampler, where the covariance of the sampling distribution is periodically updated in each complex during the evolution to the posterior target distribution using information from the sampling history induced in the transitions of the generated sequences. The following is a brief outline of the method used for generation of posterior probability distribution (see Vurgt et al.(9) for detail).

1. Select the population size s and number of complex q for model calibration.
2. Generate s set of parameters from the prior probability distribution and compute the multi-objective vector f for each individual of the population and the rank the population based on the fitness assignment strategy proposed by Vurgt et al. (9). Sort the s individual in order of decreasing fitness value and store the parameters and sorted fitness value in an array D.
3. Partition the population in array D into several complexes Cx (e.g., Cx^1, Cx^2.....Cx^q).
4. Compute the parameter covariance matrix for each complex C. Subsequently a new candidate points in k^{th} complexes Cx^k is generated according to $\theta^{t+1}=N(\theta^t, \Sigma^k)$ where θ^t is the current draw of the s.
5. Begin the metropolis algorithm
   a) Compute the acceptance ratio $r = (f^{t+1}/f^t)$, where f^t and f^{t+1} are the fitness associated with current draw and offspring parameters respectively
   b) Randomly sample a uniform label Z over the interval 0 to 1.
   c) Accept the new sampled points θ^{t+1} if $Z \leq r$
   d) Replace the worst point of C^k with θ^{t+1}
6. Unpack all complexes Cx back into D.
7. Check the convergence statistics, if convergence statistics are satisfied stop, else go to 3.

The assessment of parameter uncertainty using the Markov chain based approach has been largely used to quantify the uncertainty in model parameters at gauged site. However, there application to modeling flow in un-gauged basin is difficult due to the unavailability of observed data to construct the posterior probability distribution of model parameters. Therefore, two presumptions are made in this study to extrapolate the joint distribution of model parameters of a specified hydrological model from gauged to ungauged basin. Firstly, we presumed that the model parameter uncertainty depends on basin characteristics, input data and the structure of hydrological model, and secondly we presumed that for parsimonious model structures, the joint distribution can be sufficiently approximated by multivariate normal distribution which is the most widely used distribution for dependent continuous variables. A close examination of the work by Uhlenbrook et al. (10) and Seibert (11) reveals that the model parameter uncertainty depends on basin characteristics, input data and the structure of the hydrological model used. Therefore, formulation of relation between the characteristics parameters of joint distribution of model parameters and basin attributes is presumed plausible to estimate the joint distribution of model parameters at un-gauged site.

The outline of the regionalisation joint distribution of model parameters is described as follows (Fig. 1):

1. Select a number of gauged basin and estimate vectors of relevant catchment attribute.
2. Calibrate the parameter of the selected hydrological model for all basins using the Markov Chain Monte Carlo method. Subsequently, estimate the joint distribution of hydrologic model parameters for all basins assuming that it can be sufficiently approximated by multivariate normal distribution.
3. Identify the posterior mean vector μ, and the set of pair wise covariances Cov (X_i, X_j), including the variance Var (X_i) from the estimated joint distribution of model parameters.
4. Identify the functional relationship between the characteristic parameters of joint probability distribution of model parameters and catchment attributes.
5. Reconstruct the joint distribution of model parameters at ungauged basin from the functional relationship derived in step 4.

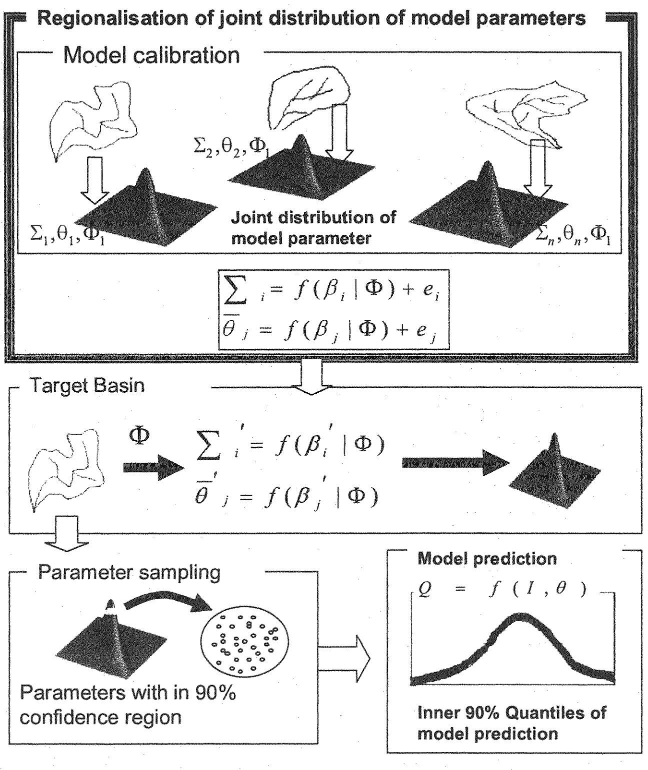

Fig.1 Schematic of regionalisation of joint probability distribution of model parameters for modeling flow in ungauged basin where Σ is the covariance matrix of the joint distribution of model parameters approximated by multivariate normal, $\bar{\theta}$ is the mean model parameter vector, Φ is the vector of catchment attributes, f(.) is the hydrological model, Q is the stream flow, I is the model input, e_i is the error.

The multivariate stochastic simulation based on multivariate normal distribution (Scheuer and Stoller (12)) is adopted to quantify the predictive uncertainties in simulated flow. The Scheuer and Stoller algorithm presupposes estimates of 1) the covariance matrix Σ containing the variance and covariance of the normally distributed random variables, and 2) the vector of their means μ. The algorithm is based on the theorem that if x is an n-component vector of variables with a multivariate normal distribution (i.e., x is distributed as N(μ, Σ) where μ is the vector of mean values and Σ is the covariance matrix), then a vector of variates X that corresponds to the multivariate normal distribution of x is given by Eq. 1.

$$X = CZ + \mu \tag{1}$$

where C is the Cholesky Factor of Σ and Z is an n-component vector of independent standard normal variates. The matrix C is the unique lower-triangular matrix such that it satisfies Eq. 2. Thus, finding C (Cholesky factorization) is equivalent to finding the "square root" of the covariance matrix.

$$CC' = \sum \tag{2}$$

Thus, once the Cholesky matrix C has been derived from the covariance matrix Σ, repeated application of Eq. 2 to successive random vectors Z yields a large sample of independent pseudorandom data vectors, each vector corresponding to an independent observation of the variables from the specified multivariate normal distribution. From larger number of parameter sets, only those falling within a pre-specified confidence region i.e. the parameters set that satisfies Eq.3 is selected.

$$(\theta - \mu)^T \sum{}^{-1} (\theta - \mu) < \frac{p(m-1)(m+1)}{m(m-p)} * F(\alpha, p, m-p) \tag{3}$$

where $F(\alpha, p, m\text{-}p)$ is the $1-\alpha$ percentile of the F distribution with p and $m\text{-}p$ degrees of freedom. This consequently results in the ensemble of simulated flow. Subsequently the distribution of model output are examined in order to provide inference on the probability that various subset of outcomes will occur.

This approach discussed above for regionalisation of parameters of hydrological model under parameter uncertainty relaxes the assumption that parameters are independent and, thereby preserves the information on the correlation among model parameters. Furthermore, the number of predictive models always remains equal to the number of characteristic parameters of joint distribution of model parameters (e.g., the predictive models required for three parameter hydrological models is nine when the joint distribution is approximated as multivariate normal distribution).

CASE STUDIES

Study area and data considerations

The 26 small to medium size basins situated in different geographic and climate zones (Table 1) were selected. The 21 catchments were considered for calibration of regional model and 5 catchments each located in 5 different countries (e.g., 395.5, Ukaibashi, K0744010, 27035 and 204016) were presumed ungauged for the appraisal of the performance of regionalisation schemes. Though the inclusion of larger number of basin are desirable, general problems with model parameter regionalisation studies are: availability of only limited data, difficulty in selecting a more general types of hydrological models that can be applied to a large number of basin characterized with different dominant hydrological processes. Therefore, we only focused on small to medium size humid basins. The subset of hydro-metrological data used

Table 1 Description of selected basins

SN	Rivers system	Catchment Id	Area Km^2	Mean elevation (m)	Mean annual Rainfall (mm)	Mean annual Runoff (cumecs)	Average topographic Index
1	West rapti	330[a]	1980	1757	1706	1336	6.26
2	Kankai	795[a]	1148	1229	2366	1833	9.47
3	Tinau	390[a]	554	909	2213	1950	6.44
4	Jhimruk	395.5[a]	683	1654	1716	1031	6.33
5	Fuji	Torinkyo[b]	1020	943	1050	823	7.52
6	Arakawa	Arakawa[b]	953	837	1115	616	6.39
7	Ukaibashi	Ukaibashi[b]	487	1102	1165	855	7.1
8	Conwy	66011[c]	344	345	2055	1704	7.8
9	Teifi	62001[c]	893.6	211	1382	1011	7.9
10	South Tyne	23006[c]	331.9	438	1331	1003	8.19
11	Ure	27034[c]	510.2	366	1342	979	8.25
12	Aire	27035[c]	282.3	233	1153	713	8.37
13	Bielsdown creek	204017[d]	82	774	1884	1202	7.86
14	Swan	302200[d]	448	342	763	281	7.81
15	Tuross	218001[d]	93	1060	946	368	7.56
16	Burnett ck	145018[d]	83	556	1411	218	7.13
17	little Murray	204016[d]	104	910	1703	1041	7.07
18	Le Guillec à Trézilidé La	J3024010[e]	43	85	1014	491	8.68
19	La Rivière de Pont-l'Ab bé à Plonéour-Lanvern	J4124420[e]	32.1	79	1236	522	8.39
20	L'Éllé au Faouët	J4712010[e]	142	201	1192	611	8.32
21	L'Yonne à Corancy	H2001020[e]	98	601	1299	930	7.61
22	Le Loup à Villeneuve-Loubet	Y5615030[e]	279	835	1159	505	7.21
23	Le Lignon du Forez à Boën	K0753210[e]	371	869	1012	485	7.5
24	L'Aix à Saint-Germain-Laval	K0813020[e]	193	751	988	493	7.27
25	Le Ternay à Savas	V3517010[e]	25.5	906	867	429	7.82
26	L'Anzon à Débats-Rivière-d'Orpra	K0744010[e]	181	758	980	443	7.47

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

in study has also been used in other parameterization and regionalisation studies (e.g., Xie and Fei (13); Shrestha et al.(14); Reichl et al.(15)).

To link the calibrated values of model parameters with easily measurable basin attributes, a number of landscape attributes were selected in this study, which are as follows: 1) area (A) (km^2), 2) average topographic index (Ave. Topographic Index or ATI), 3) Shape Factor (SF) (km/km^2), calculated as the square of the length of the longest flow-path divided by the watershed area; 4) basin average saturated transmissivity (Ave. Transmissivity or AT) (cm/h), 5) basin average maximum root zone depth (Ave. SRMAX) (m), 6) Mean Elevation (ME) (m), 7) average Basin Slope (BS) (%), 8) Drainage Density (DD) (km/km^2). Moreover, as the basins selected in the study are situated in different climate zones, a number of climate attributes were also considered such as: (1) Annual Average Rainfall (AAR) (mm), (2) Variance of monthly Rainfall (VR), and (3) Wetness Index (WI), calculated as the ratio of average rainfall to potential evaporation. In order to extract the landscape attributes for all basins, the land use data obtained from International Geosphere-Biosphere Programme (IGBP), soil data obtained from Food and Agricultural Organization (FAO), and The Shuttle Radar Topography Mission (SRTM) data that cover the entire globe with a 3-arc second (approx. 90m) digital elevation model were used.

Hydrological model

The TOPMODEL (Beven et al. (16)), which is a variable contributing area physically-conceived semi-distributed hydrological model, was selected. Based on a simple theory of hydrological similarity of points in a catchment, TOPMODEL makes distributed prediction of catchment response. These points of hydrological similarity are identified by an index that is derived from catchment topography. Flow is separated into surface runoff generated by rainfall on saturated contributing areas and subsurface downhill flow. The four basic assumptions that TOPMODEL uses to relate local down slope flow from a point to discharge at the catchment outlet are: 1) the dynamics of the saturated zone are approximated by successive steady state representations, 2) the recharge entering the water table is spatially homogeneous, 3) the effective hydraulic gradient of the saturated zone is approximated by the local surface topographic gradient, 4) the distribution of down slope transmissivity *To* with depth is a function of storage deficit. The parameters requiring calibration are the decay parameter (*m*) (m), lateral transmissivity (*To*) (m^2/h), delay factor (*Td*) (h/m), maximum root zone storage (*Srmax*) (m), and parameters related to channel routing. As we focused on modeling flow in small to medium size basins at daily time scale, the routing of flow is not implemented. Moreover, the *Srmax* parameter was estimated from the root zone depth and soil properties (Beven et al. (16)). The inputs of the model are the land cover map, digital elevation model, soil map, daily precipitation, and potential evaporation data. For model calibration only the daily stream flow measured at a basin outlet was used.

Model calibration

As multi-objective calibration eases in retrieving more information from the observed data and provides insight into parameter uncertainty and limitation of model structure, the Multi-Objective Shuffle Complex Evolutionary Metropolis algorithm (MOSCEM-UA) (Vrugt et al. (9)) was used. The primary strength of MOSCEM-UA is the estimation of the PPD of CRR model parameters and is best suited for calibration of models that have small number of model parameters (Tang et al. (17)).

For calibration the following goodness of fit criteria were used: 1) The Nash Sutcliffe efficiency (NSE), 2) NSE for the transformed flow to consider the heteroscedastic variance in flow, here the flow is transformed explicitly by using, $z = [(y+1)^{\lambda} - 1]/\lambda$, (Box and Tiao, (18)), where λ is the transformation parameter and is selected to be 0.3, z is the transformed stream flow, and y is the observed stream flow , 3) NSE for Low flow and 4) NSE for peak flow. The objective criteria were calculated using observed and simulated daily stream flow values at the basin outlets. The peak flow and low flow are defined as the periods where the observed discharge is above or below the mean flow level, respectively. The result of model calibration i.e. the most probable value of model parameter, and Coefficient of Variation (CV) of model parameters, is shown in Table 2. The coefficient of variation (CV) for model parameters is calculated as the ratio of the standard deviation to the mean. It represents a dimensionless measure of parameter uncertainty. A coefficient of variation of model parameters is calculated from the ~1000 set of Pareto optimal model parameters sets sampled from their joint probability distribution. The variation in the CV of model parameters indicates that the degree of uncertainties in model parameters varies from basin to basin depending on the basin attributes and data aspect. The model parameter *m* and *To* were found comparatively sensitive than *Td*. Furthermore, the correlation between the model parameters also varied among basins. Apart from the basin attributes, the parameter uncertainty also depends on the quality and quantity of data selected for model calibration. The required length (quantity) of the time series depends, among other things, on the complexity of the model structure used and the information content off the data. It might range from three years for a simple structure (Sefton and Howarth (19)), to up to a decade for more complex model structures (Yapo et al. (20)). However, in principle the data set should always be long enough to avoid the problem of the parameters being representative only of a particular climate sequence. With increase in the length of calibration data, the

Table 2 Most probable parameters and coefficient of variation of model parameters

Basin ID	Posterior mean			Coefficient of variation		
	m(m)	To (m^2/h)	Td (h/m)	m	To	Td
1	0.115	10	0.4	0.059	0.522	0.818
2	0.05	6	0.5	0.380	0.082	0.421
3	0.09	5	1.5	0.171	0.065	1.004
4	0.11	5	1	0.238	0.087	0.844
5	0.027	3.75	0.75	0.178	0.080	0.759
6	0.013	4.2	0.5	0.043	0.016	0.698
7	0.027	6	2.5	0.095	0.044	0.710
8	0.015	3.5	2	0.089	0.038	0.674
9	0.011	4.85	4	0.108	0.042	0.646
10	0.068	5.2	1.5	0.025	0.032	0.642
11	0.275	1.25	2.5	0.293	0.297	1.404
12	0.03	2.25	1	0.417	0.199	0.743
13	0.275	1.25	1.5	0.210	0.333	1.444
14	0.095	4.8	1	0.056	0.008	0.677
15	0.15	5.5	1.2	0.149	0.041	0.840
16	0.045	6.2	1.5	0.163	0.034	0.691
17	0.045	4.1	2.5	0.193	0.037	0.648
18	0.04	4.8	1.6	0.150	0.057	0.704
19	0.037	4.7	0.5	0.056	0.023	0.681
20	0.03	4.5	1.8	0.144	0.133	0.649
21	0.038	6.25	1	0.078	0.070	0.640

standard error of the estimate of parameter decreases (Gupta and Sorooshian (21)), but this improvement becomes smaller after nearly 1000 data points. The same paper suggested that two to three years of calibration data should be sufficient, provided that the data are of the right kind. In this study, we selected three years of hydrological data for model calibration. For an Australian basin where the hydrological variability is higher compared to other basins, we subjectively selected 3 years time series of data with the aim to include as much variety in the hydrological regimes as possible. Furthermore, the impact of input uncertainty and the presence of errors in the stream flow data were neglected. It cannot be assumed that instrumentation and observational practices will be consistent across the countries that this study covers. This may cause the same problem as structural errors because data biases are catchment-specific and will be compensated for, in a catchment-specific manner, by parameter calibration. The explicit representation of such uncertainties is beyond the scope of this study.

Joint probability distribution of model parameters

The posterior probability distribution can be entirely characterized by its mean, a set of pair wise covariance and variances if it can be sufficiently approximated by multivariate normal distribution. For the structure of model use in this study, the number of characteristic parameters of joint probability distribution to be estimated via regional link function is nine i.e., three posterior mean and six entries of covariance matrix. The covariance matrix for the parameters of TOPMODEL used in this study can be expressed as follows (Eq.4):

$$\sum_i = \begin{bmatrix} VAR(m)_i & COV(m,T_0)_i & COV(m,T_d)_i \\ COV(m,T_0)_i & VAR(T_0)_i & COV(T_0,T_d)_i \\ COV(m,T_d)_i & COV(T_0,T_d)_i & VAR(T_d)_i \end{bmatrix} \quad (4)$$

where $\sum_i$ is the covariance matrix for the ith basin, $COV(m,To)_i$, is the covariance between m and To (model parameter) for ith basin, and $Var(m)_i$ is the variance of the model parameter m for ith basin.

Regionalisation procedure

Assuming that the combinations of attributes are able to capture the exclusivity of each catchment, the method based on the regional link function (i.e. $\theta = H(\beta \mid \Phi) + e$, where θ, β and Φ are dependent variables (e.g. model parameters and the entries of covariance matrix of θ), regional parameters and catchment attributes respectively, H (.) denotes the functional relation between θ and β, and e is the error term) was used to transfer information from gauged to ungauged basin. The linear correlation between the characteristics parameters of probability distribution of model parameters and catchment attributes (see Fig.2) apparently shows the potential for developing the statistical relationship between the above mentioned dependent and independent variable (CAs). The correlation coefficient calculated between the posterior mean value of model parameters and the basin attributes reveals that basin with a higher wetness index and receiving higher average annual rainfall were characterized with lower values of model parameter *m* (the physical interpretation of the decay parameter *m* is that it controls the effective depth of the catchment soil profile). It is possible that the depth to water table is comparatively greater in basins characterized with lower wetness index. The model parameter *To* which defines the transmissivity of the catchment soil profile when saturated to the surface was found to be higher in the basin with higher mean basin elevation, steeper slope, larger drainage area and higher standard deviation of monthly rainfall. The parameter *Td* (defines the time constant for the vertical flux) was observed lower in the basin characterized with steep slope, high elevation, and high variation in monthly rainfall. Though *Td* has no direct physical meaning it has been interpreted as inverse of *To* which is also observed here in this study. Due to uncertainty in the calibrated model parameters, it is difficult to interpret the entire observed relationship/correlation solely based on physical reasoning.

Some of the predictors are strongly correlated to each other. Therefore, we combined various independent variables and selected only those combined variables for which the Variation Inflation Factors (VIF) is less than 10. The VIF is an index which measures how much the variance of a coefficient (square of the standard deviation) increased because of co linearity. The square root of VIF reflects how much the standard error is, compared with what it would be if that variable

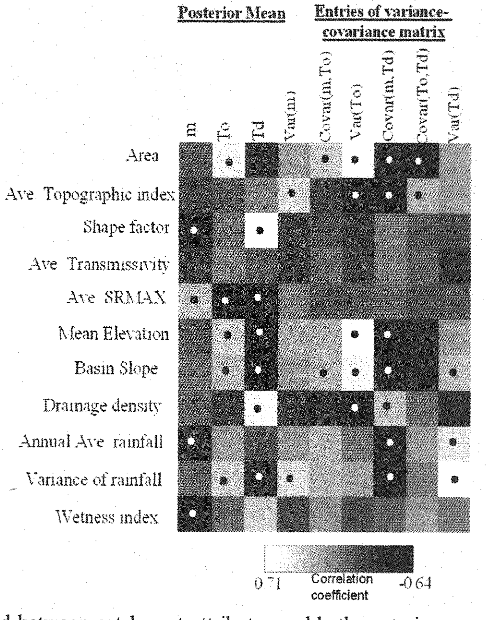

Fig.2 Linear correlation calculated between catchment attributes and both posterior mean value of model parameters and entries of covariance matrix of the posterior distribution of MPs approximated by multivariate normal distribution. The attributes that are significantly correlated with model parameters are marked with black and white marker.

were uncorrelated with the other independent variables in the equation (see Allen, Theodore T. (22)).

In this study, we used second order polynomial form of the equation to regionalize the value of posterior mean. For each dependent variable the following set of independent variables were selected: a) Ave. *SRMAX/(KS*ATI), SF*KT, WI* and *AAR* were selected as input to estimate the model parameter *m,* b) *KS, AAR/VR, ATI*√Area* and *AAR*SF* were used as independent variable for parameter *To,* and c) *DD/ATI, BS/KT* and *VR* were used as independent variable for model parameter *Td.* The close observation of the covariance between model parameters for all basins indicates that the interaction between model parameters varied among basins. The plot between the combinations of selected basins attributes and coefficient of covariance matrix is shown in Fig. 3. To regionalize the coefficient of covariance matrix, we employed an artificial neural network. Artificial Neural Networks (ANNs) are more flexible model structures that can easily account for non-linearities. For the ANN-based schemes, the popular feed forward neural network consisting of three layers: an input layer, one hidden layer and output layer. We assume *m1* input nodes (number of independent variable), *n* hidden nodes, and single output nodes. Subsequently, a total of six neural networks were trained in order to regionalize the joint distribution of model parameters. For each dependent variable the following set of independent variables are selected as input to the corresponding neural network: a) variables *WI/ATI, Area/AAR* and Ave. *SRMAX* were selected as input to estimate the variance of model parameter *m,* b) variables *Area*BS, AAR/ATI, VR/SF* and *ME*WI/DD* were selected to estimate *COV(m,To)*, c) variables *AAR*, Ave. *SRMAX* and *ATI* were selected as input to estimate the variance of model parameter *To*, d) variables *AAR, DD/ATI* and *SF*ATI/WI* were selected as input to estimate *COV(m,Td)*, e) variables *BS/WI, ATI* and *Area*ME* were selected as input to estimate *COV(To,Td)*, and f) variables *AAR*Area, VR*ATI, DD/ATI* and *BS/AT* were selected as input to estimate variance of model parameter *Td.*

The use of regional link function typically results in a number of regional estimation equations that estimate the posterior mean and entries of covariance matrix from easily measurable catchment attributes. The estimated joint probability distribution subsequently allows estimating the predictive uncertainty.

RESULTS AND DISCUSSIONS

The parameter of the TOPMODEL was calibrated using 3 years of daily hydro-meteorological data utilizing MOSCEM-UA. As the calibrated parameters could not be uniquely identified, the posterior means of model parameters were regionalized simultaneously via a regional link function that links the posterior means to catchment attributes. The simulated flow corresponding to posterior mean values accounted for much of the variability in the observed runoff in the basins considered for regionalisation. The performances of model evaluated in terms of NSE for posterior mean value of model parameters varied from 0.52 to 0.86 (see Fig. 4). A comparison between the estimated and the calibrated value of model parameters can be made to evaluate the performance of regionalisation schemes, but the accuracy of the flow simulated with regionalized parameter is a more relevant indicator. The difference between the calibrated and the estimated model performance is here after referred to as spatial loss in model performance. The predictive model performance of parameters estimated from regional link function for all basins is shown in Fig. 4. Both spatial losses in the model performance and the spatial proximity of the estimated model parameters were used to evaluate the predictive performance of regionalisation schemes. The spatial loss in the model performance was 7% for basins considered for calibration and 12% for basins presumed ungauged. The mean value of three model parameters that are calibrated are 0.079, 4.614 and 1.32 and the estimated through regional models are 0.0759, 4.17 and 1.41. Standard error in the estimated parameters is 0.009, 0.207 and 0.77 respectively for the model parameter *m*, *To* and *Td.*

The joint probability distribution of model parameters provides insights into the uncertainty and the interaction among model parameters. The extent of interaction and uncertainty in model parameter vary from basin to basin (Table 2) depending upon CAs and data aspects. The model calibration result indicates that uncertainties in the model parameter *To* was higher in larger basins, characterized with higher elevation, steeper slope and lower wetness index. Besides *To* and *m*, CV of *Td* was also significantly correlated with annual average rainfall and SD of monthly rainfall.

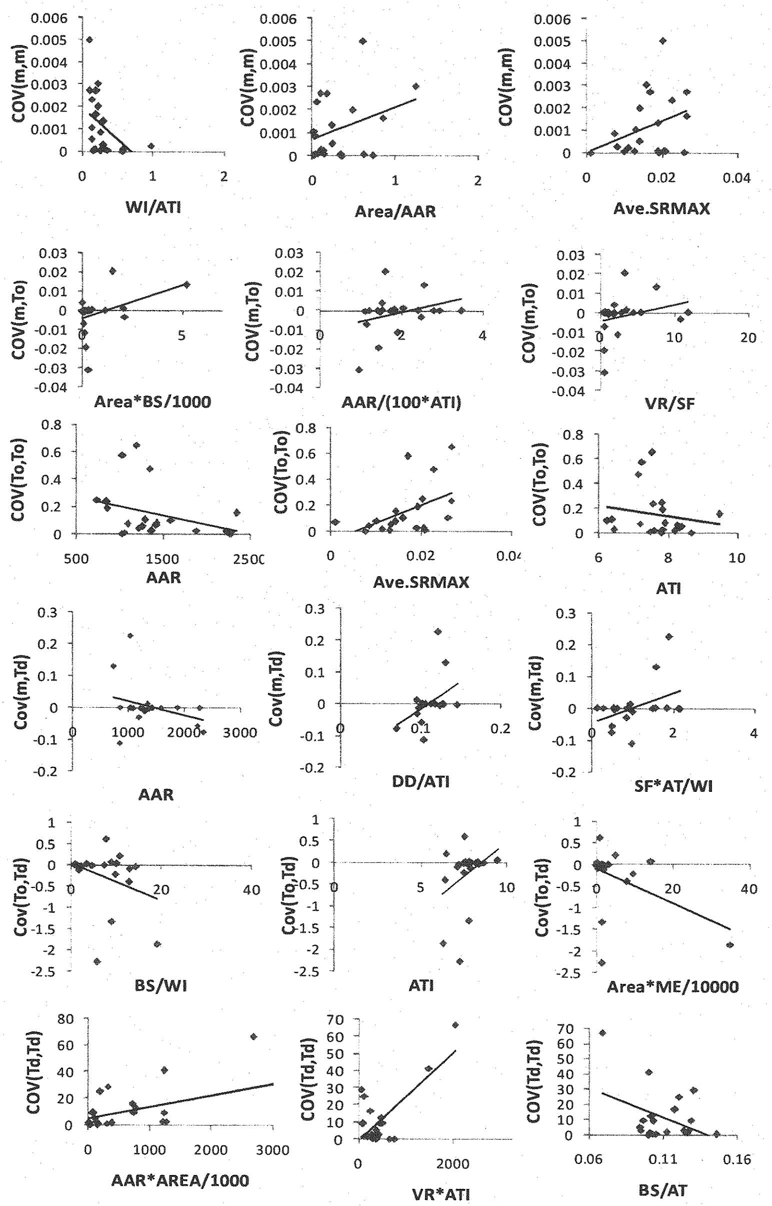

Fig.3 The scatter plot showing the relationship between the derived basin attributes and the covariance between model parameters.

To determine the uncertainty at each time step, the predicted output from the plausible value of parameters sampled (lying within 90% confidence region) from their joint distribution were equally weighted and ranked to form a cumulative distribution of the output variable. Subsequently, prediction ranges at each time step were determined corresponding to the chosen quantiles. In this study the 5, 95% quantiles were used to define the prediction bounds. From the prediction range, Average Width of the Interval of the Simulated Flow (AWISF) was calculated using Eq. 5.

$$AWISF = \sum_{i=1}^{n} (Q_{i,\alpha} - Q_{i,1-\alpha}) / \overline{Q} \qquad (5)$$

The value of AWISF (%) resulted from using numbers of plausible value of model parameter was adopted here as a measure to quantify the prediction uncertainties associated with parameter uncertainty. The value of AWISF for all basins estimated using model parameters sampled from its joint distribution conditioned on observation (referred to as calibrated in Fig. 5) is significantly (at 5% significance level) correlated with CV of model parameters. A close look into the prediction uncertainties and CAs reveals that higher prediction uncertainties are found to be associated with basins characterized by the following attributes: larger area, higher elevation, steeper slope, lower wetness index, and higher variance in monthly rainfall. Predictive uncertainties were observed lower in French and English basins compared to Nepalese and Japanese basins. For the English and French basins which are relatively wetter, the model parameter uncertainties were found comparatively less. This is in accordance with the observation of Yapo et al. (20) who also observed that the reduction in parameter uncertainty was maximal when the wettest data periods on record were used.

Once the joint distribution of model parameter are estimated using the regional estimation equations, 500 various sets of model parameters that satisfy equation 3 at 90% confidence level were sampled. The AWISF corresponding to the plausible set of model parameters sampled from estimated joint distribution was compared with the same obtained from parameters sampled from the calibrated joint distribution (Fig. 5). The uncertainties in the model prediction quantified from the parameters sampled from their regionalized and calibrated joint distribution were significantly similar to one another. The average value of AWISF obtained from the model parameters sampled from their calibrated joint distribution was approximately 28% for basins considered for calibration and 25% in basin presumed ungauged, whereas it was 34% and 33% in basins considered for calibration and validation respectively (referred to as estimated in Fig. 5). Furthermore, the ensemble of flow simulated with parameters sampled from regionalized joint distribution of model parameters for two

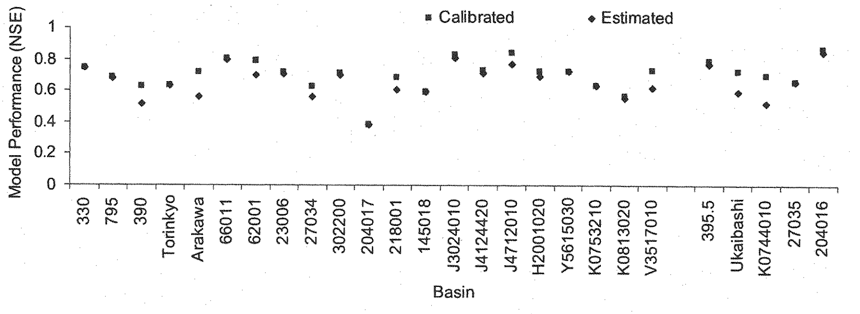

Fig.4 Predictive performance of Regional model compared to the calibrated performance.

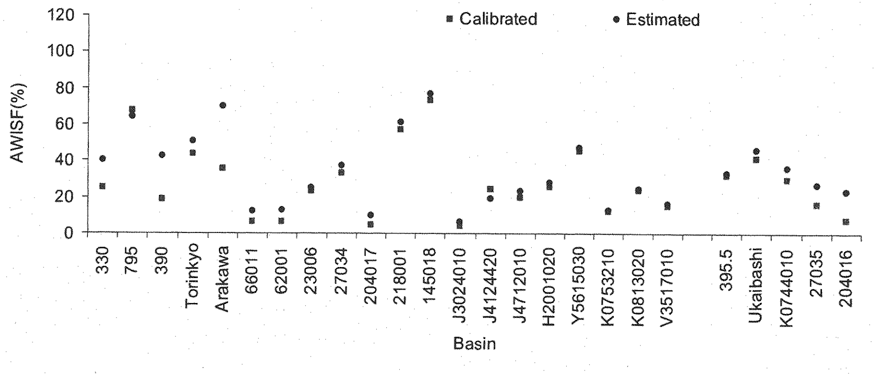

Fig.5 Average Width of the Interval of Simulated Flow (AWISF) corresponding to model parameters sampled from regionalized Joint distribution of model parameters (Estimated) and joint distribution of model parameters conditioned on observation (Calibrated).

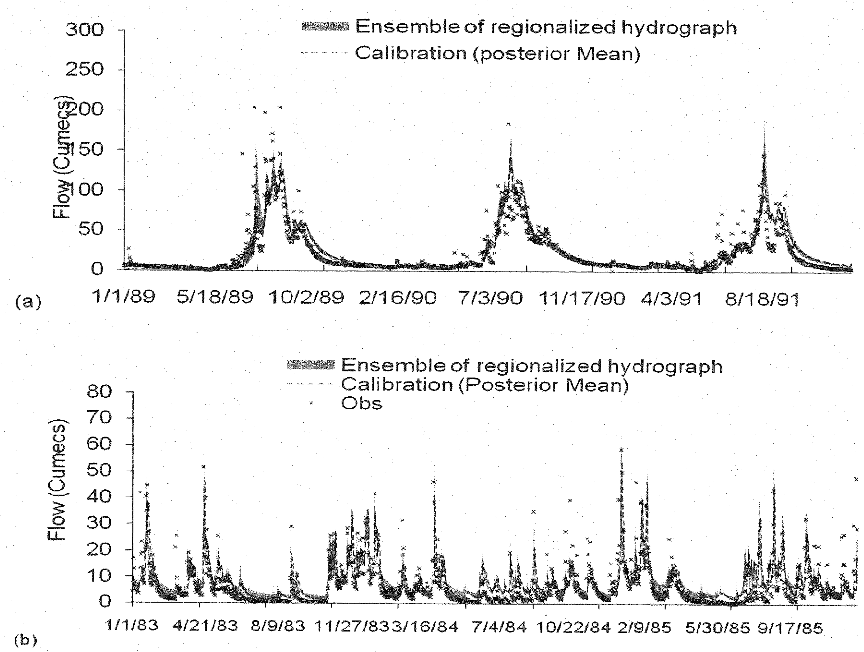

Fig.6 Ensemble of simulated hydrograph obtained from the MPs sampled from regionalized posterior probability distribution of MPs for two basins presumed ungauged (a) 330 (Nepal), (b) 27034 (UK)

selected basin is shown in Fig. 6. Even though the simulated ensemble flow encapsulated the simulated flow corresponding to posterior mean value of model parameters, it could not encapsulate completely the observed time series. The observed annual average flows for most of the basins were within the ensemble ranges except for Torinkyo (Japan), K0813020 (France) and Y5615030 (France) where it fell marginally outside the ranges of simulated annual average flow. This is most likely due to the insufficiency of model structure, because, even the ensemble of flow simulated using parameters sampled from the calibrated joint distribution of model parameters failed to encapsulate the annual average flow in these three basins.

CONCLUSIONS

Parameters of hydrological model can not be identified as a single value due to the interaction among model parameters. Close observation of the joint distribution of model parameters conditioned on observed stream flow revealed that despite using the parsimonious model with only three parameters to be calibrated, the uncertainty in the model prediction is apparent. Moreover, variations in the characteristic parameters of joint distribution of model parameters, approximated as multivariate normal distribution, were apparent among basins. Few catchment attributes were found correlated with the characteristic parameters of joint distribution of model parameters, but it is difficult to entirely define such correlation or regional relationship through physical reasoning. Apart from basin attributes, parameter uncertainty also depends on objective function used for calibration, and data quality/quantity. However, some relationship or pattern between uncertainty and basin attributes and data aspects is apparent. Therefore, an attempt to regionalize the joint distribution of model parameters is demonstrated through the case study involving a number of small to medium size basins and TOPMODEL. The prediction uncertainties quantified from the model parameters sampled from regionalized joint distribution and falling within 90% confidence region closely followed the same quantified from the model parameters sampled from their joint probability distribution conditioned on observation indicating the potential for such methods in dealing with the model parameter uncertainty while carrying out regionalisation.

ACKNOWLEDGMENTS

The authors express special thanks to Ministry of Education, Culture, Sports, Science and Technology of Japan and the 21st Century Center of Excellence (COE), Global COE program (GCOE), University of Yamanashi for providing financial support for this research.

REFERENCES

1. Merz, R. and Blöschl, G. : Regionalisation of catchment model parameters, Journal of Hydrology, Vol.287, pp.95-123, 2004.
2. Wagener, T., Wheater, H.S. and Gupta, H.V. : Rainfall-Runoff Modelling in Gauged and Ungauged Catchments, Imperial College Press, London, 300, 2004.
3. Heuvelmans, G., Bart. M. and Jan. F. : Regionalisation of the parameters of a hydrological model: Comparison of linear regression models with artificial nets, Journal of Hydrology, Vol.319, pp.245-265, 2006.
4. Bastola, S., Takeuchi, K. and Ishidahira, H. : An improved framework for the parameters regionalisation of hydrological model, Annual Journal of Hydraulic Engineering, JSCE, Vol.51, pp.43-48, 2007.
5. Parajka, j., Blöschl, G. and Merz, R. : Regional calibration of catchment models: potential for ungauged catchments, Water Resource Research, Vol.43(6), doi:10.1029/2006WR005271, 2007.
6. Kuczera, G. and Parent, E.: Monte Carlo assessment of parameter uncertainty in conceptual catchments models: the metropolis algorithm. J. Hydrol., 211, pp.69-85, 1998.
7. Marshall, L., Nott, D. and Sharma, A.: A comparative study of markov chain Monte Carlo methods for conceptual rainfall-runoff modeling. Water Resour. Res., 40, W02501, 2004.
8. Vrugt, J.A., Gupta, H.V., Bastidas, L.A., Bouten, W. and Sorooshian, S.: A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. Water Resour. Res. 39(8), 1201, doi:10.1029/2002WR001642, 2003.
9. 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 Resource Research., Vol.*39*(8), doi: 10.1029/2002WR001746, 2003.
10. Uhlenbrook, S., Seibert, J., Leibundgut, Ch. and Rodhe, A. : Prediction uncertainty of conceptual rainfall-runoff models caused by problems in identifying model parameters and structure, Hydrological Science Journal, Vol.44 (5), pp.779–797, 1999.
11. Seibert, J. : Estimation of parameter uncertainty in the HBV model, Nordic Hydrology, Vol.28, pp.247–262, 1997.
12. Scheuer, Ernest M. and David S. Stoller: On the generation of normal random vectors. Technometrics 4 (may), pp 278-281, 1962.
13. Xie Z. and Fei, Y. : A parameter estimation of VIC land surface model for the MOPEX watersheds in France, the Seventh IAHS Scientific Assembly, Foz do Iguaçu, Brazil, 3-9 April, 2005.
14. Shrestha, S., Dulal, K.N., Bastola, S., Ishidaira, H., Kazama, F. and Takeuchi, K.: The assessment of spatial transferability of a distributed hydrological model parameters in different physiographic regions of Nepal. Annual Journal of Hydraulic Engineering, JSCE, Vol.51, 2007.
15. Reichl, J.P.C., Chiew, F.H.S. and Western, A.W. : Model averaging, equifinality and uncertainty estimation in the modelling of ungauged catchments. In: Voinov, A., Jakeman, A., Rizzoli, A., (eds), Proceedings of the iEMSs Third Biennial Meeting: "Summit on Environmental Modelling and Software". International Environmental Modeling and Software Society, Burlington, USA, July. 2006.

16. 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.
17. Tang, Y., Reed P. and Wagener, 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.
18. Box, G.E.P. and Tiao, G.C.: Bayesian Inference in Statistical Analyses, Addison-Wesley-Longman, Reading, MA,USA, 1973.
19. Sefton, C. and Howarth, S. : Relationships between dynamic response characteristics and physical descriptors of catchments in England and Wales, Journal of Hydrology, 211(1-4), pp.1-16, 1998.
20. Yapo, P., Gupta, H. and Sorooshian, S.: Automatic calibration of conceptual rainfall-runoff models: sensitivity to calibration data, Journal of Hydrology, 181 (1-4), 23-48, 1996.
21. Gupta V.K. and Sorooshian S. : The relationship between data and the precision of parameter estimates of hydrologic models, Journal of Hydrology, 81 (1-2), pp.57-77, 1985.
22. Allen, Theodore T.: Introduction to Engineering Statistics And Six Sigma: Statistical Quality Control And Design of Experiments And Systems, Array London, Springer, 2006.

APPENDIX – NOTATION

The following symbols are used in this paper:

C = Cholesky Factor of Σ;

$COV(m,To)_i$ = covariance between m and To (model parameter) for i^{th} basin;

$COV(m,Td)_i$ = covariance between m and Td (model parameter) for i^{th} basin;

$COV(To,Td)_i$ = covariance between To and Td (model parameter) for i^{th} basin;

$Var(m)_i$ = variance of the model parameter m for i^{th} basin;

$Var(To)_i$ = variance of the model parameter To for i^{th} basin;

$Var(Td)_i$ = variance of the model parameter Td for i^{th} basin;

X = vector of variates;

Z = n-component vector of independent standard normal variates;

β = regional parameters;

θ = model parameters;

$\bar{\theta}$ = mean model parameter vector;

μ = vector of mean values;

Σ = covariance matrix of the joint distribution of model parameters; and

Φ = vector of catchment attributes.

(Received Jun 20, 2008 ; revised Apr 27, 2009)