Jul 1st, 12:00 AM

A Toolbox for the Identification of Parsimonious Semi-Distributed Rainfall-Runoff Models: Application to the Upper Lee Catchment

B. Orellana

I. G. Pechlivanidis

Neil McIntyre

H. S. Wheater

T. Wagener

A Toolbox for the Identification of Parsimonious Semi-Distributed Rainfall-Runoff Models: Application to the Upper Lee Catchment

B. Orellana[a], I.G. Pechlivanidis[a], N. McIntyre[a], H.S. Wheater[a] and T. Wagener[b]

[a] *Department of Civil and Environmental Engineering, Imperial College London, SW7 2AZ, London, UK (a.orellana02/i.pechlivanidis05/n.mcintyre/ h.wheater@imperial.ac.uk)*

[b] *Department of Civil and Environmental Engineering, The Pennsylvania State University, 226B Sackett Bldg., University Park, Pennsylvania 16802, USA (Thorsten@engr.psu.edu)*

Abstract: In operational hydrology identification of an appropriate model structure and suitable parameter sets for a specific catchment is a challenging task. This identification process is often based on data availability, catchment characteristics and modelling objectives, and will often result in a range of different model structures. This process of model identification becomes even more challenging when moving from lumped to distributed models as the potential number of model parameters increases proportionally to the number of spatial units considered, and due to the existence of ungauged spatial units. A Semi-Distributed Rainfall-Runoff Modelling Toolbox (RRMT-SD) has been developed to estimate continuous streamflow at points along the river system using conceptual and hybrid representations of the rainfall-runoff processes which vary from low to medium model complexity. The user can easily implement different model structures and calibration strategies, considering multiple objective functions and Monte Carlo analysis. To show the potential of the toolbox a case study on the Upper Lee catchment (1040 km^2), UK, using hourly time-steps is presented. The study area was divided into gauged subcatchments and each one of them represented through smaller spatial units of similar areas. Different model structures were applied on the spatial units using estimated a priori parameter values based on a simple regression method. The models were calibrated using spatial multipliers to adjust the a priori parameter values to the scale of the spatial units. Results showed that for different types of subcatchments (low and high base flow types) two soil moisture model structures (the Probability Distributed Moisture Model and the Catchment Wetness Index, respectively) were justified, and that parsimonious semi-distributed rainfall-runoff models on the Upper Lee catchment can perform reasonably well for a single criteria (e.g. average NSE values of 0.74).

Keywords: Rainfall-runoff models; RRMT; Semi-distributed; Parsimonious modelling; Calibration; Regionalisation.

1. INTRODUCTION

Although conceptual rainfall-runoff (RR) models are widely utilised in practice, they usually suffer from a lack of identifiability [Beven, 2001]. Research has showed that different model structures or different parameter sets within the feasible parameter space can produce virtually indistinguishable simulated streamflows [e.g. Duan et al., 1993, Spear, 1995]. Model identifiability can be improved by reducing model complexity in terms of number of model parameters considered so that all of them can be identified from the available data (parsimonious modelling) [Wheater, 2002]. There are many examples of using parsimonious, spatially lumped models, to provide reliable streamflow estimations

[e.g. Perrin et al., 2001, Wagener, 2002, among many others]. Spatially distributed models, compared to lumped, have the potential to estimate streamflow and other state variables at interior catchment points. The literature reports numerous distributed RR models and modelling approaches that address a wide variety of issues as prediction and operational forecasting environments [e.g. Turcotte et al., 2003]. Although the use of distributed models has been encouraged for many different reasons, e.g. availability of distributed rainfall data, and data on distributed catchment properties, there are issues that emerge such as scale and parameter estimation [Wooldridge et al., 2001], but also calibration strategy [Carpenter and Georgakakos, 2006]. Particularly, as the number of model parameters increases with the degree of spatial discretisation, distributed models can easily become overparameterised and subsequently ill-posed with respect to the available input-output data. Thus uncertainty in parameter estimates and hence uncertainty in model identification is a common problem [Madsen et al., 2002].

Semi-distributed RR models have been suggested to combine the advantages of lumped and distributed approaches. They can be considered as simpler distributed models that apply conceptual or hybrid metric-conceptual RR models to spatial units smaller than the catchment area. This paper presents a Semi-distributed Rainfall-Runoff Modelling Toolbox (RRMT-SD) that has been developed to estimate streamflow at points along the river system using parsimonious conceptual model structures. As an example, RRMT-SD is applied to 8 years of hourly historic data from the Upper Lee catchment, UK. Several model structures were applied and different calibration strategies investigated.

2. THE SEMI-DISTRIBUTED RAINFALL RUNOFF MODELLING TOOLBOX

The RRMT-SD has been developed to estimate continuous streamflow at points along the river system. It allows efficient building and evaluation of spatially semi-distributed rainfall-runoff models. The catchment is conceptualised through a network of subunits and a parsimonious model structure is applied to each subunit. RRMT-SD is an extension of the generic Rainfall-Runoff Modelling Toolbox (RRMT) [Wagener et al., 2002] developed previously to produce parsimonious lumped model structures estimating streamflow at the catchment outlet. RRMT has been applied for prediction and to investigate model identifiability in gauged and ungauged catchments [e.g. Wagener and McIntyre, 2005, Lee et al., 2006].

The user builds up a rainfall-runoff model on each subunit (Fig. 1(A)) by selecting from a list of soil moisture accounting (SMA) models, routing models and channel routing models. The user also has the ability to add new modules providing additional flexibility. The SMA module determines effective rainfall (ER), actual evapotranspiration (AET) and an estimation of soil moisture status; the runoff routing module considers a fast and slow runoff routing; and the channel routing module estimates discharge at the outlet of the subunits. The SMA and runoff routing models available correspond to the ones use within the established RRMT framework. Topologically, RRMT-SD simulates streamflow for the uppermost stream subunits first and accumulates the flow down the channel network.

Each subunit is considered as lumped, with the hydrological processes and climatologically forcing data homogeneous within each one of them. Then the degree of spatial distribution is represented mainly through the number of subunits, which can be subcatchments or hydrological response units.

Using the toolbox interfaces the user defines: (1) the catchment network specifying the number of subunits to consider and the corresponding identification number for each one of them; (2) the mode of the modelling environment, i.e. calibration, simulation or uncertainty, and the modelling timescale in seconds per time step which must be consistent with the temporal resolution of the input-output data; (3) the model structure for each subunit which can be specified individually for each subunit or considering homogeneity across subunits, i.e. the same model structure and parameter values. The toolbox allows for different optimisation methods for calibration: a uniform random search, the Shuffled Complex Evolution [Duan et al., 1992], and local nonlinear multi-constrained method based on simplex searching [Cormen et al., 2001].
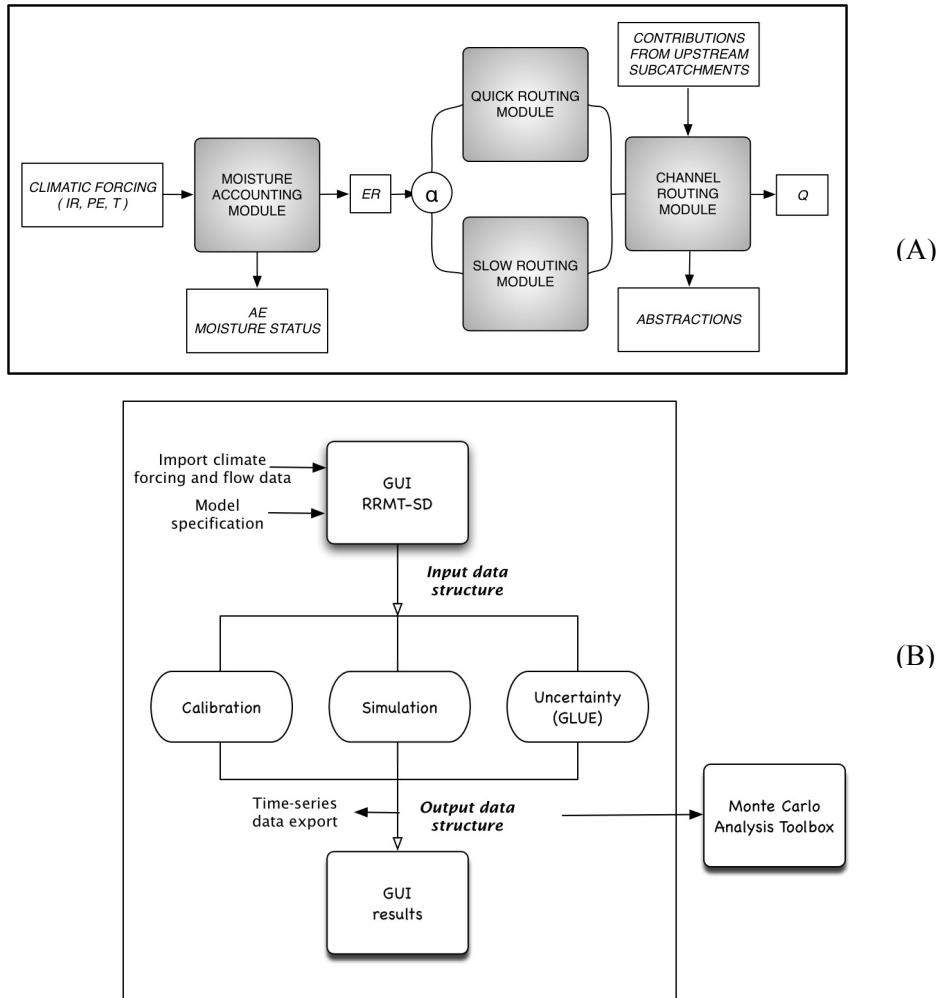
Figure 1. (A) Subcatchment module structure; (B) General Toolbox representation

The uniform random search can be used considering the same model structure and parameter sets among the subunits (URS-LP) or applying independent model structure and/or parameter sets (URS-SD). Also the user can use a priori parameter estimates (which may be non-uniform across subunits) and calibrate the model structures using calibration multipliers (URS-MP), which can be uniform or not across subunits. Multiple objective functions for measuring calibration and validation performance can be selected which are computed in all subunits where streamflow data are provided. Calibration of ungauged subunits is based on the closest downstream gauged outlet.

After finishing the modelling computations the input data and simulated variables in every subunit can be analyzed by selecting a variety of visualisation tools: input and output time series, and estimated state variables (i.e. rainfall, effective rainfall, streamflow, soil wetness, potential and actual evapotranspiration), water balances, flow duration curve, double mass plot, residual analysis to identify trends, normality and autocorrelation, and plots which show sensitivity of outputs over the parameter space.

RRMT-SD has so far been applied in a range of catchment types. A semi-distributed IHACRES model was applied to hourly data from the 734 km^2 Wadi Ahin catchment in Oman. This arid zone catchment was split into 20 topographically defined subunits in order to represent the strong spatial variability in input rainfall, and associated travel time distributions in the channel network, while parameters were assumed uniform across the catchment. The model was used for flash flood peak and volume estimation. RRMT-SD has also been applied to the small Rhosasflo catchment (1.3 km^2) in Wales, UK. Here, the 73 subunits were fields of average area 0.018 km^2 and the spatial distribution of parameter

values was defined according to land use. The model was applied to investigating the effects of land use change on flood flows. RRMT-SD is presently being applied to the large Illinois catchment, USA, as part of the international DMIP2 project, exploring calibration strategies for distributed models. Finally, RRMT-SD has been applied to the Upper Lee catchment, UK, and this application is described in some detail below.

3.   THE UPPER LEE CASE STUDY – CATCHMENT DESCRIPTION

The RRMT-SD toolbox is demonstrated here in application to the Upper Lee catchment (1040 km^2) located in the Thames region, UK, see Fig. 3. The catchment is characterised as humid temperate with mean annual precipitation of 630 mm, while the elevation varies between 20 and 250 meters above UK ordnance datum (AOD). The study area corresponds to the whole Upper Lee catchment at Feildes Weir but also considers three inner subcatchments called Upper Lee at Luton Hoo, Mimram at Panshanger, and Stort at Glen Faba (see Fig. 3).
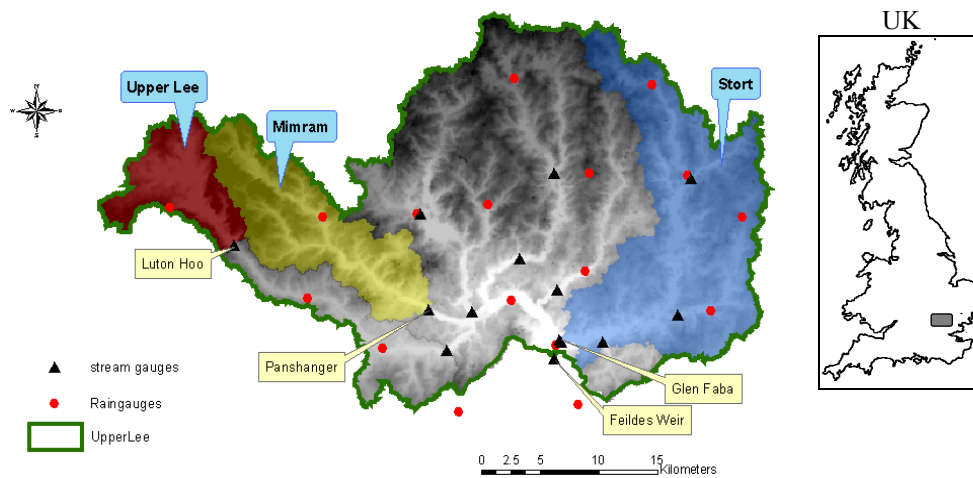
Figure 3. Map of the Upper Lee catchment, the subcatchments considered in this study are displayed in colour.

The Upper Lee subcatchment at Luton Hoo (70 km^2) is mainly chalk with high permeability. However, the subcatchment is highly urbanised with more than half covered by the city of Luton, while the land cover in the rest of the area is arable and horticulture. The subcatchment is dominated by high baseflow.

The Mimram subcatchment (130 km^2) is also mainly of chalk geology with high baseflow. Although the catchment is homogeneous and mainly rural, there is a small urban segment due to the city of Welwyn Garden. The upper part of the Stort subcatchment (278 km^2) is mainly rural, whereas the valley in the lower part is urban. Three cities of Bishop Stortford, Sawbridgeworth, and Harlow affect the runoff. The geology is chalk and clay. Overall, the Upper Lee at Feildes Weir (1040 km^2) is mainly rural, characterised by arable farming. The area has seen significant growth in housing, with urban areas covering 15 % of the total area. The Upper Lee is classified as medium baseflow dominated catchment.

A relatively fine subdivision of the subcatchments into subunits of about 10-25 km^2 was applied (Table 1), to investigate the significance of spatial rainfall on runoff generation (details can be found in Pechlivanidis et al. [2008]).

Historical hourly flow, rainfall and monthly mean potential evapotranspiration data were provided by the Environment Agency of England and Wales for the period 1991-2002. For each subcatchment, mean areal precipitation was estimated interpolating raingauge records (Fig. 3) based on the Inverse Distance Weighting method. The precipitation is assumed homogenously distributed over the area of each subcatchment. Only the stream gauges at the outlet of the study areas were considered.

Table 1. Catchment spatial discretisation into subunits

Sub Catchments	Upper Lee @ Luton Hoo	Mimram @ Panshanger	Stort @ Glen Faba	Upper Lee @ Feildes Weir
Area (km^2)	70	130	280	1040
No. Sub units	4	9	22	74
Range subunits area (km^2)	15 - 25	15 - 20	15 – 20	15 - 25

4. THE UPPER LEE CASE STUDY - MODEL DESCRIPTION AND METHODS

4.1 Semi-distributed model structure

Different RR models were built up within RRMT-SD and applied to each subunit of the Upper Lee catchment. Here only the models with better performance are included. The two SMA modules considered are (1) The Probability Distributed Moisture (PDM) model [Moore, 2007], and (2) The Catchment Wetness Index (CWI) model [Jakeman and Hornberger, 1993].

The PDM assumes that rainfall during each time step accumulates in the soil moisture store, where a specific function is used to describe the distribution of storage capacity over the catchment. The soil moisture storage is depleted by evaporation as a linear function of PE and the degree of saturation. The soil moisture storage capacity, in this study, is described by a 2-parameter Pareto distribution. In catchments with high base flow a bypass mechanism (BP) was introduced to represent groundwater runoff generation due to highly permeable soils, such as chalk. The CWI model uses an empirical model to simulate an index of wetness. The losses are proportional to the value of this index, while the gain in the index at each time-step is linearly related to the input rainfall. The effective rainfall is linearly related to the value of this index and rainfall.

The routing models applied were two linear reservoirs in parallel (2 par), 3 linear reservoirs in parallel (3par), and 2 linear reservoirs in parallel including macro-pores (2pmp).

A linear reservoir is used as channel routing for all the combinations of SMA and routing models.

4.2 Calibration strategies

Based on the Monte Carlo approach, a uniform random searching procedure was used to explore the feasible parameter space and to investigate parameter identifiability. Eight years of data were used for calibration (1991-1998) and four for validation (1999-2002). Two calibration strategies were considered. One strategy assumed that the rainfall-runoff process is homogeneous among the subcatchments thus the same model structure and parameter sets are applied on all subunits (URS-LP), and a single "optimal" parameter set was used.

The other calibration strategy estimates a priori parameters values at the subunit scale (SD-prior) by regionalisation and downscaling of calibrated parameters (see Pechlivanidis et al. [2008]). These a priori estimates were used to define the spatial variability of the parameters across subunits, and then the value of each parameter was adjusted uniformly across subunits using a multiplier which was calibrated to optimise the performance. This second calibration strategy was only applied using the PDM model because a priori parameter estimates have not yet been developed for the CWI model.

All catchments were calibrated using streamflow data at the catchment outlet using two objective functions (OFs) (Eq. 1): the Nash-Sutcliffe Efficiency (NSE) and a root mean square error based on low flows only (FSBM).

$$NSE = 1 - \frac{\sum_{i=1}^{n}(Qobs_i - Qsim_i(\theta))^2}{\sum_{i=1}^{n}(Qobs_i - Q\overline{obs})^2} \quad FSBM = \frac{\sqrt{(1/n)\sum_{i=1}^{n}(Qobs - Qsim(\theta))^2}}{(1/n)\sum_{i=1}^{n}Qobs} \quad (1)$$

where *Qsim* is the calculated flow using the parameter set *θ*, *Qobs* is the observed flow, and *n* is the length of the time series in hours. The threshold considered for low flows corresponds to 25% of the mean observed flows.

5. THE UPPER LEE CASE STUDY - RESULTS

Table 3 summarises the model performance in terms of the NSE of the various models applied during calibration. Although the PDM-2par model is performing better than the other structures in the case of low-medium baseflow catchments (Stort at Glen Faba and Upper Lee at Feildes Weir), the structure does not seem to be able to describe the hydrological processes of the high baseflow areas (Upper Lee at Luton Hoo and Mimram at Pangshanger). The CWI-2par structure appears to be the most suitable for these areas. This result is difficult to explain by the representation of the hydrological processes and it is thought to be related to the ability of CWI model to compensate for mass balance errors. In all cases the "best" model structures, as measured by NSE, represent the runoff routing as two linear reservoirs in parallel.

Fig. 3(A) shows the variability of one of the PDM's calibrated parameter values and Fig. 3(B) the relationship between the model performance (based on NSE) and the calibration multipliers used in the PDM-2par model in the Upper Lee at Feildes Weir catchment. The results demonstrate that even though an adjustment factor is assumed homogeneous over the catchment, the application of a priori estimates which are consistent with observed catchment characteristics [Pechlivanidis et al., 2008] can represent the spatial variation of model parameters in a parsimonious manner. In the case of the Upper Lee catchment at Felides Weir, some refinement of the ranges of multipliers values seem to be needed (results for Kchannel and %q in Fig. 3B).

Fig. 4 shows observed streamflow in the Stort catchment at Glen Faba and the "best" simulated streamflow series, based on the NSE criteria, using calibration that considered the same parameter sets in each subunit (URS-LP), different parameter sets (URS-Dist), and based on prior parameter values (SD-prior).

Table 3. Model performance based on the NSE criterion, on grey the selected model structure for each subcatchment based on this criterion

Models	Upper Lee @ Luton Hoo 0.87*	Mimram @ Panshanger 0.72*	Stort @ Glen Faba 0.49*	Upper Lee @ Feildes Weir 0.56*
PDM-2par	-0.09	0.51	0.86	0.84
PDM-2par(BP)	0.1	0.52	0.80	0.81
PDM-3par	0.13	0.54	0.83	0.82
PDM-3par(BP)	0.16	0.57	0.78	0.81
CWI-2par	0.61	0.72	0.67	0.66
CWI-3par	0.56	0.71	0.7	0.67
CWI-2pmp	0.57	0.68	0.69	0.65

* represents BFIHOST number

(A)

(B)
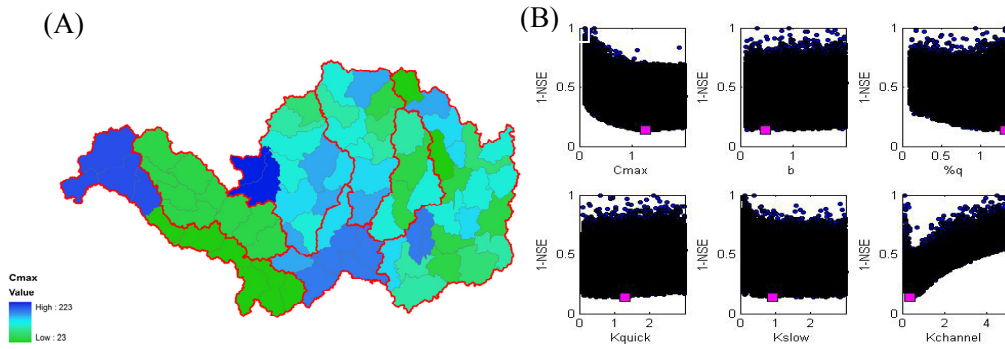
Figure 3. (A) Example of spatial parameter variability in the Upper Lee catchment; (B) Identification of calibration multipliers for PDM-2par in the Glen Faba subcatchment
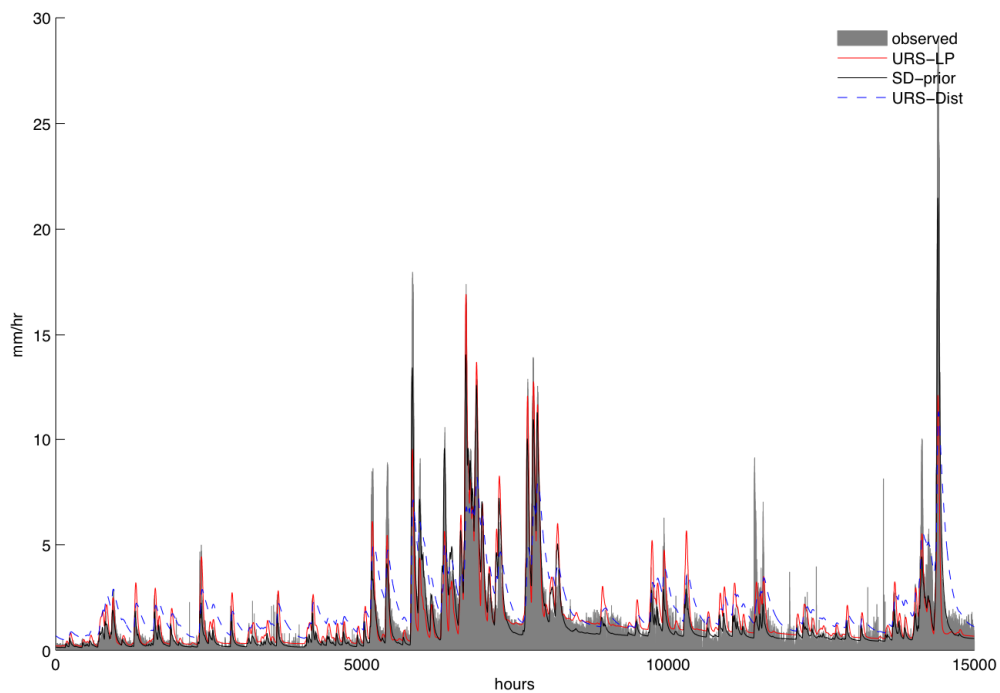
Figure 4. Simulated and observed streamflow in at Stort Glen Faba

6. CONCLUSIONS

The RRMT-SD toolkit presented here facilitates the development, analysis and application of semi-distributed parsimonious rainfall-runoff model structures. RRMT-SD has a high degree of model flexibility, which allows the implementation of established or new structures with low or medium level of complexity quickly and efficiently.

The case study of hourly data from the Upper Lee catchment showed some of the applicability of RRMT-SD to investigate calibration strategies using different criteria and inter-comparisons of conceptual model structures. Also, the case study has shown the potential of the toolbox for developing regional equations for a priori estimation of model parameters and subsequent optimisation using multipliers, hence maintaining spatial variations which are consistent with catchment characteristics while maintaining identifiability. Various other applications of RRMT-SD at Imperial College London have demonstrated the toolbox potential applicability to a range of hydrological problems, including flash flood simulation in arid regions and simulating effects of land use change. Further applications of the toolbox are in development.

REFERENCES

Ajami, N.K., H. Gupta, T. Wagener, and S. Sorooshian, Calibration of a semi-distributed hydrologic model for streamflow estimation along a river system, *Journal of Hydrology*, vol. 298, no. 1-4, pp. 112-135, 2004.

Beven, K., *Rainfall-Runoff modelling. The Primer* John Wiley and Sons, Chichester, UK, 2001.

Boyle, D.P., H.V. Gupta, S. Sorooshian, V. Koren, Z.Y. Zhang, and M. Smith, Toward improved streamflow forecasts: Value of semidistributed modeling, *Water Resources Research*, vol. 37, no. 11, pp. 2749-2759, 2001.

Carpenter, T.M., and K.P. Georgakakos, Intercomparison of lumped versus distributed hydrologic model ensemble simulations on operational forecast scales, *Journal of Hydrology*, vol. 329, no. 1-2, pp. 174-185, 2006.

Cormen T.H., C.E. Leiserson, R.L. Rivest, and C. Stein, *Introduction to Algorithms*, Second Edition. MIT Press and McGraw-Hill, Section 29.3: The simplex algorithm, pp.790–804, 2001.

Duan, Q., V.K. Gupta, and S. Sorooshian, A shuffled complex evolution approach for effective and efficient global minimization, *Journal of Optimization Theory and Applications*, vol. 76, no. 3, pp. 501-521, 1993.

Gupta, H.V., S. Sorooshian, and P.O. Yapo, Towards improved calibration of hydrologic models: Multiple and noncommensurable measures of information, *Water Resources Research*, vol. 34, no. 4, pp. 751-763, 1998.

Koren, V., M. Smith, and Q. Duan, Use of a priori parameter estimates in the derivation of spatially consistent parameters sets of rainfall runoff models, in *Calibration of Watershed Models*, vol. 6 American Geophysical Union, pp. 239-254, 2003.

Lee, H., N. McIntyre, H.S. Wheater, and A.R. Young, Predicting runoff in ungauged UK catchments, *Water Management*, vol. 159, no. WM2, pp. 129-138, 2006.

Madsen, H., G. Wilson, and H.C. Ammentorp, Comparison of different automated strategies for calibration of rainfall-runoff models, *Journal of Hydrology*, vol. 261, no. 1-4, pp. 48-59, 2002.

Pechlivanidis, I.G., N. McIntyre, and H.S. Wheater, Relating conceptual rainfall-runoff model structure and parameters to catchment characteristics: A case study on the Upper Lee catchment, in *HYDRA Postgraduate Research Poster Meeting*, 17 December, University College London, London, UK, 2007.

Pechlivanidis, I.G., N. McIntyre, and H.S. Wheater, The significance of spatial variability of rainfall on runoff, in *International Congress on Environmental Modelling and Software* (iEMSs 2008), 7 July, Barcelona, Spain, 2008.

Perrin, C., C. Michel, and V. Andreassian, Does a large number of parameter enhance model performance? Comparative assessment of common catchment model structures on 429 catchments, in *Journal of Hydrology*, vol. 242, pp. 275-301, 2001.

Wagener, T., D.P. Boyle, M.J. Lees, H. S. Wheater, H. V. Gupta, and S. Sorooshian, A framework for development and application of hydrological models. *Hydrology and Earth System Sciences*, vol. 5, no. 1, pp. 13-26, 2001.

Wagener, T., M.J. Lees, and H.S. Wheater, A toolkit for the development and application of parsimonious hydrological models, in *Mathematical models of large watershed hydrology - Volume 1*, 1 edn, V. P. Singh & D. Frevert, eds., USA, pp. 87-136, 2002.

Wagener, T., and N. Mcintyre, Identification of rainfall-runoff models for operational applications, in *Hydrological Sciences,* 50(5), pp. 735-751, 2005

Wheater, H.S., A.J. Jakeman, and K.J. Beven, Progress and directions in rainfall-runoff modelling, in *Modelling change in environmental systems*, A. J. Jakeman, M. B. Beck, and M. J. McAleer, eds., New York, pp. 101-132, 1993.

Wheater, H.S., Progress in and prospects for fluvial flood modelling, in *Philosophical Transactions of the Royal Society of London Series A - Mathematical Physical and Engineering Sciences* 360 (1796), 1409-1431, 2002.

Wooldridge, S., J. Kalma, and G. Kuczera, Parameterisation of a simple semi-distributed model for assessing the impact of land-use on hydrologic response, in *Journal of Hydrology*, vol. 524, pp. 16-32., 2001