Modelling ungauged basins with the Sacramento model

TERRI HOGUE[1], KORAY YILMAZ[2], THORSTEN WAGENER[3] & HOSHIN GUPTA[2]

1 *Dept of Civil and Environmental Engineering, University of California, Los Angeles, California, USA*
thogue@seas.ucla.edu

2 *Dept of Hydrology and Water Resources, University of Arizona, Tucson, Arizona, USA*

3 *Dept of Civil and Environmental Engineering, Pennsylvania State University, College Station, Pennsylvania, USA*

Abstract This paper evaluates two contrasting approaches to parameter estimation for ungauged basins using the US National Weather Service's SACramento Soil Moisture Accounting (SAC-SMA) model. An automatic calibration scheme (Multi-Step Automatic Calibration Scheme, MACS) provides deterministic parameter estimates using a three-step, multiple objective approach. The MACS estimates are then transferred to similar or "sister" watersheds for basins in the French MOPEX data set. Physically-based parameter estimates are also developed for the same basins based on the *a priori* approach of Koren *et al*. (2000). In general, the two methods, the transfer and the *a priori* approaches, show similar overall performance. Parameter estimates appear more consistent between basins using the *a priori* approach, but statistically the regionalized MACS parameters and the *a priori* parameters show very similar model performance for the three basins investigated in this study. Model simulated hydrographs are also very similar between the two methods, with both methods tending to underpredict most events (peak and volume) but matching the shape and pattern of flow well. However, both methods have worse performance than a calibrated model for the same basin, indicating the possibility for further refinement and adjustment of the techniques presented here.

Key words parameter estimation; rainfall–runoff modelling; regionalization; Sacramento model

INTRODUCTION

A priority of MOPEX is to improve the estimation of parameters, primarily through the development of *a priori* methods. The goal of this paper is to compare two fairly recent, but diverse, methods for estimating parameters for ungauged basins. One method is based on the transfer of parameters from calibrated, neighboring catchments using an automated procedure for estimation in the gauged basins, and the second method is based on the estimation of *a priori* estimates from relationships developed between soil physics and model parameters. This study uses the conceptual rainfall–runoff SACramento Soil Moisture Accounting (SAC-SMA) model used by the National Weather Service (NWS) for forecasting river flows in the United States. We test the *a priori* and regionalization methods on three basins in France. This work is directly tied to the goals of MOPEX, which are to advance the state of knowledge of parameter estimation techniques and provide guidance on procedures for improvement of *a priori* estimates of parameters for land surface and hydrological models.

METHODS

The two methods used here were developed for the case in which streamflow observations are not available for calibration of the SAC-SMA model (i.e. the basins under analysis are ungauged). These parameter estimation methods include: (1) a direct estimate of parameters for the ungauged basin using theoretical understanding of (small-scale) soil physics proposed by Koren *et al*., (2000); and (2) transfer of parameters calibrated on neighboring basins to the ungauged basin. A third possible approach of parameter regionalization using statistical regression equations (e.g. Wagener *et al*., 2004) was not tested here due to the small number of basins available. The second method (sister approach) involved calibrating the model to streamflow observations in each basin using an automated procedure (MACS; Hogue *et al*., 2000, 2006), and applying the mean parameter values of the calibrated estimates to the ungauged basin. Ideally, and in our future work, one could use physically-derived relationships (i.e. watershed size, precipitation/evaporation (PE) ratio, mean flow ratios, etc.) to adjust the parameters before application to the ungauged basin. The sister approach assumes that there will be some regional basins with observational data for calibration, and that geographically proximal basins behave in a similar manner in terms of their hydrological response. The advantage of the *a priori* method lies in the "physical" relationship of the derived parameters to actual watershed properties. In principle, the *a priori* approach should result in more consistent parameter distributions within (distributed models) and between watersheds. A disadvantage of this method, however, is the difficulty of finding reliable and accurate watershed information (soils, land use, etc.). Other challenges include unexplored problems of scale between the watershed data and model parameters.

Study sites

Table 1 lists the watershed identifiers along with their soil and/or watershed characteristics for our three study basins. Each basin was treated as ungauged and parameters were derived using (1) *a priori* estimates, and (2) a mean parameter set from two regional sister-basins. Each of the methods, along with the model, is outlined in more detail below.

Table 1 Study basins with identifier, size, land cover and soil type.

Basin ID	Size (km^2)	Land cover	USDA soil texture class
J3024010	43	Agricultural (50%) Arable land (40%) Other (10%)	4 (100%)
V6035010	150	Forest (70%) Scrub and herbaceous (30%)	6 (80%) 4 (20%)
Y5615030	279	Forest (40%) Natural grassland (36%) Urban (10%) Other (14%)	4 (55%) 6 (45%)

Model

The SAC-SMA is a conceptual model using a two-layer soil moisture system to continuously account for storage and flow through the soil layers. The upper layer represents surface soil regimes and interception storage, while the lower layer represents deeper soil layers and groundwater storage (Brazil & Hudlow, 1981). Each layer consists of fast components (free water), driven mostly by gravitational forces, and slow components (tension water), driven by evapotranspiration and diffusion. The SAC-SMA, with a total of 16 parameters (Table 2), is a saturation excess model; when precipitation amounts exceed percolation and interflow capacities, upper zone storage will overflow and overland flow will occur. Inputs to the model are Mean Areal Precipitation and Potential Evapotranspiration (PET). An evapotranspiration demand curve (or adjustment curve) is used for estimating the potential evaporation for the watershed. Output from the model, channel inflow, is routed through a unit hydrograph for forecasting basins in the USA.

The model was run at the hourly time-step for each of the study basins. Channel routing was performed using a series of Nash-cascade linear reservoirs. A Monte Carlo based sensitivity analysis was used to determine the optimum number of reservoirs; for the basins under study five reservoirs (each using the same recession coefficient, k_route) was found to be optimal. In the MACS procedure, k_route was optimized along with 13 SAC-SMA parameters. In the *a priori* method, k_route was estimated by averaging the MACS obtained value for the other two basins.

A priori parameter estimation

Koren *et al*. (2000) present equations (hereafter called Koren equations) for deriving *a priori* estimates for the 11 major SAC-SMA parameters from soil texture, hydrological

Table 2 List of parameters within the SAC-SMA model and their description.

SAC-SMA	Parameter description
UZTWM	Upper zone tension water max. storage (mm)
UZFWM	Upper zone free water max. storage (mm)
LZTWM	Lower zone tension water max. storage (mm)
LZFPM	Lower zone free water primary max. storage (mm)
LZFSM	Lower zone free water suppl. max. storage (mm)
UZK	Upper zone free water lateral depletion rate (day^{-1})
LZPK	Lower zone prim. free water depletion rate (day^{-1})
LZSK	Lower zone suppl. free water depletion rate (day^{-1})
ADIMP	Additional impervious area (fraction)
PCTIM	Impervious fraction of the watershed (fraction)
ZPERC	Maximum percolation rate (dimensionless)
REXP	Exponent of the perco. equation (dimensionless)
PFREE	% of water percolating directly to lower zone free water storage
RIVA	Riparian vegetation (fraction)
SIDE	Ratio of deep recharge to channel baseflow (fraction)
RESERV	% of lower zone free water not transferable to lower zone tension water

soil group, soil depth and vegetation. The physical basis of the Koren equations and their derivation are given by Duan *et al*. (2001) and Koren *et al*. (2003), therefore only a brief overview is given here. The Koren equations assume that the SAC-SMA tension water storages are related to "available" soil water and that the free water storages are related to "gravitational" soil water. Available and gravitational soil water can be derived from soil properties (i.e. saturated moisture content, field capacity, and wilting point). Following Koren *et al*. (2003), regression equations derived by Cosby *et al*. (1984) were used together with Campbell's matric water potential equation to determine soil hydraulic properties from the USDA soil texture class information.

Soil texture classes in the French MOPEX Data set (available in GIS), were mapped to the USDA soil texture classes (Table 1). A soil depth of 2.5 m was assumed. Soil hydraulic properties for each basin were defined as the area average of soil property values of soil texture polygons in a basin. The combined thickness of the upper and lower layers was assumed to be equal to the soil profile depth. The thickness of the upper layer was estimated using a concept of initial rain abstraction based on the Curve Number (CN) classification system developed by The Natural Resources Conservation Service (NRCS) (McCuen, 1982). Following Koren *et al*. (2003), it was assumed that under the average soil moisture conditions stipulated by NRCS, the upper layer tension water storage is full and the free water storage is empty. In this case initial rain abstraction should satisfy the upper layer free water capacity (Koren *et al*. 2003). The upper layer thickness can then be calculated based on a CN for the soil profile. NRCS soil hydrological group was assumed to be "B" for all soils in the study basins. Information on land cover was available through the CORINE land cover project (http://reports.eea.eu.int/COR0-part1/en) launched by the Commission of the European Communities. For each soil texture class polygon, a dominant land use was assigned and the corresponding CN was estimated. For each basin an average CN was calculated as the area weighted average of the CNs in a specific basin.

Under these assumptions the Koren equations were used (see Koren *et al*., 2003) to estimate the SAC-SMA storages (UZTWM, UZFWM, LZTWM, LSFSM, LZFPM) and runoff depletion rates (UZK, LZPK, LZSK). The impervious fraction (PCTIM) of the watersheds was estimated from the percent of watershed area with land cover types of urban and bare rocks. One of the major limitations of the *a priori* methods is the scale difference between the soil hydraulic properties (point measurements) and model parameters (representative of a spatially heterogeneous model grid area of several kilometres). This limitation is being investigated in ongoing research and is not a focus of the current study.

Automated parameter estimation

MACS is a departure from previously developed single-step, single-criterion automatic calibration techniques and is based on a progressive evaluation of objective function values throughout the optimization procedure. The procedure uses the Shuffle Complex Evolution-University of Arizona (SCE-UA) algorithm developed by Duan *et al*. (1992, 1993). The method has been tested in a wide variety of hydro-climatic regimes in the USA and has been shown to produce model simulations as good as, or better in some circumstances, to traditional manual calibration techniques. Details are presented in Hogue *et al*., (2000, 2006); and therefore only a brief overview is given here.

In *step one* of MACS, 13 of the parameters of the SAC-SMA and the linear reservoir parameter (routing parameter) are selected and optimized using the LOG criterion (see below). This first run places strong weighting on the low-flow portions of the hydrograph and gives good estimates of the lower zone parameters.

$$LOG = \sum (LOG_{Qsim,t} - LOG_{Qobs,t})^2$$

where $Q_{sim,t}$ = simulated flows, and $Q_{obs,t}$ = observed flows at time step t.

The *second step* of MACS emphasizes the estimation of parameters that influence higher flow events. Lower zone parameters estimated in the first step are held constant, and a second optimization is run using the RMSE function using the upper zone parameters and routing parameter:

$$RMSE = \sqrt{1/n(\sum_{t=1}^{n}(Q_{sim,t} - Q_{obs,t})^2)}$$

Step three is run using the LOG function to fine-tune the parameters which affect the lower zone processes (upper zone parameters from step two are held constant).

The MACS procedure was run for each basin to find a set of "calibrated" parameters for the SAC-SMA. A total of 14 parameters were estimated: 13 SAC-SMA and one routing parameter (RIVA, SIDE and RSERV are set to book values). A split-sample approach was undertaken with an initial period used for calibration and a separate evaluation period run with the calibrated parameters to assess calibration adequacy. Calibrated parameters from each of the two neighbouring basins were then averaged to find parameters for the third basin, assumed to be ungauged. In this study we had the benefit of comparing this "sister" parameter set with a parameter set calibrated specifically to the basin (theoretically our ungauged basin), along with the parameter set determined via the *a priori* method.

RESULTS

Three parameter sets were compared for each study basin, including *a priori* estimation, parameters calibrated specifically to the basin (MACS), and parameters estimated using the sister approach (mean of two neighboring basins). Figure 1 shows a normalized plot of the three parameter sets for each of the three basins (J302, V603 and Y561). In general, a larger variability is seen between the study basins for the MACS parameters (parameters specifically calibrated to each basin). There is slightly better consistency between basins with the *a priori* method. Intuitively this makes sense. The basin specific calibration would tend to pick up nuances and differences in flow regimes between basins and adjust parameters accordingly. As each of the basins have fairly similar soil types (all have type 6 and two have type 4), the *a priori* estimates should have some similarity and the estimated values reflect this. The mean values transferred to each basin (sister approach) lie somewhere in between these two extremes (slightly more consistent between basins than the MACS, but not as consistent as the *a priori* values). Also of interest, there is much better consistency between all three approaches in the lower zone parameters indicating similar baseflow patterns while the parameters controlling surface flow (UZTWM through REXP) show more variability between methods.

Statistics for the model simulations are presented in Figs 2 and 3. Statistical results from the calibration and evaluation period simulations for each study basin are shown in Fig. 2. Four statistics are presented: Root Mean Squared Error (RMSE), square root RMSE (SQRT RMSE), Nash-Sutcliffe Efficiency (NSE) and percent bias (%Bias). The MACS calibration tends to outperform the *a priori* during the calibration period, showing better statistics in most cases except for a slightly lower NSE for V603.

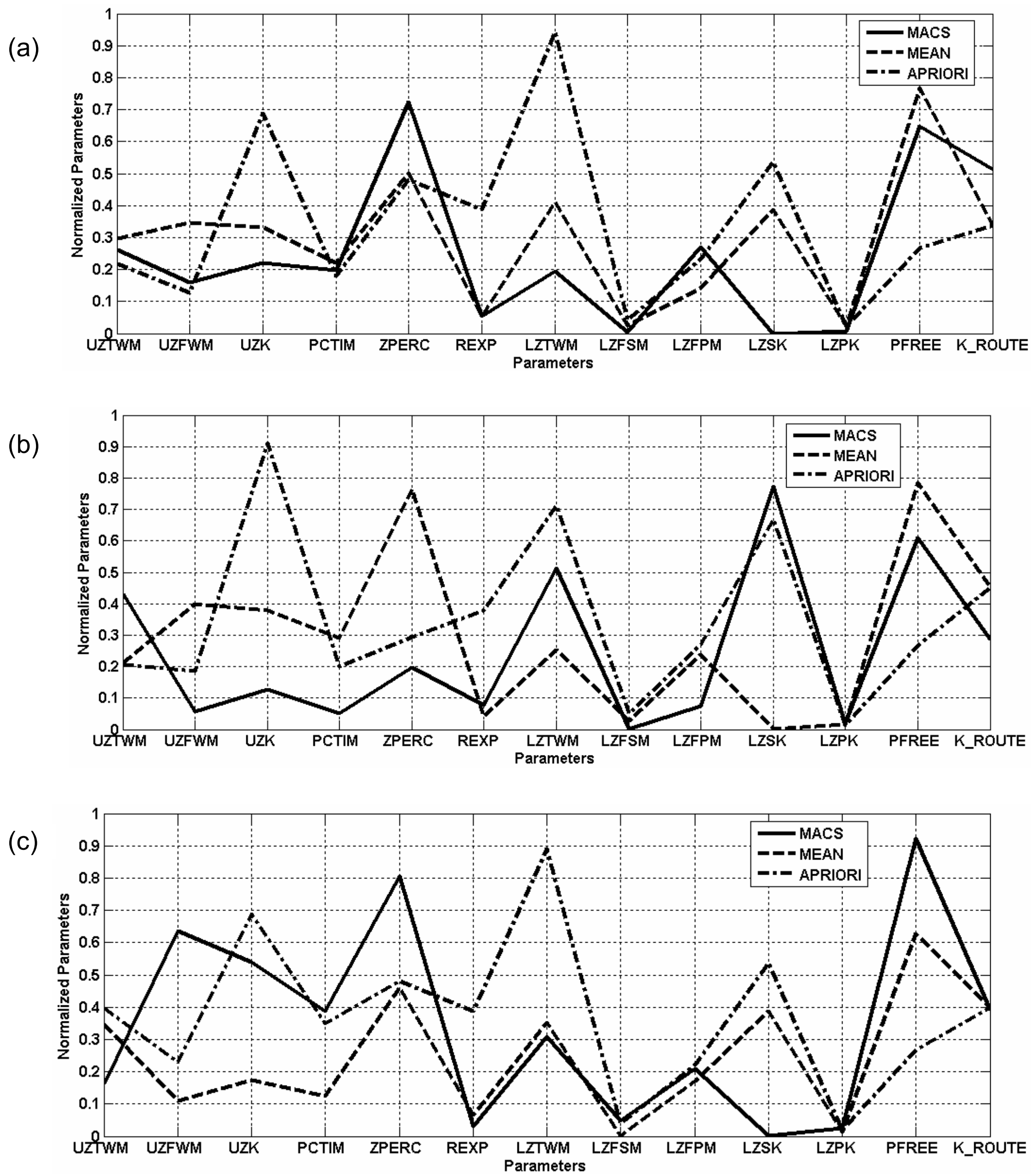

Fig. 1 Comparison of parameters from the *a priori* method, mean (or sister approach) and parameters calibrated at the specific basin (MACS) for each of the study basins: (a) J302, (b) V603, and (c) Y561.

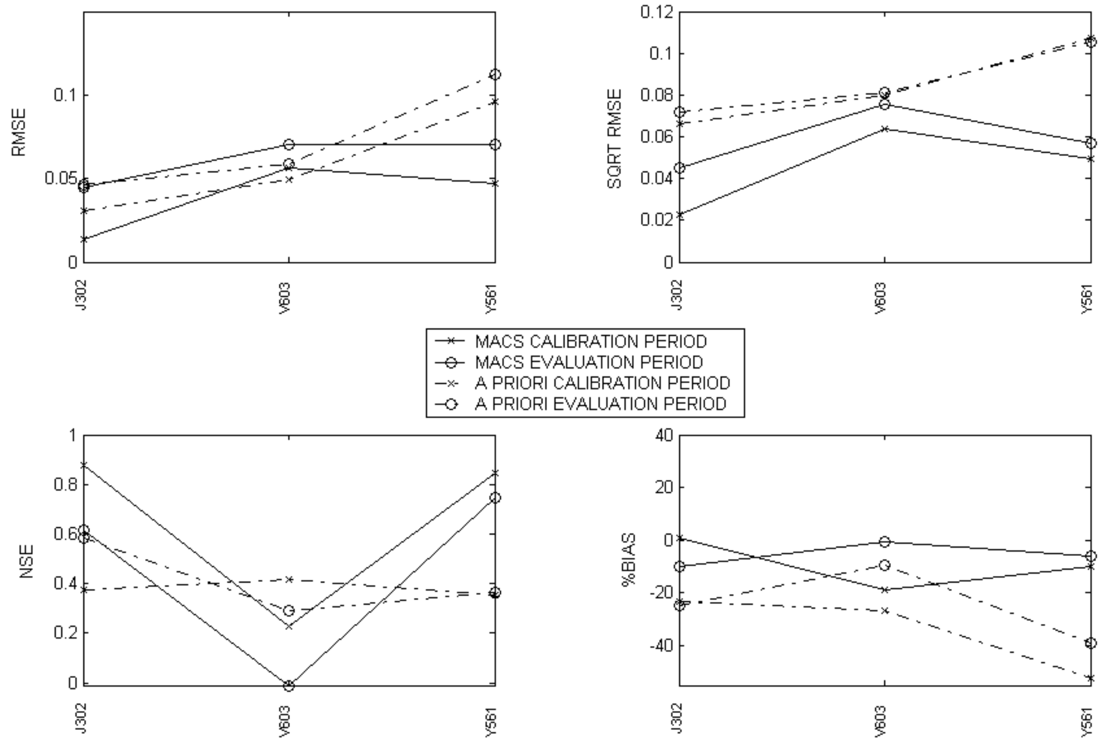

Fig. 2 Comparison of statistics for calibration and evaluation period for each study basin with MACS and *a priori* estimation methods (root mean square error (RMSE), square root RMSE (SQRT RMSE), Nash-Sutcliffe efficiency (NSE), and percent bias (%BIAS).

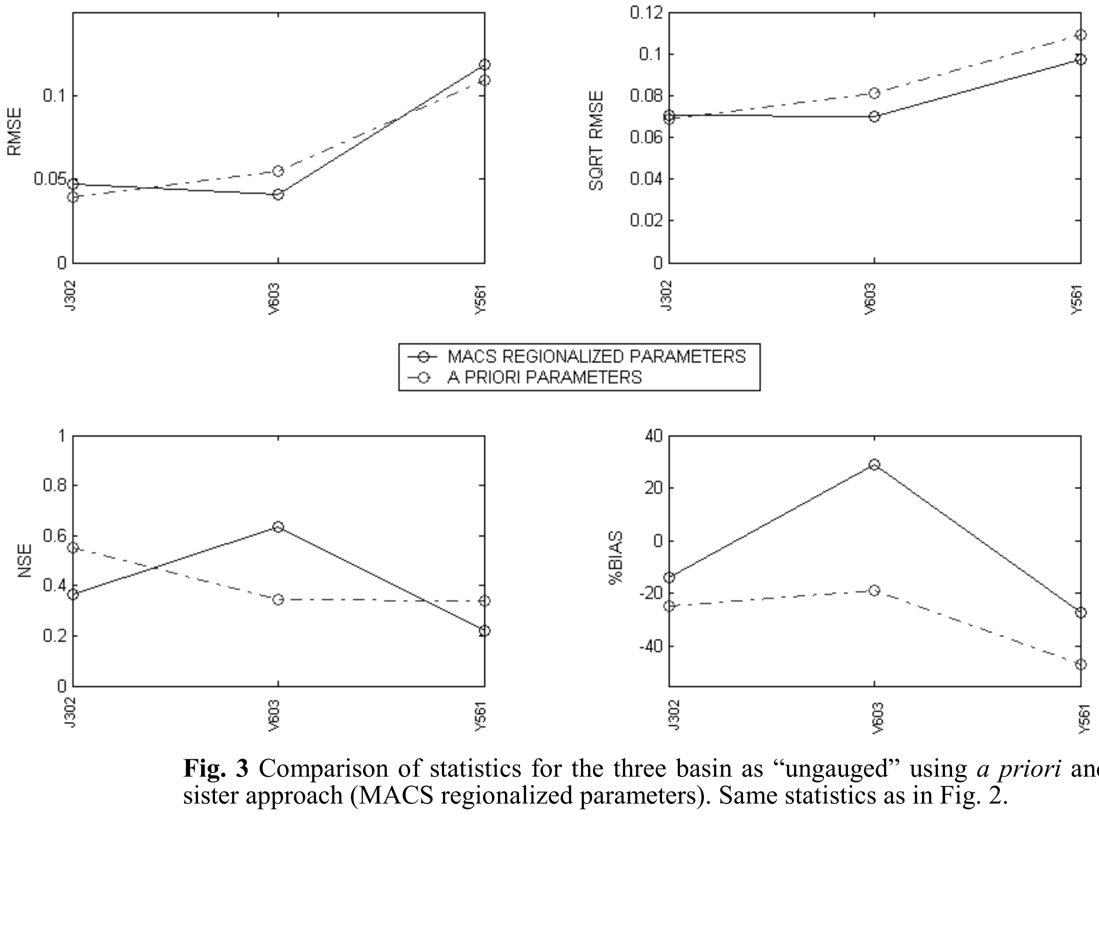

Fig. 3 Comparison of statistics for the three basin as "ungauged" using *a priori* and sister approach (MACS regionalized parameters). Same statistics as in Fig. 2.

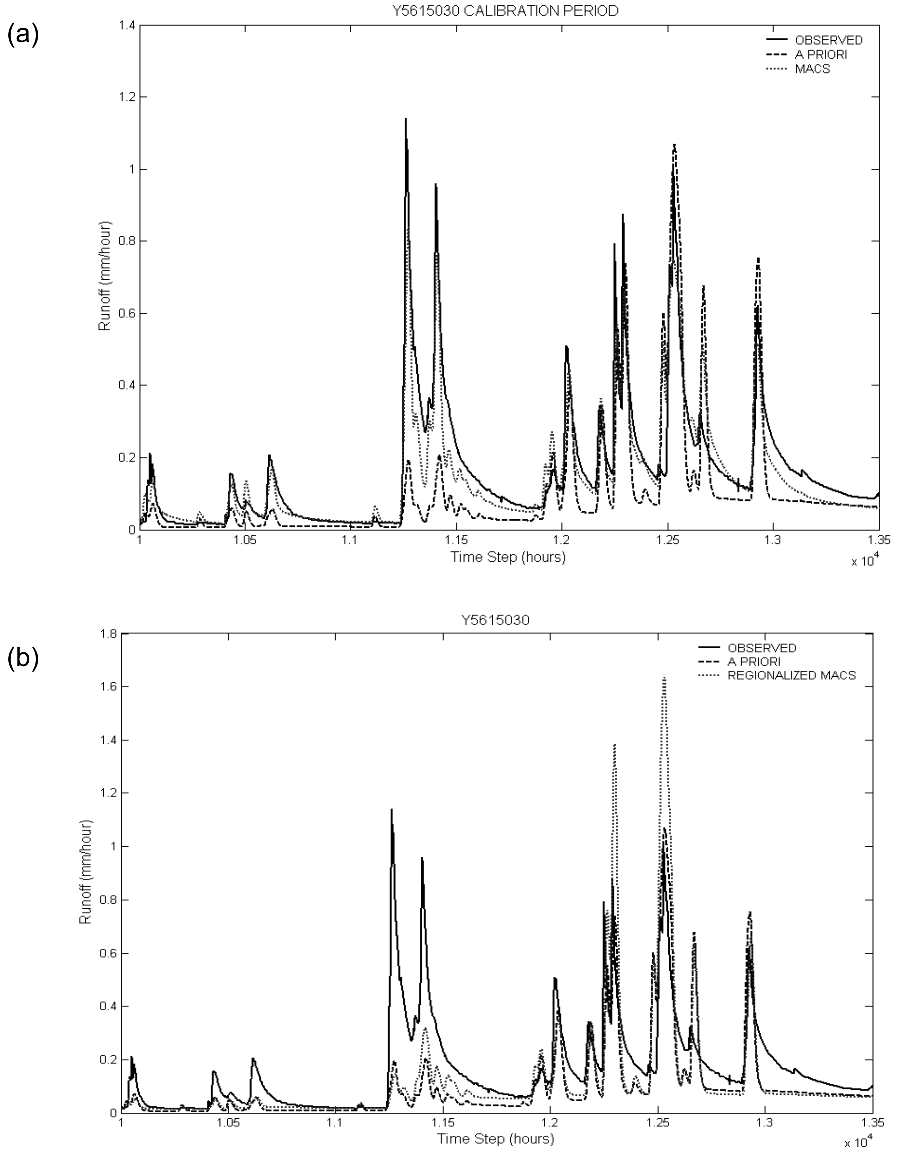

Fig. 4 Model simulations for basin Y561 as: (a) gauged basin (calibration period), and (b) ungauged basin with only *a priori* and regionalized parameters.

basin. However, the MACS performance declines somewhat during the evaluation period, showing more similarity to the *a priori* performance and even slightly worse for the V603 basin (NSE and %Bias). Figure 3 presents the same set of statistics for the basins treated as ungauged. Interestingly, performance is very similar between the two methods. *A priori* estimates tend to have a negative %Bias for all basins, while the transferred parameters show a positive percent bias only for the V603 basin. The *a priori* method does slightly better in the NSE statistic for two of the basins (J302 and Y561) while the MACS has a much better NSE in the V603 basin.

An example of model simulations for basin Y561 (our largest basin) is shown in the hydrographs presented in Fig. 4. Two simulations are shown against observed flows for the calibration period (i.e. gauged basin) in Fig. 4(a) (*a priori*, and MACS calibration) and the same period is shown in Fig. 4(b) with the basin treated as ungauged (showing observed, *a priori* and MACS regionalized parameters). In the gauged case, simulations from both methods, MACS and *a priori*, tend to fit the pattern fairly well. The *a priori* underestimates the first large event around time-step 11 200 but does better catching the peaks in the remaining period. The MACS catches most of the peaks very well and also matches the recessions and volume better than the *a priori*. In the ungauged case of the basin (Fig. 4(b)), both methods (*a priori* and transferred parameters) miss the same first large event (around time-step 11 200). Both methods do better in the remaining time period, however the regionalized parameters drastically overpredict two of the large events (around time-steps 12 200 and 12 500). Also, neither method matches the recessions during most of the simulations. In the case of the *a priori* method, this is possibly due to the lack of deep soil property information (i.e. groundwater).

SUMMARY AND CONCLUSIONS

In summary, it is observed (and is consistent with intuitive reasoning) that better model performance is obtained when the model is specifically calibrated to a basin with observational data. However, the goal of this paper was to evaluate and compare methods for estimating parameters when streamflow data are not available, and for further insight, to compare these methods to the ideal case of having calibration data. Interestingly, both the *a priori* and sister approach show similar performance for the case of the ungauged basins in our study. The simple method of taking the mean of calibrated parameters from two neighboring basins with observational data and directly applying these estimates within the ungauged basin model, showed surprisingly good results. The performance (statistics and hydrograph visualization) was nearly identical to the *a priori* estimates for two of the basins. However, this is not to say that both methods are ideal in their present form. Both have poorer statistics (and poorer hydrograph fit) than if observational data were available, indicating the need for improvement in both methods. Some of the issues to be overcome include the scaling of the point-scale soil physics to the watershed scale and the accuracy of watershed-scale soil surveys when using *a priori* methods. The sister approach could be improved by development of relationships to scale or relating calibrated parameters from basins with observational data to ungauged basins. We continue our work in both of these research areas and plan to provide more insight on these topics in future publications.

REFERENCES

Brazil, L. E. & Hudlow, M. D. (1981) Calibration procedures used with the National Weather Service river forecast system. In: *Water and Related Land Resource Systems* (ed. by Y. Y. Haimes & J. Kindler), 457–566. Pergamon Press, New York, USA.

Cosby, B. J., Hornberger, G. M., Clapp, R. B. & Ginn, T. R. (1984) A statistical extrapolation of the relationships of soil moisture characteristics to the physical properties of soil. *Water Resour. Res*. **20**(6), 682–690.

Duan, Q., Gupta, V. K. & Sorooshian, S. (1992) Effective and efficient global optimization for conceptual rainfall–runoff models. *Water Resour. Res*. **28**, 1015–1031.

Duan, Q., Gupta, V. K. & Sorooshian, S. (1993) A shuffled complex evolution approach for effective and efficient global minimization. *J. Optimization Theory Application* **76**, 501–521.

Duan, Q., Schaake, J. & Koren, V. (2001) *a priori* estimation of land surface model parameters. In: *Land Surface Hydrology, Meteorology, and Climate: Observation and Modeling*, (ed. by V. Lakshmi), 77–94. Water Science and Applications, vol. 3. AGU, Washington, DC, USA.

Hogue, T. S., Sorooshian, S., Gupta, H., Holz, A. & Braatz, D. (2000) A multi-step automatic calibration scheme for river forecasting models. *J. Hydrometeor.* **1**, 524–542.

Hogue, T. S., Gupta, H. V. & Sorooshian, S. (2006) A "user-friendly" approach to parameter estimation in hydrologic models. *J. Hydrol*. (in press).

Koren, V. I., Smith, M., Wang, D. & Zhang, Z. (2000) Use of soil property data in the derivation of conceptual rainfall–runoff model parameters. In: *Proc. 15th Conf. on Hydrology* (AMS, Long Beach, California, USA), 103–106.

Koren, V., Smith, M. & Duan, Q. (2003) Use of *a priori* parameter estimates in the derivation of spatially consistent parameter sets of rainfall–runoff models. In: *Calibration of Watershed Models, Water Science and Applications* (ed. by Q. Duan, S. Sorooshian, H. Gupta, H. Rosseau & R. Turcotte), vol. 6, 239–254. AGU, Washington, DC, USA.

McCuen , R. H. (1982) *A Guide to Hydrologic Analysis using SCS Methods*. Prentice Hall, Englewood Cliffs, New Jersey, USA.

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