A Bayesian Framework for Assessing Dynamic Hydrological Systems

[1]Marshall, L. A., A. Sharma and [2]D. Nott

[1]School of Civil and Environmental Engineering, [2]School of Mathematics, University of New South Wales,
E-Mail: lucy@civeng.unsw.edu.au

Keywords: *Rainfall-runoff modeling; Model averaging; Hierarchical Mixtures of Experts.*

EXTENDED ABSTRACT

Hydrological models are a popular tool for simulating catchment processes in response to a rainfall event. The wide range of available models means that hydrologists are faced with the problem of determining which model is best applied to a catchment in any modeling exercise. An attractive alternative to selecting a single hydrological model is to combine the results from several models, thereby providing a model performance that is substantially better than any model alone. Methods based on Bayesian statistical techniques provide an ideal means to compare and combine competing models as they explicitly account for model uncertainty. Bayesian model averaging combines individual models by weighting model outputs proportional to their respective posterior probability of selection. However, the necessity of having fixed weights over the length of the simulation period means that the relative usefulness of different models at different times is not considered.

In hydrological modeling, evidence exists of the catchment responding differently under different conditions so that relying on a rigid modeling structure can lead to significant inaccuracies and biases, particularly when used for prediction. This study combines the benefits of model aggregation and a dynamic model structure via a framework where each model is adopted at each time step with a different probability. The framework is known as a Hierarchical Mixture of Experts (HME, Figure 1). The HME framework consists of individual models (known as experts) which may have different structures and parameter values. These are grouped by mathematical functions (known as the gating function).

The probabilistic nature of the HME framework means it is ideally specified using Bayesian inference. Bayesian methods incorporate uncertainty in our observed data and design parameter values, but to date little work has been done in assessing the impact of the model structure on model uncertainty. HME is a shift towards incorporating model structural uncertainty.

Previous work has shown that an innovation of the HME framework is that it provides a way of assessing the mechanisms of our existing models to see which structures or parameter sets are preferred to describe hydrological processes under different conditions. However, the challenge still remains to apply the HME framework to catchments for predictive purposes. This problem lies in determining how to calculate which model should be selected depending on the state of the catchment.

In the study, the HME framework is applied to a catchment in NSW using two parameterizations of a simple conceptual rainfall runoff model. We investigate the usefulness of different predictors and gating functions. More complex choices for the gating function involving nonlinear or nonparametric functional terms are illustrated. The study shows that given careful comparison of predictors and gating function, HME can be a useful predictive tool, giving an aggregated model simulation that is better than any individual model.

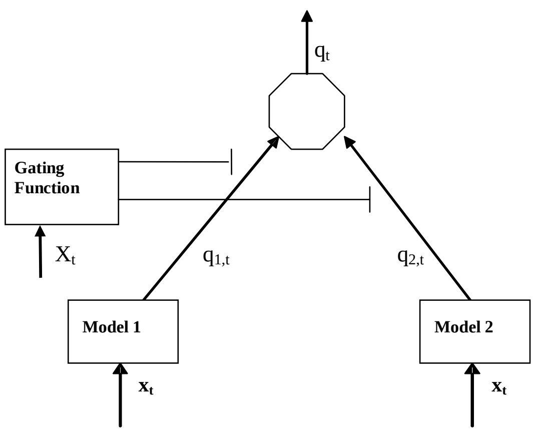

Figure 1. A Single Level HME Framework.

1. INTRODUCTION: THE PROBLEM OF MODEL STRUCTURE UNCERTAINTY

A range of hydrological models are available to simulate the runoff response to a rainfall event. Hydrological processes are very complex, with numerous associated variables. This reality, coupled with advances in computing power has fuelled the desire for many model developers to build models of ever increasing complexity. Despite this, the practicing hydrological community will tend to favour simpler models, often due to data constraints and the desire to apply a model with a wider range of applicability. As a result, hydrological models vary considerably in terms of their complexity and scope of application.

Despite the variety of available models, no single model has been identified as ideal over all possible situations and conditions. As a result, hydrological modellers must choose which model is best applied to a catchment, a task which can be problematic. It is generally desirable to choose a model that provides the best fit to the available data, but assessing the relative predictive performance of competing models can be difficult. Recently, it has been of increasing importance to allow for uncertainty on the model outputs, which can be significant considering the frequent lack of data and ever-increasing complexity of new hydrological models.

A number of approaches exist in the hydrological literature to incorporate model uncertainty in the modelling process. These include likelihood based methods such as GLUE (Beven and Binley 1992), or stochastic modelling approaches. In more applied studies, informal methods that put user-defined probability estimates on the model inputs or outputs are more popular. Uncertainty is generally expressed in terms of interval estimates, ensemble forecasts, or with predictions defined by probability distributions. In recent years, classical and hybrid Bayesian techniques have emerged as in ideal means of formally incorporating uncertainties in the hydrological modelling process, the end product being a probability distribution (the posterior distribution) on the model unknowns (parameters and model outputs) describing uncertainty after the data have been observed. Many studies have applied Bayesian techniques in hydrology. These have often focused on computational aspects (Kuczera and Parent 1998, Bates and Campbell 2001, Thiemann et al. 2001, Marshall et al. 2004), whilst others have explored how Bayesian methods can better characterise the sources of hydrological model uncertainty (Beven and Freer 2001, Kavetski et al. 2002).

Four major sources of uncertainty are well recognised in hydrological modelling (Butts et al. 2004) and Bayesian techniques explicitly incorporate these sources of uncertainty with varying success:

1. Parameter Uncertainty. The posterior distribution describes uncertainty about parameters and serves as a basis for selecting appropriate values for use in modelling applications. The advent of Markov Chain Monte Carlo (MCMC) methods has helped address some of the computational difficulties in summarizing and exploring the posterior distribution in hydrological modelling applications.

2. Calibration Data Errors. The data affects a parameter's posterior through the likelihood function. In applying different likelihoods, we are able to make assumptions about the statistical distribution of errors in the data.

3. Input Data Errors. Input data uncertainty can be incorporated in hydrological modelling under the addition of a model parameter, such as a rainfall depth multiplier (Lamb 1999) or random error term (Butts et al. 2004). Recently, likelihood based methods have been introduced that allow for input uncertainty in a Bayesian framework (Kavetski et al. 2002).

4. Model Structural Uncertainty. The impact of model structure assumptions and structural uncertainty has been paid less attention in the Bayesian hydrological literature. Nearly all studies make the assumption that the model structure is a reasonable approximation of the processes occurring. The structural uncertainty is usually incorporated via the distribution of the data noise assumed by the likelihood function. Biases arising from an incorrect structure are less frequently taken into account as it is assumed that the structural errors are random.

With the large number of variables involved in describing hydrological processes and the non-linearity of catchment mechanisms, hydrological models will never completely describe the processes occurring, especially given current data limitations. This means that one of the more significant sources of model uncertainty is that arising from the assumption of the (imperfect) model structure. There is no consensus about how to approach model structural error in hydrological modelling. Most studies are aimed at comparing model structures by changing the model itself, recalibrating and assessing the model performance based on some goodness of fit criterion.

2. IMPROVING PREDICTIVE PERFORMANCE BY COMBINING MODEL STRUCTURES

One way of reducing predictive uncertainty about the model structure is to incorporate the information of several different models at once. Each model provides different information about the process being considered so that overall the process mechanisms are better captured. The advantages of combining model predictions are well documented in many disciplines. In hydrology, established methods of combining models might include a simple or weighted average of models' results, or via the use of neural networks (Shamseldin et al. 1997, Georgakakos et al. 2004). Methods based on Bayesian statistical techniques also provide a way within which model outputs may be combined.

In comparing two models, the traditional Bayesian approach requires calculation of the Bayes factor, the odds of one model versus the other after observing the data. This logic can be extended to combining models. Say we wish to aggregate a set of models $M=\{M_1,\ldots M_n\}$, given data y for implementing the model. Let $p(M_i)$ be the prior probability of model Mi, and θ_i be the set of uncertain model parameters corresponding to model Mi. In Bayesian model averaging, the model outputs are weighted, with weights defined by:

$$p(M_i \mid y) = \frac{p(y \mid M_i)\,p(M_i)}{\sum_{j=1}^{n} p(y \mid M_j)\,p(M_j)} \qquad (1)$$

where $P(M_i \mid y)$ is the posterior probability of model Mi, $P(M_i)$ is the prior probability of model Mi, and $p(y \mid M_i)$ is the model's marginal likelihood.

Model weights as defined by Bayesian Model Averaging allow us to determine the probability for a particular model in comparison to others. We are able to compare different model structures whilst incorporating the uncertainty associated with the model outputs. However, the necessity of having fixed weights for each model over the entire length of the simulation period means that the relative usefulness of different models at different times is not considered. We can only estimate the overall performance ranking of individual model structures, not the times or conditions one model structure might be preferred to another. An appealing alternative may be to allow for the weights to change over time. This adaptation could depend on (say) applicable catchment antecedent soil moisture conditions.

3. HIERARCHICAL MIXTURES OF EXPERTS

Such an alternative is presented in a modelling framework known as hierarchical mixtures of experts or HME (Jordan and Jacobs 1994). HME models provide an improvement on simple combinations of models, by allowing the way that model predictions are combined to depend on predictor variables. HME models aim to combine the output from two or more models in a probabilistic sense. Each model configuration is adopted at a given time with a probability that depends on the current hydrologic state of the catchment. A HME approach to model development in hydrology gives greater flexibility to specification of the model structure (by allowing multiple model structures to exist in a single framework) and to the specification of model errors (by allowing different assumptions to apply depending on the input data or predictor variables).

The HME architecture is organized into a tree-like structure (Figure 1). The framework consists of individual model structures (known as experts or component models) that are grouped by nodes (known as gating functions). Figure 1 is the simplest HME framework, consisting of a single level and combining only 2 component models. The architecture shown may be expanded by recursively dividing the branches to include further levels or adding component models.

The HME networks should be considered probabilistically. The overall output is generated based on a probabilistic weighting of the output of each of the component models that is updated at each time step. The probability is based on current catchment indicators that are specified by the modeller to describe the state of the catchment. The respective probabilities of selecting each of the component models are estimated through use of the gating function, a mathematical function that is specified by the user.

The probabilistic nature of the model framework means that it is ideally specified using Bayesian inference. The difficulties in applying Bayesian techniques in to the HME framework in a hydrological setting lie in calculating the posterior distribution. Markov chain Monte Carlo (MCMC) is routinely used for estimating the posterior distribution in applied Bayesian statistics in complex problems. It has been shown to apply well to rainfall-runoff models and has previously been applied to the HME framework in hydrology (Marshall et al. 2005b). The HME framework presented here is specified via a mixture of Gibbs and Metropolis updating.

3.1. The HME Framework- Importance of Predictors

Implementing the HME approach requires specification of the probability of selecting each component model at each time step. This probability is calculated from two elements: a catchment descriptor that summarises the "state" of the catchment at a time step (the predictor) and the gating function that relates the probability of selecting a model to the predictor.

An important and indeed beneficial part of implementing the HME framework is assessing the effectiveness of different catchment predictors and gating functions in reproducing the switch from one component model to another. Each coupled component model and error model can be thought of as reproducing a different catchment "state", where different dominant hydrological processes are driving the catchment's response to rainfall. By assessing different predictors, modellers can interpret what physical processes are related to (or are forcing) the switch from one catchment "state" to another.

3.2. Interpretation of the Final HME Architecture vs. HME as a Predictive Tool

The HME approach has two main innovations for hydrological modellers:

1. Interpretation of the final model structure. In a hydrological setting, the HME approach can be an ideal tool for model building and for assessing individual model components and has been shown (Marshall et al. 2005b) to illustrate the conditions under which different model structures and model parameters are preferred .

2. Improvement of predictive performance. Given the difficulties in selecting a single model structure for predicting streamflow, HME provides a more sophisticated way to combine models than achieved by simple or weighted averages of model outputs. Achieving a model performance in prediction that is better than that of the individual component models can be hard. The HME approach shows an overwhelming improvement to the model in the calibration period (Marshall et al., 2005b). This is due to conditioning the probability of selecting one component model on the observed calibration data. In predictive mode, it must be assumed that this data is not available (indeed it is what the model is seeking to reproduce). The predictive ability of the HME approach is determined via investigation of different catchment predictors used to weight the individual models.

3.3. Case Study

The HME framework was applied in a case study to assess the predictive performance of different catchment descriptors and gating functions. It was desirable to attain a predictive performance from a combination of models that was better than a single model (given the increase in model complexity). This requires careful selection of appropriate catchment predictors.

The selected study area was the Never Never River at Glenniffer Bridge, a 51 km^2 catchment located in New South Wales with annual rainfall of 2036mm and runoff of 1114mm. Ten years of daily rainfall and runoff data were used for calibration in the study.

The simplest HME architecture was selected, consisting of only 2 component models and a single level (Figure 1). Based on earlier studies (Marshall et al., 2005b) it was observed that a range of catchments were well modelled as a mix of only two states. These results motivated the desire to show the method's predictive performance for a catchment modelled by two component models.

Unlike earlier studies, different error models were specified for each component model. It was observed that when conditioned on the observed data, one component model would tend to fit to the model peaks, with the other model fitting the recession curve. It is also recognised in a number of hydrological studies that model errors often exhibit heteroscedasticity proportional to flow magnitude (Sorooshian and Dracup 1980). Based on these results, it was likely that the component model fitting the peak flows would have a greater variance. Use of separate error models would allow better justification of the assumptions made on the distribution of model errors.

Each of the component models was specified to have the same model structure, but was calibrated to have different parameter values. Both components were set to be the simplified 3-parameter Australian Water Balance Model

(AWBM, Boughton, 2004). The model (Figure 2) consists of 3 parameters: S (surface store capacity), K (recession constant) and BFI (base-flow index), and uses a soil moisture accounting process to generate streamflow from daily rainfall and evapotranspiration data.

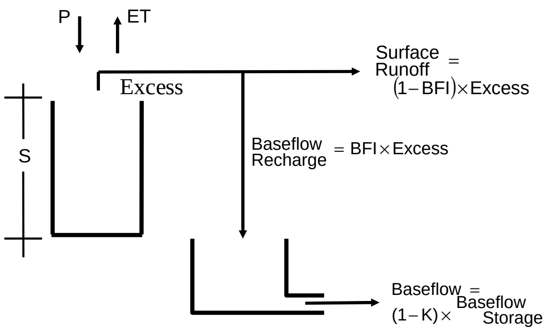

Figure 2. The Australian Water Balance Model

The distribution of model parameters was determined using a MCMC algorithm that was a modified version of the Metropolis algorithm. The component model parameters and gating function coefficients were sampled in separate blocks using a multivariate normal proposal distribution in which the covariance had been pre-tuned.

The popular logistic regression model was used for the gating function of the form:

$$g_{t,1} = \frac{e^{g(X_t,\beta)}}{1+e^{g(X_t,\beta)}} \quad (2)$$

where $g_{t,1}$ is the probability of selecting component model 1, X_t are the catchment predictors and β is the vector of logistic regression parameters. Different options for the function $g(X_t,\beta)_{t,1}$ were implemented in the study. A simple linear regression function was used of the form:

$$g(X_t,\beta)_{t,1} = \beta_0 + \beta_1 X_t \quad (3)$$

This was also extended to a polynomial regression function:

$$g(X_t,\beta)_{t,1} = \beta_0 + \beta_1 X_t + \beta_2 X_t^2 \quad (4)$$

A spline was also implemented of the form:

$$g(X_t,\beta)_{t,1} = \sum_{m=1}^{5} \beta_m h_m(X_t) \quad (5)$$

having continuous linear basis functions $h_m(X_t)$ and three knots. The knots were intuitively set at the 25th, 50th and 75th percentiles of the range of predictor values.

The performance of different catchment predictors is compared using an appropriate criterion. Comparison of models using Bayes Factors in previous hydrological studies (Marshall et al. 2005a) has effectively compared models, but has high computational demands. The Bayesian Information Criterion (BIC) of Schwarz (1978) is an asymptotic approximator of the marginal likelihood of a model. The criterion holds that the model log-marginal likelihood is approximately -0.5 BIC, where (if N is the size of the sample):

BIC = -2(log maximized likelihood) + (log N)(number of parameters) (6)

4. RESULTS AND DISCUSSION

The suitability of the HME framework for prediction was assessed in comparison to a single model structure. The resulting BIC for different gating functions and predictors are given in Table 1. A ‘null’ predictor was first implemented, so that the probability of selecting each model does not change in time. Note that this produces a BIC value that is worse than a single component model. Hence, a simple weighted average of the two component models does not give a better prediction than that from a single model. In a preceding related study (Marshall et al. 2005b) it was observed that the parameter BFI was most sensitive in describing the switch from one catchment state to another. This parameter describes the proportion of excess runoff returning to the base storage component of the model. Hence, the modelled base storage (obtained as a weighted average of the two models) was initially selected to describe the catchment ‘state’ when determining the probability for each model.

Table 1. Comparison of HME Predictors

Predictor	Gating Function	-0.5BIC
Single Component Model	N/A	-11233
Null Predictor	Linear Logistic	-11537
Modelled Base storage	Linear Logistic	-10209
Log-Antecedent Cumulative Rainfall (7 days preceding)	Linear Logistic	-8482
	Polynomial Logistic	-8480
	Spline Logistic	-8212

Table 2. HME and AWBM model parameters, estimated as the mean posterior value.
The bracketed values give the 90% posterior intervals for the parameters.

AWBM			HME Component 1			HME Component 2		
K	BFI	S	K	BFI	S	K	BFI	S
0.930 (0.92- 0.94)	0.518 (0.49- 0.54)	149 (139- 188)	0.929 (0.90- 0.95)	0.506 (0.47- 0.54)	175.8 (139-196)	0.974 (097- 0.98)	0.293 (0.29- 0.30)	37.7 (36- 42)

When different predictors are introduced, the model performance improves. The antecedent rainfall proves to be the best predictor, with a spline gating function.

Calibration of the component models showed the existence of two distinct states that were best modelled by different parameterisations of the AWBM model. Note the detailed results in Table 2 and the implications of the parameter values for each HME component in reference to the AWBM structure illustrated in Figure 2. A high BFI (HME component 1) implies more rain is stored in the baseflow storage, and correspondingly a small fraction of the rainfall enters the stream as direct runoff. A low BFI (HME component 2) implies the reverse. It was observed illustrated that the "quick-flow" process (HME component 2) dominates at the high flow periods and early recession.

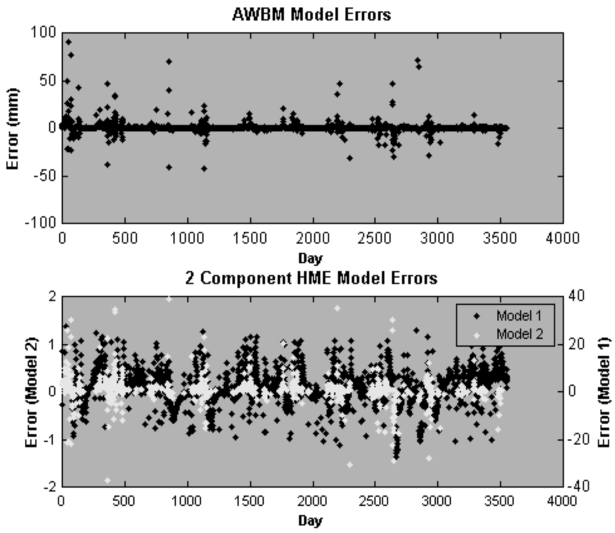

Figure 3. Distribution of (a) AWBM Errors (b) HME Errors

Table 2 also indicates the 90% posterior intervals for the model parameters. For each of the parameters in the model, diffuse prior distributions were defined (in the interval 0 – 1 for the parameters K and BFI and the interval 0 – 1500 for the parameter S). The posterior intervals show that the estimated parameters are not largely influenced by these priors. It must also be noted here that for each component model the posterior intervals for the parameter BFI do not coincide. Two distinct models are observed.

It is important to note the characteristics of the errors of the fitted model. If we fit a single (AWBM) component model using a likelihood function that assumes independent, identical normally distributed errors, Figure 3 is the resulting error plot. Note the dominance of low errors, with few high errors. The errors are not well summarised by a normal distribution and are highly heteroscedastic, hence the assumptions taken when applying the error model are not well satisfied. When we model the errors by a combination of two structures, the approach allows greater flexibility in the specification of the model errors. Rather than using a multiobjective approach (where the entire length of the data is used to specify each objective), the data is divided and different assumptions about the distribution of the model errors in different sections of the data may be made. We can allow the structure of errors to change depending on which part of the catchment's response we are modelling. The errors are now better summarised by a normal distribution, and the individual error structures do not show a dominance of very low values.

5. CONCLUSION AND FUTURE WORK

In hydrological modeling, using a single model with a rigid model structure can lead to significant bias, as evidence exists of the catchment responding differently under certain antecedent conditions. To form a better prediction of catchment behaviour than would be provided by a single model, a model can be approximated through the combination of a number of different modelling configurations. Each model is adopted at a given time with a probability that depends on the current hydrologic state of the catchment. This framework is known as a Hierarchical Mixture of Experts (HME).

Application of the HME framework shows that the catchment is well modelled as two different "states", rather than by a single static model. The two HME component models corresponded to different catchment mechanisms: a high recharge state where the baseflow storage is increasing, and a low recharge state in low flow times.

The challenge in applying the HME framework for predictive purposes lies in determining which model should be selected depending on the state of the catchment. Estimating this probability is reliant on choosing appropriate variables to characterise the catchment state, and a mathematical function that can relate the predictor to the dominant model. In this study several different gating functions and predictors were investigated and the performances of these were compared via the BIC, a Bayesian-like comparison criterion. Results showed that by comparing different predictors, the modeller can assess which variables are most likely forcing or related to a 'switch' in the catchment state. By selecting different predictors, the HME framework can give a better simulation than from a single model (taking into account the increase in model complexity). The cumulative antecedent rainfall (a measure of the catchment's wetness) proved to be the most appropriate predictor.

Much of the current literature concerning HME is interested in finding the optimum topology of the network's architecture. Although this may prove a promising way to extend the simple architecture used in this study, computational difficulties will likely arise and there is a desire to keep the model parsimonious in hydrological applications. By dividing the calibration data space, the individual component models may become over-identified. This is of particular importance in hydrological modelling, where there may be insufficient data available to identify each model.

6. REFERENCES

Bates, B. C., and E. Campbell (2001), A Markov chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall-runoff modeling, *Water Resources Research* 37, 937-947.

Beven, K., and A. M. Binley (1992), The future of distributed models: Model calibration and uncertainty prediction, *Hydrological Processes* 6, 279-298.

Beven, K., and J. Freer (2001), Equifinality, data assimilation, and uncertainty estimation in mechanistic modeling of complex environmental systems using the GLUE methodology. *Journal of Hydrology*, 249, 11-29.

Boughton, W. C. (2004), The Australian Water Balance Model. *Environmental Modelling & Software*, 19, 943-956.

Butts, M. B., J. T. Payne, M. Kristensen, and H. Madsen (2004), An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation, *Journal of Hydrology*, 298, 242-266.

Georgakakos, K. P., D. Seo, H. Gupta, J. Schaake, and M. B. Butts (2004), Characterising streamflow simulation uncertainty through multimodel ensembles, *Journal of Hydrology*, 298, 222-241.

Jordan, M. I., and R. A. Jacobs (1994) Hierarchical mixture of experts and the EM algorithm, *Neural Computation*, 6, 181-214.

Kavetski, D., S. W. Franks, and G. Kuczera (2002), Confronting input uncertainty in environmental modelling. Pages 49-68 in Q. Duan, H. Gupta, S. Sorooshian, A. N. Rousseau, and R. Turcotte, editors. *Calibration of Watershed Models*, AGU Water Science and Applications Series.

Kuczera, G., and E. Parent (1998), Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm, *Journal of Hydrology*, 211, 69-85.

Lamb, R. (1999), Calibration of a conceptual rainfall-runoff model for flood estimation by continuous simulation, *Water Resources Research*, 35, 3103-3114.

Marshall, L., D. J. Nott, and A. Sharma. 2004. A comparative study of Markov chain Monte Carlo methods for conceptual rainfall-runoff modeling, *Water Resources Research*,40,1-11.

Marshall, L., D. J. Nott, and A. Sharma (2005a), Hydrological model selection: a Bayesian alternative, *Water Resources Research*, In press.

Marshall, L., D. J. Nott, and A. Sharma (2005b) Towards dynamic catchment modelling: a Bayesian Hierarchical Mixtures of Experts framework, *Hydrological Processes*, Under review.

Schwarz, G. (1978), Estimating the dimension of a model, *The Annals of Statistics*, 6, 461-464.

Shamseldin, A. Y., K. M. O'Connor, and G. C. Liang (1997), Methods for combining the outputs of different rainfall-runoff models, *Journal of Hydrology*, 197, 203-229.

Sorooshian, S., and J. A. Dracup (1980), Stochastic parameter estimation procedures for hydrologic rainfall-runoff models: correlated and heteroscedastic error cases, *Water Resources Research*, 16, 430-442.

Thiemann, M., M. Trosset, H. Gupta, and S. Sorooshian (2001), Bayesian recursive parameter estimation for hydrologic models, *Water Resources Research*, 37, 2521-2535.