

IMPROVING STREAMFLOW SIMULATIONS AND FLOOD FORECASTS WITH MULTIMODEL ENSEMBLES

MICHAEL B. BUTTS, JEFFREY T. PAYNE, JESPER OVERGAARD

*DHI Water & Environment, Agern Allé 5,
Horsholm, DK 2970 Denmark*

Methods for improving the hydrological simulation through the use of multi-model ensembles (MME) are demonstrated. In recent years, the meteorological community has exploited several MME combination techniques as a means for improving short-term weather and seasonal climate forecasts. Within the hydrological community, little work has been carried out to explore the benefits of MMEs for streamflow simulations. This study examines the use of MMEs for improving streamflow simulation, including their potential benefits to flood simulation. Multimodel ensembles are generated using ten distinct model structures derived using a new hydrological modelling tool. These model structures were applied to a US NWS test catchment, the Blue River and evaluated using a split-sample procedure. The resulting ensemble can be used to make probabilistic simulations that characterise model structure uncertainty. Furthermore it is shown that the ensemble average of all 10 models performs better than any single model in split sample test. Using regression methods, improvements in ensemble simulations using linear combinations of the ensemble members were explored. The performance of the resulting weighted ensemble is similar to the simple ensemble average but uses a smaller ensemble. This may provide a means to identify which model structures provide significant contributions to accurate hydrological simulation.

INTRODUCTION

Any evaluation of the accuracy of deterministic hydrological models for hydrological simulations and forecasting must address four sources of uncertainty. They are the random or systematic errors in the model inputs or boundary condition data, random or systematic errors in the recorded output data, uncertainty due to sub-optimal parameter values and errors due to incomplete or biased model structure. A critical choice in applying mathematical models to hydrological problems is the selection of model structure. Selecting an appropriate model structure encompasses determining the most significant hydrological processes to be described and the mathematical description of these processes.

While the selection of model structure is crucial in determining the reliability and accuracy of the resulting model very few studies have examined the impact of model structure uncertainty in the context of simulation uncertainty. In most practical engineering problems, hydrologists attempt to identify the important processes and select a model system accordingly. Once this selection is made, further efforts to examine the impact of adopting different model structures are seldom made. Instead emphasis is

usually placed on evaluating the impact of parametric uncertainty. Very few studies have examined in a systematic way the significance of model structure uncertainty in relation to the other sources of uncertainty. However some recent studies (Georgakakos et al. [5], Butts et al. [2]) suggest that model structure uncertainty is at least as important as the parametric uncertainty.

The essential difficulty for hydrological practitioners is that most current modelling tools provide very limited options for exploring different model structures. The alternative of applying a number of different models is impractical in many cases. Furthermore there is a need to match the model structure to the level of complexity to the particular application, the available data sources and the accuracy requirements. To address these requirements a new hydrological modelling tool has been developed, Butts et al. [2]. The idea behind this tool is to allow the modeller to select which processes require modelling, the level of conceptualisation, the degree of lumping and hence the level of complexity that is necessary for a particular application. The tool provides a process-based description of the rainfall-runoff and channel routing based on the MIKE SHE and MIKE 11 hydrological modelling concepts. However the MIKE SHE modelling paradigm of Freeze and Harlan [3] is extended to include lumped and conceptual descriptions to supplement the original distributed physics-based flow equations, Butts et al. [2].

In this paper this tool is used to develop multimodel ensembles and to examine the potential benefits of applying ensemble-based simulations for hydrological applications and forecasting. Meteorologists have applied ensemble modelling using a number of different models or model structures to both long-term predictions (e.g. WMO [8]) and short-term operational forecasts (Krishnamurti et al. [6]). Ensemble modelling can also be used to provide probabilistic simulations from deterministic models that can be used in an evaluation of the variability or uncertainty in their predictions. In the hydrological community ensemble flow forecasting has focussed on investigating the effect of parametric and input error using several realisations from single models, Beven and Freer [1], Georgakakos [4]. However few systematic studies of the impact of model structure uncertainty on simulation uncertainty have been carried out. Furthermore there is a need within hydrological modelling to examine the benefit of using ensemble averages to provide consensus forecasts and to improve simulation accuracy.

STUDY AREA

The study area used here is one of the test basins within the Distributed Modeling Intercomparison Project (DMIP) organised by the Hydrology Lab of the National Weather Service (NWS). The purpose of the DMIP study was to evaluate the capabilities of existing distributed models and identify avenues for model improvements, <http://www.nws.noaa.gov/oh/hrl/dmip/>. We have chosen here to focus on the 1232 km² Blue river basin in Oklahoma, Figure 1. It is of interest for distributed hydrological

modelling because of its unusual aspect ratio, soil variability and the availability of distributed radar-based rainfall.

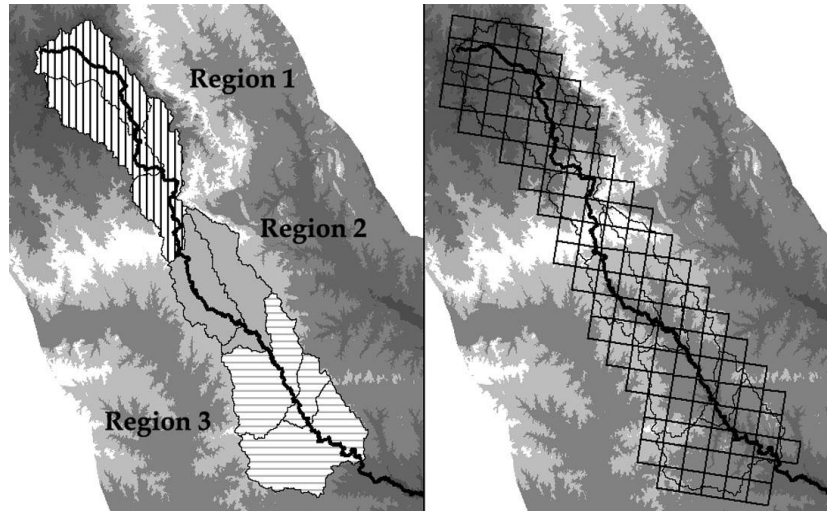


Figure 1. Spatial discretization used in this study for the Blue River basin. The figure on the left shows the 8 subcatchments used in conceptual modelling and the parameter regions used for calibration. The figure on the right shows the 4-km NEXRAD grid used for the grid-based modelling.

The Blue river basin is located in south-central Oklahoma and flows into the Red River at the Texas-Oklahoma border. The watershed is semi-arid, with significant convective rainfall events. Distributed rainfall data is available in the form of NEXRAD gridded data provided at hourly intervals at a spatial resolution of 4 km by 4 km. The new modelling tool was used, firstly to test the performance of the different model structures and secondly to evaluate the model structure uncertainty compared to other sources of uncertainty, Butts et al. [2]. To this end 10 different model structures were identified as plausible model structures. The different model structures included both lumped and distributed routing, lumped, subcatchment-based and distributed rainfall-runoff models, grid-based modelling using physics-based flow equations, different conceptual process descriptions and lumped, subcatchment-based and gridded radar-rainfall input, Table 1.

To evaluate their performance each model structure was fitted to data in the calibration period and then evaluated against measurements in the validation period. Each model structure was fitted to the calibration period using automatic parameter estimations based on the Shuffled Complex Evolution and multiple objectives. (Madsen [7], Butts et al, [2]). The average error and RMSE were used as the calibration objectives. The split sample testing approach used here can then be used as a means of selecting the best model structure for the purpose – in this case flood forecasting.

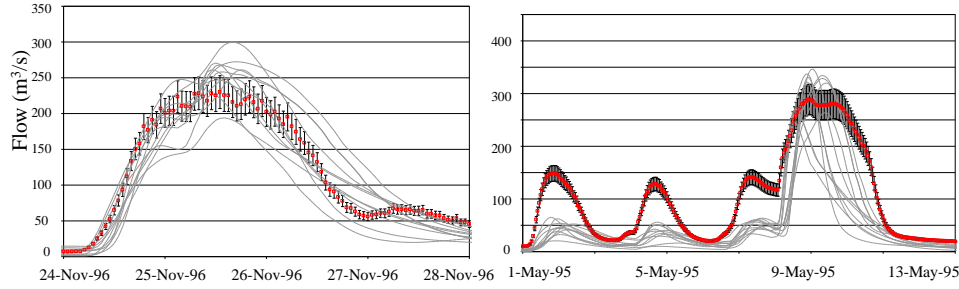


Figure 2. Examples of the model structure ensemble (gray lines) and the estimated measurement uncertainty (error bars) for two major peaks, one in the calibration period and one in the validation period.

Table 1. Short description of model structures used in this study.

ID	Short Name	Description
s1	Lumped	Completely lumped using the MIKE 11 NAM model
s2	Distributed routing	Completely lumped with fully dynamic distributed routing, using the MIKE 11 NAM model concept.
s3	Muskingum	Sub-basin distributed rainfall with Muskingum-Cunge routing, using the MIKE 11 NAM model concept.
s4	Distributed rainfall	Sub-basin distributed rainfall with fully dynamic routing, using the MIKE 11 NAM model concept.
s5	3 regions	Sub-basin distributed rainfall with fully dynamic routing, with 3 independent parameter regions, using the MIKE 11 NAM model concept.
g1	Aggregated rainfall	Sub-basin distributed rainfall using the fully distributed grid-based MIKE SHE model concept
g2	Gridded rainfall	4 km NEXRAD gridded rainfall using the grid-based fully distributed MIKE SHE model concept.
g3	No drains	Excluding drain flow in g1.
g4	Linear reservoir	Sub-basin distributed rainfall using the grid-based fully distributed MIKE SHE model concept for surface and unsaturated soil processes and a semi-distributed model conceptual model for the sub-surface processes.
g5	Bypass infiltration	A bypass conceptual model for rapid infiltration in the unsaturated zone used in g4.

MULTIMODEL ENSEMBLES

The simulations using these calibrated model structures form an ensemble, albeit small, of model simulations. This ensemble can be viewed as a measure of the uncertainty arising from model structure. Figure 2 shows the ensemble simulations for two major peaks. Significant variability is observed amongst the different model structures even though they were calibrated on the same data to the same objectives. The purpose of this paper is to explore the use of the ensemble to improve hydrological simulations. The simplest approach is to use simple averages of the different models. Figure 3 shows the simulations made using two different model structures, denoted s3 & g1, (see table 1). Using a simple average of these two models apparently provides a much better simulation of the flood hydrograph. Taking this one step further and using a weighted linear combination of the two models, a further improvement in simulated hydrograph can be obtained.

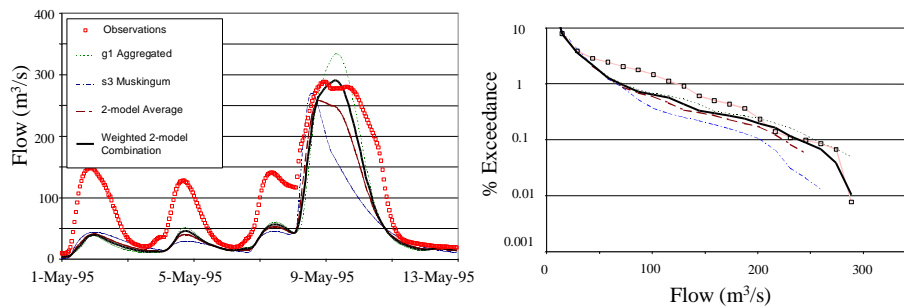


Figure 3 Simulated hydrograph and flow duration curves for a simple multimodel ensemble with two members, g1 and s3.

It is also of interest to determine the ability of the hydrological model to represent the full range of flows occurring within the catchment. One way to examine this is to compare the flow duration curves predicted by the model with the measured flow duration curve, Figure 3. We found that while the performance of the s3 model is better than the g1 model structure in terms of the RMSE (see table 2.), the g1 model better captures the range of flows within the catchment. It appears that neither model is able to accurately represent the distribution of intermediate flows, (50 –150 m³/s). For the higher flows, which were the focus of the DMIP study, it appears that a much better representation of the distribution of flows is obtained using a weighted mean of the two model structures, Figure 3.

From this simple analysis it appears that there is some potential for improving the accuracy of hydrological simulations using ensemble averages and linear combinations of model structures from an ensemble. It can be argued that different model structures represent different aspects of the flow hydrographs, such as the magnitude and width of the peak and the rising and falling limbs of the flood hydrograph. Combining these

different structures could be expected to improve the accuracy of hydrological simulations when compared to a single model structure.

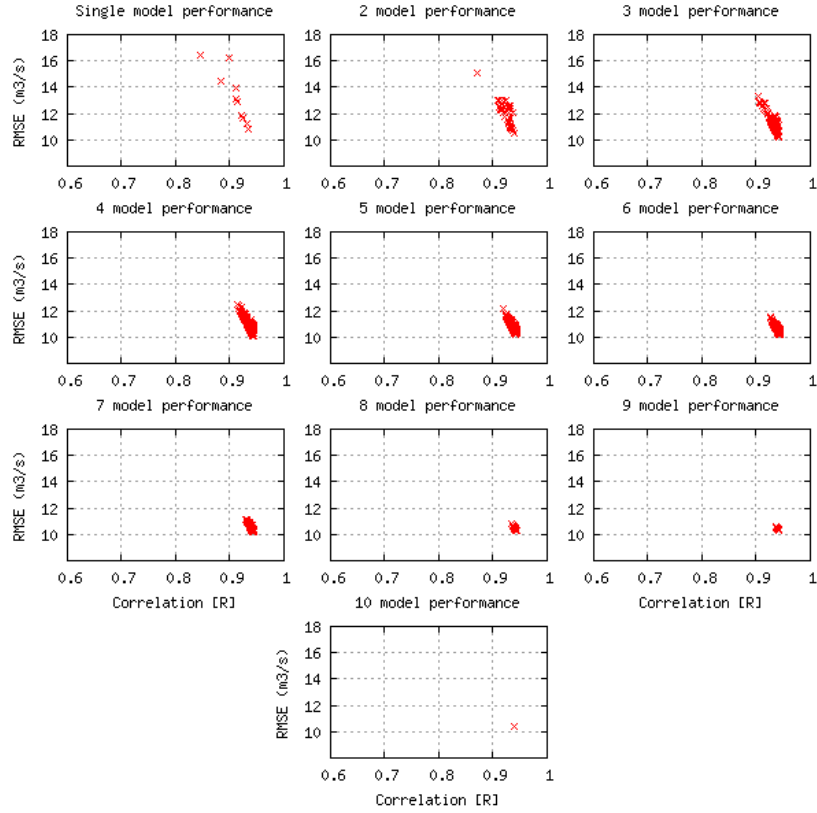


Figure 4 Performance (RMSE and correlation R) of simple averages of multi-model ensembles for the calibration period.

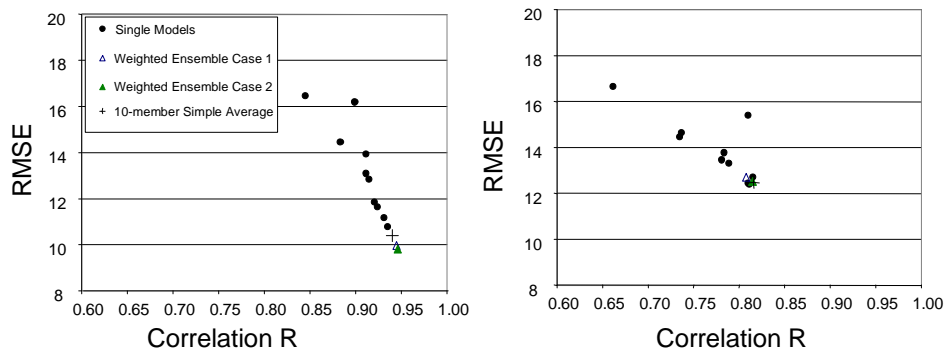


Figure 5 Performance measures for the individual ensemble models and ensemble averages using different weighting schemes in the calibration and validation periods.

Table 2. Performance measures for the individual model structures. The shading indicates the best performance.

	s1	s2	s3	s4	s5	g1	g2	g3	g4	g5
RMSE (m ³ /s) Calibration	16.1	11.8	11.2	11.7	10.8	12.9	13.1	14.0	16.5	14.5
RMSE (m ³ /s) Validation	15.4	14.4	12.7	13.3	13.5	12.5	12.4	13.8	16.7	14.7
Correlation R Calibration	0.90	0.92	0.93	0.92	0.93	0.91	0.91	0.91	0.84	0.78
Correlation R Validation	0.81	0.73	0.81	0.79	0.78	0.81	0.81	0.78	0.66	0.74

Figure 4 shows the performance (RMSE and correlation R) of simple averages of different combinations of 2,3,4, ... 10 models. It appears from this figure that for any number of ensemble members there exists combinations that perform better than the individual models. It was also found that the individual models that performed best in the calibration period did not necessarily perform best in the validation period, table 2. By contrast the 10-model ensemble average performs well in both the calibration and validation period, Figure 5.

Since the simple average of the 10-model ensemble improves the accuracy of flow simulation, two interesting questions are raised. Firstly is there an ensemble consisting of a smaller number of members, that outperforms the individual models in both the calibration and validation period? Secondly it is possible to further improve the simulation accuracy by using a linear combination of the ensemble members? To answer these questions, linear combinations of the 10 ensemble members were investigated. Optimisation methods were used to find the “best” linear combination of model structures from the amongst the 10 model structures during the calibration period. The “optimised” ensemble time series were found using Marquardt-Levenberg optimization and penalty functions to ensure the weights lie within the selected range. Since only non-negative weights were permitted non-linear regression is required. Two cases were investigated, the first (Case 1.) assuming that the sum of the weights is equal to unity and the second (Case 2.) without this restriction. This second approach was evaluated as it expected that this might correct for any biases in the model structures. In both cases an ensemble of 7 members was found which provided similar accuracy to the simple ensemble average of all 10 member during the validation period, Figure 5.

CONCLUSIONS

A new modelling framework is used to generate multimodel ensembles using different calibrated model structures. The resulting ensemble can be used to make probabilistic forecasts that characterise model structure uncertainty. The performances of different ensemble averages were evaluated using RMSE, correlation and flow duration curves. We found that a simple ensemble average of the 10 model structures used performs better

than any single model in the split sample test. One explanation for this is that different model structures better represent different parts of the catchment flow response to rainfall. This suggests that ensembles derived from different models provide an opportunity to improve hydrological simulation and forecasts that deserves further investigation. Using regression methods it was found that similar performance levels could be achieved with a subset of model structures which would reduce computation time in real-time applications. This could also provide a means to identify which model structures provide significant contributions to accurate hydrological simulation.

ACKNOWLEDGEMENTS

This work was in part funded by the EU 5th framework project FLOODRELIEF. The second author would like to acknowledge the support provided through the University of Washington's Valle Exchange and Scholarship Program.

REFERENCES

- [1] Beven K.J. and J. Freer Equifinality, "Data Assimilation and Uncertainty Estimation in Mechanistic Modelling of Complex Environmental Systems Using the GLUE Methodology". *Journal of Hydrology* Vol. 249, (2001), pp 11-29
- [2] Butts M., J. T. Payne, M. Kristensen and H. Madsen. "An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow predictions", *Journal of Hydrology*, DMIP Special Issue. 2004 (Submitted).
- [3] Freeze, R.A. and Harlan, R.L., 1969, "Blueprint for a physically-based digitally-simulated hydrological response model", *Journal of Hydrology*, Vol. 9, (1969), pp237-258.
- [4] Georgakakos, K.P., "Hydrometeorological Models for Real Time Rainfall and Flow Forecasting". Chapter 18 in *Mathematical Models of Small Watershed Hydrology and Applications*, V.P. Singh and D.K. Frevert, eds., Water Resources Publications, LLC, (2002), pp 593-655.
- [5] Georgakakos K.P., Seo, D-J, Gupta , H, Schaake, J. and Butts, M.B., " Characterising streamflow simulation uncertainty through multimodel ensembles", *Journal of Hydrology*, DMIP Special Issue 2004 (Submitted)
- [6] Krishnamurti, T.N, C.M. Kishtawal, T. T. E. LaRow, D. R. Bachiochi, Z Zhang, C. E. Williford, S Gadgil, S Surendran. "Improved Weather and Seasonal Climate Forecasts from Multimodel Superensemble", *Science* Vol. 285, (1999), pp 1548-1550.
- [7] Madsen, H. "Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives", *Adv. Water Resources*, Vol. 26, (2003), pp 205-216.
- [8] World Meteorological Organisation WMO, "Report on the Operational Use of EPS to Forecast Severe Weather and Extreme Events". *WMO Commission on Basic Systems, Meeting of Expert team on Ensemble Prediction Systems*, Tokyo, Japan, 15-19 October 2001, CBS ET/EPS/Doc. 3(7) (9.X. 2001), Geneva Switzerland, (2001), 7pp.