2 Model Evaluation and Comparison: Uncertainty Analysis and Diagnostics

The commitment to the quantification and subsequent reduction of uncertainty in hydrological flux predictions lies at the heart of the PUB initiative (Sivapalan *et al.*, 2003; Wagener *et al.*, 2004). An uncertainty framework in which models, data sources and methods can be tested is needed to fully implement this effort. Two fundamental science questions are central to this effort: (1) How can we (explicitly) estimate and propagate all sources of uncertainty in hydrological modelling? (2) What is an appropriate framework for (model/method) evaluation under uncertainty?

Recent research suggests that the use of an inadequate model structure can be even more problematic than the use of sub-optimal parameter values in many hydrological modelling studies. Additionally, it will often have a considerable affect on predictive uncertainty. However, despite these results, model structural uncertainty has largely been ignored in uncertainty analysis and it is common to adopt a single model structure with little justification for this selection. The PUB session in which the following papers were presented focused on the evaluation of hydrological and environmental model uncertainties (Theme 3). Papers related to theory and tools for evaluation and diagnostics were presented.

There is an urgent need to improve the theory and techniques used in model comparison and evaluation. Examples of current science questions, which require answering to achieve the overall objective of a suitable framework, are: How can new or innovative data sources (e.g. soft data or tracers) be used in model comparison and evaluation? What is the effect of differences in scale between observed and modelled variables when these are used for model evaluation? How far do internal model states relate to real world variables? How can model structural uncertainty be considered when estimating predictive uncertainty? What evaluation frameworks provide feed-back regarding potential model structural improvement? How can we evaluate the "realism" of a model structure?

The papers in this section were selected from the contributions to the Theme 3 session and cover three main areas:

(1) Uncertainty in the ***observations*** of input and output variables and their impact on model uncertainty (Hughes; Post & Hartcher; Arheimer; Rode & Wriedt; Aronica *et al.*; Xavier *et al.*; Chiang *et al.*).
(2) The evaluation and reduction of ***model structure*** uncertainty (Koren; Ren *et al.*; Wyatt & Franks; Mo *et al.*).
(3) The estimation and propagation of uncertainties into model ***predictions*** for different applications (Romanowicz *et al.*).

The first group of papers highlights the problems of attempting the modelling of natural systems with data that are insufficient in quality or quantity. The studies by Post & Hartcher, and by Hughes demonstrate very clearly the limitations of modelling studies in less developed countries. While remotely sensed information can bring some relief, ground truthing is still required to provide reliable estimates of important variables such as ground cover factors in the sediment transport study of Post & Hartcher. Hughes concludes that improvements with respect to model structures could be irrelevant if better forcing (precipitation) data is not available in Southern Africa. Similarly, Arheimer's study demonstrates that reliable predictions are infeasible if representative rainfall is lacking. Xavier *et al.* tried to address this problem by combining different sources of precipitation observations (radar and raingauges) using kriging in a Generalized Likelihood Uncertainty Estimation (GLUE; Beven & Binley, 1992) framework. Finally Chiang *et al.* define different quantitative performance measures to assess the impact of increasing precipitation uncertainty on the model output.

The studies by Rode & Wriedt, and Aronica *et al.* on the other hand, evaluate the use of observations of the system response during model calibration. Aronica *et al.* tested the impact of rating curve knowledge on GLUE predictions of a rainfall–runoff model and found that lack of this knowledge can have serious impacts on prediction uncertainty, particularly if the rating curves are overestimated. Rode & Wriedt tested the value of data in a multi-objective framework used to calibrate a water quality model and tried to identify necessary amounts of observations for their situation.

The second group contains four papers that emphasize the evaluation or improvement of model structures. The study by Koren demonstrates that processes which are often neglected in hydrological modelling can have significant impacts on the result. Koren introduced a physically-based parameterization of seasonally frozen soil into a lumped conceptual rainfall–runoff model (Sacramento). This addition led to significant improvements in the model's performance with respect to reproducing spring and summer flood events. Ren *et al.* tested the performance of a selected conceptual model (Xinanjiang) in reproducing soil moisture in addition to the prediction of streamflow only. They discovered that a relatively high correlation between soil moisture observations and the model's soil tension water storage state could be found in their case. This paper is an interesting contribution to the ongoing discussion on how realistic conceptual models represent features of hydrological systems. Wyatt & Franks present a multi-model approach in which a wide variety of lumped model structures can be implemented from a library of process descriptions and used simultaneously. Within a modified GLUE framework, they tested two million (uniform) randomly sampled parameter sets for each of the 45 model permutations contained. Wyatt & Franks also found that utilizing multiple (synthetic) observations of fractional saturated area for model conditioning considerably reduced the runoff prediction uncertainty envelopes for an extreme event period, regardless of the selected model structure. Mo *et al.* performed a detailed analysis of a land surface model using multiple output variables (states and fluxes), again in a GLUE framework. They found that, even though the multi-objective approach is very efficient in constraining the model space, a serious equifinality problem remained. This is particularly problematic since they also found little interaction between the model parameters.

Romanowicz *et al.* present an interesting study in which they calculate the propagation of uncertainty in a sequential, multiple-step-ahead flood forecasting system based on stochastic transfer functions in connection with the GLUE algorithm. They found that model structural uncertainty is of particular importance in achieving high prediction performance (defined as lead times in their study).

In conclusion it is interesting to note that many of the authors of the above discussed papers opted for using the GLUE procedure to estimate uncertainty (Aronica *et al.*; Mo *et al.*; Romanowicz *et al.*; Xavier *et al.*; Wyatt & Franks). While there might be different reasons for the adoption of this approach, one is clearly the ease of its implementation. The adoption of other approaches often improves computational efficiency, but also includes additional assumptions about error distributions that might be difficult to test or justify (see discussion in Beven & Freer, 2001).

REFERENCES

Beven, K. J. & Binley, A. M. (1992) The future of distributed models: model calibration and uncertainty prediction. *Hydrol. Processes* **6**, 279–298.

Beven, K. J. & Freer, J. (2001) Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *J. Hydrol* **249**(1–4), 11–29.

Sivapalan, M., Takeuchi, K., Franks, S., Gupta, V. K., Karambiri, H., Lakshmi, V., Liang, X., McDonnell, J. J., Mendiondo, E. M., O'Connell, P. E., Oki, T., Pomeroy, J. W., Schertzer, D., Uhlenbrook, S. & Zehe, E. (2003) IAHS Decade on Predictions in Ungauged Basins (PUB), 2003–2012: Shaping an exciting future for the hydrological sciences. *Hydrol. Sci. J.* **48**(6), 857–880.

Wagener, T., Sivapalan, M., McDonnell, J. J., Hooper, R., Lakshmi, V., Liang, X. & Kumar, P. (2004) Predictions in Ungauged Basins (PUB)—A catalyst for multi-disciplinary hydrology. *Eos, Trans. AGU* **85**(44), 451–452.

Thorsten Wagener
Department of Civil and Environmental Engineering,
The Pennsylvania State University, 226B Sackett Bldg.,
University Park, Pennsylvania 16802, USA
thorsten@engr.psu.edu