18th World IMACS / MODSIM Congress, Cairns, Australia 13-17 July 2009
http://mssanz.org.au/modsim09

Ensemble predictions of hydro-biogeochemical fluxes at the landscape scale

Exbrayat, J.-F.[1], Breuer, L.[1], Viney, N.R.[2], Seibert, J.[3], Wrede, S.[4], and Frede, H.-G.[1]

[1] *Institute for Landscape Ecology and Resources Management, Justus-Liebig-University Giessen, Heinrich-Buff-Ring 26, 35392 Giessen, Germany*

[2] *CSIRO Land and Water, Canberra, Australia*

[3] *University of Zurich, Switzerland, and Stockholm University, Stockholm, Sweden*

[4] *Delft University of Technology, Delft, Netherlands*

Email: jean-francois.exbrayat@umwelt.uni-giessen.de

Abstract: Model predictions of biogeochemical fluxes on the landscape scale are highly uncertain, both with respect to stochastic (parameter) and structural uncertainty. The idea of our ensemble modelling approach is to reduce the predictive uncertainty by covering part of the parameter and model structural uncertainty. In this study 4 different models (LASCAM, a modified INCA model, SWAT and HBV-N-D) designed to simulate hydrological fluxes as well as mobilization and transport of one or several nitrogen species are applied over the meso-scaled River Fyris catchment in Mid-Eastern Sweden.

Hydrological calibration against 5 years of recorded discharge at two stations gives highly variable results from Nash-Sutcliffe Efficiency (NSE) values above 0.80 to values around 0.50. SWAT and HBV-N-D gives alternatively the best simulation result at each station respectively. Alteration of nitrogen parameters following Monte-Carlo or Latin-Hypercube stratified sampling schemes is realized in order to cover the parameter uncertainty of predictions for 3 nitrogen species: nitrate (NO_3), ammonium (NH_4) and total nitrogen (Tot-N) in terms of exported loads.

For each model and each nitrogen species, predictions are ranked in two different ways regarding the performance indicated by two different objective functions: the coefficient of determination R^2 and the Nash-Sutcliffe Efficiency (NSE). Model ensembles were compiled in various ways. A total of 396 Single Model Ensembles (SME) are generated using an increasing number of model members. Finally, 78 Multi-Model Ensembles (MME) are combined by using the best SME for each model, nitrogen species and station. The evolution of the two aforementioned objective functions is used as performance descriptor of the ensemble procedure.

In each studied case, there is always at least one compilation scheme which outperforms any of its members. The best SME are multiple-linear regression models with R^2 selected members, increasing the best NSE values from negativity up to very high ones (0.83). The uncertainty bounds of the SME are almost always smaller than the one introduced by the whole set of selected single model runs still including most of measurements and even more (half of the cases) than the bounds of the selected single runs set.

In the same way, there is always at least one MME combination scheme which outperforms all the SME, but the increase in model performance is pronounced than the difference between single model runs and SME. The best MME are the ones with the most members and both R^2 and NSE values are reaching 0.89 in the best case. Uncertainty areas described by MME are alternatively increased or reduced compared to the bounds delineated by their members. No general trend is deduced for the studied cases.

Keywords: *Ensemble modelling, Nitrogen, Fyris, Multi-model ensembles, Single-model ensembles, Uncertainty*

1. INTRODUCTION

In every modelling system predictive uncertainty results from a combination of at least three uncertainty sources: model structure uncertainty, forcing uncertainty and parameter uncertainty (also referred to as stochastic uncertainty). It is usually difficult to assess the contribution to the total uncertainty from each of these elements. However, ensemble approaches have been proposed to investigate part of this contribution (Breuer *et al.*, 2009).

Several global methods to assess parameter uncertainty have been described, e.g. the Monte-Carlo sampling based Generalized Likelihood Uncertainty Estimation (GLUE) approach (Beven & Binley, 1992). As parameter interactions are usually a sensitive source of uncertainty, a high number of realisations is required to cover a representative number of feasible parameter combinations and corresponding model simulations. Different combinations of parameter sets for a given model, based on a random sampling of parameter values (e.g. Monte-Carlo procedures or Latin-Hypercube stratified sampling, McKay *et al.*, 1979) in realistic ranges, are a common way to compile single-model ensembles (SME), i.e. combinations of distinct predictions obtained by perturbation of parameters, input data or initial conditions. SME built from random sampling are direct descriptions of the possible range of outcomes and illustrate part of the stochastic uncertainty.

Multi-model ensembles (MME) are based on the combination of several deterministic model outputs following different statistical post-processing of model outputs. They have been widely used in climatic and atmospheric sciences where MME usually outperform individual models and SME. However, it has only received little attention in hydrology even though initial MME studies of hydrological simulation were published in the mid 1990s (Shamseldin *et al.*, 1997). Lately some new initiatives have started to explore ensemble modelling in hydrology, such as the Distributed Model Intercomparison Project (DMIP, Smith *et al.*, 2004) and assessing the impact of Land Use Change on Hydrology by Ensemble Modelling (LUCHEM, Breuer *et al.*, 2009). However, research into hydro-biogeochemical applications remains scarce. MME are a state-of-the-art option to consider the structural model uncertainty in the total predictive uncertainty.

This study proposes to compile different SME and MME by merging the nitrogen outcomes of four models following different methods previously used in the LUCHEM project (Viney *et al.*, 2009). The models involved are LASCAM (Sivapalan *et al.*, 1996a,c; Viney & Sivapalan, 2000) and a Python language self written tool inspired by the snow and soil moisture routines of HBV (Bergström, 1976 & 1992) for flow generation and coupled to INCA (Whitehead *et al.*, 1998) for nitrogen predictions. The main modification of HBV concerns the separation of soil and groundwater which is made at the land-use scale instead of the sub-catchment scale. All the equations were adapted from literature references. The two remaining models are SWAT (Arnold *et al.*, 1998) in its 2005 version and HBV-N-D (Wrede, 2006) a distributed version of the original HBV water routines combined to the conservative solute transport model concepts of the TAC^D model (Wissmeier & Uhlenbrook, 2007). Evolution of the quality of predictions with respect to two different objective functions is done as well as analyses of uncertainty bounds. The final objective of this study is to answer two different questions:

- Is there an optimal merging scheme?
- Is there an optimal number of ensemble members?

2. MATERIALS AND METHODS

2.1. The River Fyris Catchment

The Fyris catchment is located in the mid-eastern part of Sweden, 90 km north of Stockholm. The Fyris River has a catchment area of 2000 km² and flows into the Lake Ekoln, a northern part of Lake Mälaren (Sweden's third largest lake) which drains into the Baltic Sea. It is a lowland catchment which elevation ranges between 15 and 115 metres a.s.l. Streams are draining from the north, east and west to the outlet at Flottsund (Figure 1).

Land-use is dominated by forest (mainly coniferous trees) with about 59% of the catchment while crop lands cover 33% of the area. Other minor land-use types are wetlands (4%), urban areas (~2%) and lakes (~2%). Forests are mainly associated with till and crop lands with clay soils (Lindgren *et al.*, 2007).

Daily records of 8 rain gauges and 3 temperature stations were available from SMHI (Swedish Meteorological and Hydrological Institute) for the 5 years study period (2000 to 2004). During this time mean annual precipitation was about 640 mm. The warmest and wettest month on average was July (>80 mm precipitation, +17°C mean daily temperature) while driest month was April (<40 mm precipitation) and coldest months were December and January (-1°C).

Over the studied period (2000 to 2004), two daily discharge series were available for two sub-catchments: Vattholma and Sävja, which contributing areas are 281 km² and 699 km² respectively (Figure 1). There is no gauging station at the catchment outlet to Lake Mälaren. High flows usually occur from late autumn to early spring. Inter-annual variability of discharge is high and thaw-refreezing events lead to high temporal variability of winter discharge in some years. Annual runoff was 219 mm at Vattholma and 189 mm at Sävja.

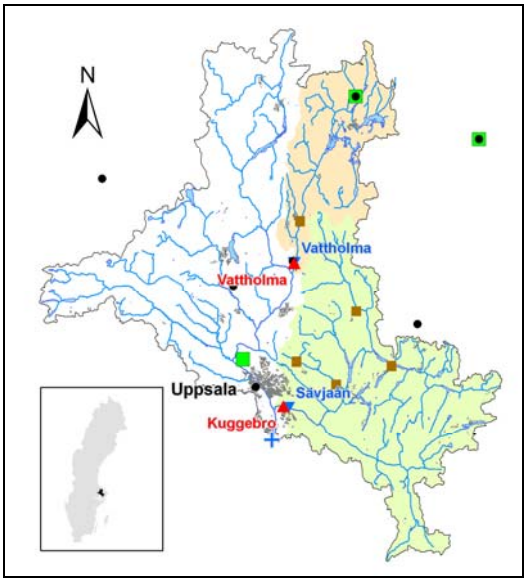

Figure 1. The River Fyris catchment: stream network, outlet (blue cross), Uppsala (shaded area), precipitation stations (black full circles), temperature stations (green squares), discharge gauges (blue triangles), water quality stations (red triangles) and treatment plants (brown squares).

For the same period, stream chemistry data from 4 long-term measurement stations of the SLU (*Swedish University of Agriculture*) was available for model applications. Monthly measurements of NO_3+NO_2, NH_4 and Tot-N concentrations resulted in a total of 60 measurements for each station. The water quality sampling stations Vattholma and Kuggebro are located closely to the gauging stations Vattholma and Sävja, respectively (Figure 1). Estimates of daily exported loads were computed for these gauging stations using concentrations measured at the aforementioned sampling stations. For NO_3, concentrations were estimated to be equivalent to the NO_3+NO_2 ones.

In-stream nitrogen input data from sewage treatment plants was also available on a daily time step for the largest plant in Uppsala, and with a biweekly or monthly resolution for the four smaller ones. The observed point source discharges were interpolated to a daily time step as described in Lindgren *et al*, (2007).

2.2. Models and Setup

The four models LASCAM, INCA, SWAT and HBV-N-D are able to simulate both runoff and the mobilization and transport of different N-species (see Table 1) at the landscape scale and at a daily time step. The models vary considerably in their smallest spatial units as well as the required input data and provided a good structural variability among the cohort.

LASCAM, INCA and SWAT are semi-distributed models. The same sub-catchment disaggregation scheme was adapted for each of them. The fully distributed HBV-N-D is implemented in the PCRaster (Karssenberg *et al.*, 2001) modeling environment and was already applied on the same watershed in the frame of a model comparison for nitrogen source apportionment (Lindgren *et al.*, 2007). We used the same setup of the HBV-N-D model based on 250 x 250 m grid cells.

Table 1. Model characteristics.

Model	Smallest spatial unit	Climate forcings	Outputs	N forcings
LASCAM	Sub-catchment	Daily P and PET	NO_3, NH_4, Tot-N	Rainfall concentration, fertilizer application
INCA	Land-Use	Daily P and T	NO_3, NH_4	Wet and dry deposition, fertilizer application, STP effluents
SWAT	HRU	Daily P, maximal and minimal daily T	NO_3, NO_2, NH_4, Organic-N	Rainfall concentration, fertilizer application, STP effluents
HBV-N-D	Grid cell	Daily P and T, monthly PET	Tot-N	Rainfall concentration, leaching coefficients, STP effluents

HRU: Hydrological Response Unit, Unique combination of a land-use with a soil type, P: Precipitation, T: Temperature, PET: Potential Evapotranspiration, STP: Sewage Treatment Plant

The hydrological components of each model were calibrated against two available runoff records over the period 2000-2004: Vattholma and Sävja (Figure 1). When available, integrated calibration tools were

Table 2. Calibration characteristics for hydrology

Model	Calibration Method	Objective function	Integrated tool	Reference
LASCAM	SCE-UA	Weighted sum of efficiencies[a]	Yes	Duan *et al.*, 1993
INCA	SCE-UA	Weighted sum of the square residuals	No	Duan *et al.*, 1993
SWAT	ParaSol	Independent sums of the square residuals	Yes	van Griensven *et al.*, 2002
HBV-N-D	PEST	Weighted sum of the square residuals	No	Doherty, 2005

[a] Efficiency for each flux is automatically calculated as 1-Var(residuals)/Var(observations)

preferred while PEST was used with the time-consuming HBV-N-D model. Calibration methods are summarized in the Table 2. Weights were set as the inverse of the standard deviation of observations. The integrated ParaSol method utilized with SWAT considers two independent sums of the squared errors (SSE).

In order to compare the goodness-of-fit resulting from the calibration efforts, calibration results were expressed as the Nash-Sutcliffe efficiency (NSE, Nash & Sutcliffe, 1970). Good NSE values can be achieved with the sole good representation of high peaks. Therefore, the quality of low flow simulations was also checked by computing logNSE, the efficiency based on logarithmic values of predictions and observations, which puts more weight also to lower flows than the usual NSE.

2.3. Ensembles construction and assessment

Model runs with fixed calibrated hydrology and varied parameters for the simulation of N modules were created for the set up of SME. Monte-Carlo and Latin-Hypercube stratified sampling (McKay *et al.*, 1979) procedures were used for the alteration of the N parameters. Several SME for each model, nitrogen species and measurement station were built. The members used in the ensembles construction were selected following two criteria:

1. The 2, 5 and 10 best runs considered using the determination coefficient R^2 as a goodness-of-fit indicator for the simulation of exported loads, and
2. The 2, 5 and 10 best runs are considered using the NSE as a likelihood estimate for the same fluxes.

The difference of these two criteria is that R^2 requires only the dynamics, or relative differences, to be simulated correctly, while NSE also evaluates the absolute values. For each case (i.e. N species, station and model) different merging schemes were then applied to each of the 6 sets of single model runs. It provided a total of 396 different SMEs (22 per model for each simulated N species at each station). SME combinations were realized following some of the methods used by Viney *et al.* (2009), including:

- daily mean of the predictions for each day,
- daily median of all ensemble members,
- multiple linear regression ensembles using the single runs as independent variables and the observations as dependent variables and
- multiple linear regression ensembles compiled in the same way but with a zero intercept.

The feasibility of each regression model was also checked by extrapolating the coefficients to the whole time series in order to check the occurrence of negative values.

One best SME regarding R^2 and NSE for each model, N species and measurement station was selected to set up MME. Following the previous schemes, we obtained 78 different combinations, alternatively using 2 or 3 models as members for each N species (INCA being not able to simulate Tot-N, and HBV-N-D considering only it), before examining the evolution of both objective functions for every generated MME and SME.

We used the P-factor to characterize the proportion of observed values bracketed by the uncertainty area of the full set of SME and MME (Abbaspour, 2007). The D-factor, defined as the ratio of the average width of the uncertainty bounds to the standard deviation of the observations, was the second computed descriptor. Best results are obtained with P-factor values close to 1 and D-factor values close to 0 (i.e. thin uncertainty ranges including most of the actual observations).

3. RESULTS

3.1. Hydrology

A summary of calibration results is presented in Table 3. A high variability is observed between the different NSE

Table 3. Goodness-of-fit indicators for calibrated runs

	Vattholma		Sävja	
Model	NSE	logNSE	NSE	logNSE
LASCAM	0.54	0.49	0.47	**0.76**
INCA	0.74	0.60	0.65	0.69
SWAT	**0.83**	**0.69**	**0.76**	0.64
HBV-N-D	0.65	0.67	**0.76**	**0.76**

values. The semi-distributed models perform better at Vattholma than at Sävja. Considering the second objective function, logNSE, which was not used in the automatic calibration process for hydrology itself, results are better for Sävja than for Vattholma, except for SWAT. While LASCAM showed the worst NSE for both discharge stations, it obtains the highest logNSE value for Sävja.

Considering these two objective functions as likelihood estimates of two different parts of flow dynamics, a global preferable model in flow simulation can be pointed out for each station, i.e. SWAT for Vattholma and HBV-N-D for Sävja, respectively.

3.2. Nitrogen

3.2.1 Single Runs Overview

Results are summarized in Table 4. Except for INCA, the best results are obtained at Sävja station. SWAT performs the best for NO_3 simulations regarding both likelihood estimators. For NH_4 the best R^2 and NSE are provided by INCA at Vattholma and by LASCAM at Sävja. For Tot-N, the best R^2 values are obtained with SWAT and the best NSE with LASCAM. NSE takes negative values 10 times, sometimes even when the R^2 values can be very high. It seems that the trends are well caught by the models with negative NSE, but with wrong absolute values.

3.2.2 Single-Model Ensembles

In this study, more than 70% of regression models are discarded due to the occurrence of negative values in the course of extrapolation of coefficients. There is a trend that if more members are used to build a SME, the more often these SME are discarded. We therefore alternatively used constrained regression models to circumnavigate this problem. For all investigated N compounds, improvements of at least one objective function occur. However, the median never increases both objective functions and only rare and weak improvements are achieved by using the mean of models. Best SME results are always obtained by using regression with R^2 selected members in unconstrained models. NSE remains negative only once (NH_4 prediction by SWAT at Vattholma).

In almost every case, the D-factor of the area described by a full set of SME is smaller than the one bounded by its ensemble members. It is accompanied by both, increasing and decreasing P-factors (data for R- and D-factors not shown). The D-factor is only increasing when the P-factor is really poor for the whole set of selected single runs (NH_4 for SWAT).

Table 4. Nitrogen results summary

Models	Vattholma						Sävja					
	NO_3		NH_4		Tot-N		NO_3		NH_4		Tot-N	
	R^2	NSE	R^2	NSE	R^2	NSE	R^2	NSE	R^2	NSE	R^2	NSE
LASCAM												
Best Single Run[a]	0.511	0.336	0.188	< 0	0.606	0.249	0.603	0.550	0.502	0.109	0.615	0.545
Best SME[b]	0.513	0.511	0.272	0.272	0.622	0.622	0.616	0.574	0.517	0.517	0.639	0.639
INCA												
Best Single Run[a]	0.301	< 0	0.471	< 0			0.229	< 0	0.323	0.029		
Best SME[b]	0.301	0.301	0.473	0.473			0.229	< 0	0.354	0.354		
SWAT												
Best Single Run[a]	0.678	0.502	0.32	< 0	0.84	< 0	0.839	0.720	0.164	< 0	0.821	< 0
Best SME[b]	0.686	0.686	0.326	0.326	0.864	0.864	0.833	0.827	0.175	0.175	0.831	0.830
HBV-N-D												
Best Single Run[a]					0.385	< 0					0.624	< 0
Best SME[b]					0.393	0.393					0.752	0.752
Best MME[b]	0.754	0.728	0.618	0.618	0.876	0.876	0.867	0.849	0.645	0.645	0.890	0.883

[a] Best single runs regarding R^2 and NSE are not necessarily obtained with the same parameter set
[b] Best SME and MME R^2 and NSE values are obtained with the same ensemble

3.2.3 Multi-Model Ensembles

Similar to SME, the feasibility of the regression models is checked prior to further model evaluation. On the 24 MME combined by using regression, 9 are discarded. Simple mean increases both R^2 and NSE in 25% of cases compared to the best SME values. Median outperforms R^2 in only one case.

All the available multiple linear regression models present an improvement of both likelihood estimators compared to their members. Better results are obtained by including the maximum of members (here 3) in unconstrained models. Performance of the MME is more improved when the members that are combined in an ensemble present weaker result (e.g. NH_4 at Vattholma) compared to already high performing model members (e.g. Tot-N).

D-factor is lower for MME than for their respective members in 3 cases in which the P-factor is only increased once (Tot-N at Vattholma). For the remaining MME, improvements of P-factor occur twice.

4. DISCUSSION

4.1. Hydrology

As illustrated by the hydrological calibration results in Table 3, the ranking of model performances using a single objective function can be misleading. Structural differences involve a high variability in the calibration results. For instance, a bad performance during high flow simulation for LASCAM is illustrated by overall weak NSE values. The lack of a snow module in LASCAM may be the clue to explain the bad fit during high flow periods, as hydrographs in this geographical area are dominated by snow melt events. However, this weakness is compensated at Sävja by a good simulation of low flows. An already planned improvement of LASCAM is the integration of a degree-day approach to handle snow pack and snow melt simulation. This study focuses on N predictions and therefore, the effects of ensemble modelling on hydrological predictions for the River Fyris catchment are presented elsewhere.

4.2. Nitrogen

A good match of observed and simulated N loads is not necessarily required to achieve good R^2 values. Selecting the best single runs regarding this criterion is therefore probably the source of large predictive uncertainty reflected in large D-factors. Best results for SME are achieved by combining the best R^2 selected members in multiple linear regression models. Effects of this procedure are especially obvious for the improvements of NSE. The more members are combined into linear models; the better is the performance thanks to more precise adjustments. But the risk of unrealistic values increases with an increasing number of members. Uncertainty bounds are usually reduced between single runs and SME but still include most of the measurements, reflected by good P-factors.

MME results always show the best overall model performances but improvements are not very high compared to SME. Due to variability of results obtained in this study, this type of ensemble generation cannot be demonstrated as a definitive way to reduce the prediction uncertainty. This is mainly attributable to the already good performances of the SME, as there is only limited space for further improvement of the overall model performance. Uncertainty ranges which are introduced are also smaller and then harder to decrease than in the case of the SME. Combining MME from the best single runs rather than the SME could be another way to take account of the global prediction uncertainty linked to the full set of considered models.

The compilation of ensembles was only realized in a re-prediction exercise and regression coefficients were computed by using the full set of measurements. Further investigation may be to check their applicability in split-sample and proxy-basins approaches.

5. CONCLUSION

Numerous SME and MME were compiled. In every studied situation numerous combination schemes show improvements compared to the performance of its single members. Regression schemes are the most efficient combinations when members are selected by their R^2 values. Ensembles are then only adjustment of the absolute values of already well-trended predictions. The risk of unrealistic values remains high with higher numbers of ensemble members and has to be checked.

At the same time, the evolution of uncertainty bounds between members and full sets of ensembles show variable patterns. SME uncertainty bounds are smaller than uncertainty described by their single members. Results for MME do not follow a general trend, but often, uncertainty is still though often only slightly reduced compared to the whole set of selected single runs.

We demonstrated in this paper that the ensemble approach might allow reducing simulation uncertainty and increasing simulation performance. Multiple linear regression models are clearly the best merging schemes when applied to R^2 selected members. Better results are obtained with more members. Nevertheless an

optimal number of members cannot clearly be defined and it is rather the maximal number of members preventing from unrealistic values. However, this could be also the case as only 4 different model structures were available for N predictions. As compared to hydrological predictions in the DMIP or LUCHEM experiment, less potential model structures are at hand for predicting N fluxes on the catchment scale.

REFERENCES

Abbaspour, K.C., 2007. 'User Manual for SWAT-CUP, SWAT Calibration and Uncertainty Analysis Program', Swiss Federal Institute of Aquatic Science and Technology, Eawag, Dübendorf, Switzerland. http://www.eawag.ch/organisation/abteilungen/siam/software/swat/index_EN (Last accessed March 2009)

Arnold, J.G., Srinivasan, R., Muttiah, R.S. and Williams, J.R., 1998. 'Large area hydrologic modelling and assessment part I: model development', J. Am. Water Resour. Assoc., 34, 73-89.

Bergström, S., 1976. 'Development and application of a conceptual runoff model for Scandinavian catchments', SMHI, Report No. RHO 7, Norrköping, 134 pp.

Bergström, S., 1992. 'The HBV model – its structure and applications', SMHI Hydrology, RH No. 4, Norrköping, 35pp.

Beven, K.B. and Binley, A., 1992. 'The future of distributed models: models calibration and uncertainty prediction', Hydrol. Proc., 6: 279-298.

Breuer, L., Huisman, J.A., Willems, P., Bormann, H., Bronstert, A., Croke, B.F.W., Frede, H.-G., Gräff, T., Hubrechts, L., Jakeman, A.J., Kite, G., Lanini, J., Leavesley, G., Lettenmaier, D.P., Lindström, G., Seibert, J., Sivapalan, M. and Viney, N.R., 2009. 'Assessing the impact of land use change on hydrology by ensemble modelling (LUCHEM). I: Model Intercomparison with current land use.', Adv. in Wat. Res., 32: 129-146.

Duan, Q., Gupta, V.K. and Sorooshian, S., 1993. 'A shuffled complex evolution approach for effective and effeicient global minimization', J. Optim. Theory Appl., 76: 501-521.

Doherty, J., 2005. 'PEST: Model-independent parameter estimation, user manual', Watermark numerical computing, Brisbane.

Karssenberg, D., Burrough, P.A., Sluiter, R. and de Jong, K., 2001. 'The PC raster software and course materials for teaching numerical modelling in the environmental sciences', Trans. GIS, 5(2): 99-110.

Lindgren, G.A., Wrede, S., Seibert, J. and Wallin, M., 2007. ‚Nitrogen source apportionment modeling and the effect of land-use class related runoff contributions', Nordic Hydrology, 38(4-5): 317-331.

McKay, M.D., Beckman, R.J. and Conover, W.J., 1979. 'A comparison of three methods for selecting values of input variables in the analysis of output from a computer code', Technometrics, 21 (2): 239-245.

Nash J.E. and Sutcliffe, J.V., 1970. 'River flow forecasting through conceptual models. Part I: A discussion of principles', J. Hydrol., 10: 282-290.

Shamseldin, A.Y., O'Connor, K.M. and Liang, G.C., 1997. 'Methods for combining the outputs of different rainfall – runoff models', J.Hydrol, 197: 203-229.

Sivapalan, M., Ruprecht, J.K. and Viney, N.R., 1996. 'Water and salt balance modeling to predict the effects of land-use changes in forested catchments. 1. Small catchment water balance model', Hydrol. Proc., 10, 393-411.

Sivapalan, M., Viney, N.R. and Jeevaraj, C.G., 1996. 'Water and salt balance modeling to predict the effects of land-use changes in forested catchments. 3. The large catchment model' Hydrol. Proc., 10, 429-446.

Smith, M.B., Georgakakos, K.P. and Liand, X., 2004. 'Special issue: the distributed model intercomparison project (DMIP)', J. Hydrol., 298:1-335.

van Griensven, A., Francos, A. and Bauwens, W., 2002. 'Sensitivity analysis and autocalibration of an integral dynamic model for river water quality', Water Science and Technology, 45 (9): 325-332.

Viney, N.R., Bormann, H., Breuer, L., Bronstert, A., Croke, B.F.W., Frede, H.-G., Gräff, T., Hubrechts, L., Jakeman, A.J., Kite, G., Lanini, J., Leavesley, G., Lettenmaier, D.P., Lindström, G., Seibert, J., Sivapalan, M. and Willems, P., 2009. 'Assessing the impact of land use change on hydrology by ensemble modelling (LUCHEM). II: Ensemble combinations and predictions', Adv. in Wat. Res., 32: 147-158.

Viney, N.R., Sivapalan, M., Deeley, D., 2000. 'A conceptual model of nutrient mobilization and transport applicable at large catchment scales', J. of Hydrol., 240, 23-44.

Whitehead, P.G., Wilson, E.J. and Butterfield, D., 1998. 'A semi-distributed integrated flow and nitrogen model for multiple source assessment in Catchments (INCA): Part I – model structure and process equations', The Science of the Total Environment, 210/211: 547-558.

Wissmeier, L. and Uhlenbrook, S., 2007. 'Distributed, high-resolution modelling of ^{18}O signals in a meso-scale catchment', Journal of Hydrology, 332(3-4): 497-510.

Wrede, S., 2006. 'Distributed solute transport modelling in the Fyrisån catchment in Sweden', MSc Thesis, Institute of Hydrology, University of Freiburg, Freiburg.