Water Quality and Sediment Behaviour of the Future: Predictions for the 21st Century (Proceedings of Symposium HS2005 at IUGG2007, Perugia, July 2007). IAHS Publ. 314, 2007.

Modelling catchment-scale nitrate transport using a combined process-based and data-driven approach

RAJESH RAJ SHRESTHA & MICHAEL RODE

Department of Hydrological Modelling, UFZ – Helmholtz Centre for Environmental Research, Brueckstrasse 3a, D-39114 Magdeburg, Germany

rajesh.shrestha@ufz.de

Abstract Diffuse nitrate pollution in catchments is mainly driven by hydrological flow components characterised by complex relationships with streamflow nitrate concentration. This paper demonstrates a combined process-based–artificial neural network (ANN) approach for the simulation of streamflow nitrate concentration based on the relationships between driving and resultant variables. The simulated hydrological flow components from a process based WaSiM-ETH model, together with observations, are used to train two different ANNs. The results show a reasonable match between observed and simulated streamflow and nitrate-N concentration. The ANN with temperature as an input performed better than the ANN without it, indicating the effect of seasonal variability. Nash-Sutcliffe coefficients of 0.746 and 0.856 were obtained for streamflow in calibration and test periods, respectively, while these coefficients were 0.819, 0.629 and 0.627 for nitrate-N concentration in training, cross-validation and test periods, respectively. Hence, the combined approach offers an effective and efficient methodology for modelling catchment scale nitrate dynamics.

Key words artificial neural network; hybrid model; nitrate transport; water balance

INTRODUCTION

The occurrence of diffuse nitrate pollution in rivers is driven by complex biochemical and hydrological processes related to nitrogen and water cycles. The deterministic modelling of these processes requires a realistic description of different processes and their interactions at spatial and temporal scales. Although the hydrological processes can be simulated relatively reliably using process-based models, the simulation of nutrient transport is associated with large uncertainties due to a limited understanding of the processes and restrictive data availability.

An alternative approach for modelling catchment-scale nitrate dynamics is the treatment of processes as a relationship between driving and resultant variables. A number of studies concerning temporal and spatial patterns of nutrient transport processes indicate that the hydrological connectivity via subsurface flow is a key factor in biochemical dynamics (Hornberger *et al.*, 1994; Stieglitz *et al.*, 2003). However, the relationship between driving variables such as discharge and nitrate concentration is highly nonlinear, which might be characterised by anticlockwise hysteresis with lower nitrate concentrations occurring on the rising limb of discharge hydrograph compared to the receding limb (Van Herpe & Troch, 2000). The analysis undertaken by Creed *et al*. (1996) and Martin *et al*. (2004) showed intermediate or inverse cycles with higher nitrate concentrations in streams during summer months. The seasonal variability might be an important factor since the neural network analysis by Lischeid & Langusch (2004) identified mean daily temperature as an important driving variable for modelling stream nitrate concentration.

Data-driven models, such as artificial neural networks (ANNs), provide a framework for capturing such highly complex and nonlinear relationship between different variables. The ANNs are a connectionist computing paradigm, inspired by the human brain, that can abstract knowledge from physical processes and make it available for use. The capability of the ANNs to learn from the relationships between causative and resultant variables using data sets characterised by noise and uncertainties makes them suitable for modelling catchment scale nutrient transport processes. Several researchers have demonstrated the application of ANNs for rainfall–runoff modelling and streamflow simulation (e.g. Dawson & Wilby, 1999; Shrestha, *et al.*, 2005). In recent years, the application of ANNs for nitrate transport problem has gained in popularity, and includes the prediction of daily streamflow and nitrate load based on rainfall (Yu *et al.*, 2004) and weekly nitrate-N forecasts using artificial neural networks (Markus *et al.*, 2003). The performances of the ANNs have also been found to be superior compared to process-based models such as INCA for nitrate-N simulation (Lischeid & Langusch, 2004) and SWAT for monthly water quality forecasting (Srivastava *et al.*, 2006).

The ANNs also allow integration of results of other models and measurements in a hybrid model structure for a synergetic combination of traditional process-based and data-driven models (Krasnopolsky & Fox-Rabinovitz, 2006). The hybrid approach provides a framework for combining the strengths of both systems based on the prevailing understanding of relationships between different driving variables on catchment scale nitrate dynamics. The aim of this study is to demonstrate an application of a hybrid process-based ANN approach for the simulation of catchment-scale nitrate transport at a high resolution daily time step. The reliable simulations of hydrological flow components from a water balance model are combined with observations in an ANN for mapping the relationships with streamflow nitrate concentration.

STUDY AREA AND DATA

The Weida study catchment comprises an area of 100 km^2 in the lower mountain range in the state of Thuringia, Germany. The Weida stream is a small tributary to the Weisse Elster River in the Elbe River basin which flows through a cascade of Zeulenroda and Weida drinking water reservoirs. Hence, water quality in the Weida stream is an important consideration for the long-term drinking water supply in the region. As two-thirds of the catchment is agricultural land, diffuse nutrient pollution is the major contributor to the quality of Weida stream water.

The spatial data obtained from the catchment included a digital terrain model of 25 m grid resolution with elevation differences between 357 and 552 m. Land-use data are based on the classification of Landsat ETM images from the year 1999 for the entire Saale basin. The land-use data consist of six classes with three predominant classes: arable land (40%), forest (29%) and grassland (26%). Soil data are based on a 1:25 000 soil map which consists of four main classes and 23 sub-classes in total dominated by sandy loam (40%) and silt loam (36%).

The temporal data used in the study consist of precipitation data from five gauging stations, with average annual precipitation of 640 mm between 1988 and 2004. In addition, relative humidity, relative sunshine hours, temperature and wind velocity were obtained from two climatic stations in the vicinity of the catchment. High resolution time series of discharge and water quality data at one minute intervals are available from the Laewitz gauging station at the catchment outlet between 1998 and 2004, which is aggregated to daily time steps for modelling purposes. It should be noted that both the discharge and nitrate-N information comprise unprocessed data, which include a considerable number of missing values. The nitrate-N data, in particular, also contain noticeable noise.

WATER BALANCE MODEL

The deterministic Water balance Simulation Model (WaSiM-ETH; Schulla & Jasper, 2001) was used in this study for the simulation of different hydrological flow components. The WaSiM-ETH is a process-based distributed modelling system which can simulate different hydrological processes at spatial and temporal scales using a modular, object-oriented architecture. WaSiM-ETH uses a Green and Ampt/TOPMODEL approach for the simulation of runoff generation from infiltration excess and saturated areas. The Green and Ampt equation calculates infiltration based upon soil moisture conditions and surface runoff occurs when soil infiltration capacity has been exceeded. The TOPMODEL (Beven & Kirby, 1979) is a variable contributing area approach based upon the distribution of saturation deficit. In this WaSiM-ETH implementation of TOPMODEL, soil water balance and runoff generation is simulated separately for each grid cell based on the spatial distribution of soil topographic index. The WaSiM-ETH generates surface runoff when the unsaturated zone is filled (Schulla & Jasper, 2001).

The WaSiM-ETH was set up for the Weida catchment using a digital elevation model, soil and land-use classes for a 25 m grid cell size. Temporal data at a daily time step consisting of time series of precipitation, relative humidity, relative sunshine hours, temperature and wind velocity were used. Discharge from Loessau Reservoir was introduced as external inflow. Data for three hydrological years (November–October, 1998–2000) were used for model calibration and from a further four years (2001–2004) for independent evaluation (test) of model results. The model simulation was started one year in advance so that initial conditions do not affect the model outputs.

The WaSiM-ETH model was calibrated by comparing the simulated and observed discharge data from the Laewitz gauging station. The model calibration was carried out by adjusting the nine parameters using the Parameter Estimation (PEST; Doharty, 2004) programme, which consists of a gradient-descent parameter estimation algorithm based on the Gauss-Marquardt-Levenberg optimisation method. The overall performance of the model results was evaluated using the statistical criteria of the Nash-Sutcliffe coefficient of efficiency (NSCE), the coefficient of determination (R^2) and mean absolute error (MAE).

ARTIFICIAL NEURAL NETWORK NITRATE TRANSPORT MODEL

The ANN-based nitrate transport modelling comprised a universal approximation of the system without explicitly taking account of the physical processes. Multilayer feedforward networks (MFNs) were used for this purpose, and these can be trained in a supervised manner to solve highly nonlinear problems. The capability of the MFNs can be enhanced by using feedback loops subjected to time delays, which provide a dynamic state to the networks. For more details on ANNs the readers are referred to Haykin (1994).

The first step in setting up an ANN model was the selection of appropriate driving variables as model inputs. The criteria for the selection were their effect on nitrate transport processes and availability of data at the daily time step used in this study. Based on these criteria, four different variables were identified as suitable inputs, which included the subsurface flow, surface runoff, external inflow from the Loessau Reservoir and mean daily temperature. The subsurface and surface flow components were obtained from WaSiM-ETH simulation, and the external inflow and mean daily temperature were observation data sets. Two different combinations of input variables were considered with all four input variables (ANN 4 inputs) and three inputs without temperature (ANN 3 inputs). Variables that affect the nitrate entry, such as fertilizer input and crop cycles, were not included as these data were not available at a daily time step.

The structure of the ANN consists of 20 neurons in the first hidden layer, 10 neurons in the second hidden layer and one neuron in the output layer. Input and output data sets were normalised using the mean and standard deviation of training data sets. It was observed during the preliminary trials that the use of recurrent feedback in the output layer enhanced the performance of the ANNs. Hence, recurrent networks were used, although this considerably slowed down the training process. The network consisted of hyperbolic tangent activation functions in the hidden layers and linear activation functions in the output layer. The ANN model was developed using the procedure of the MATLAB Neural Network Toolbox (Demuth *et al.*, 2006). The backpropagation algorithm with Bayesian regularisation of the Levenberg-Marquardt method was used for ANN training. Early stopping criteria provided by the validation data sets were used to prevent overtraining. The test data sets were used independently for the evaluation of model performance. For further analyses, the simulated nitrate-N concentration from the ANN and discharge from the water balance model were used to calculate daily nitrate-N load in the stream. The performance of the ANNs was assessed using the same criteria of NSCE, R^2 and MAE.

RESULTS AND DISCUSSION

Water balance model

The results obtained from the calibration of the WaSiM-ETH water balance model using the parameter estimation programme PEST for the calibration (1998–2000) are shown in Fig. 1. The result for the test data sets (2001–2004) are shown in Fig. 2.

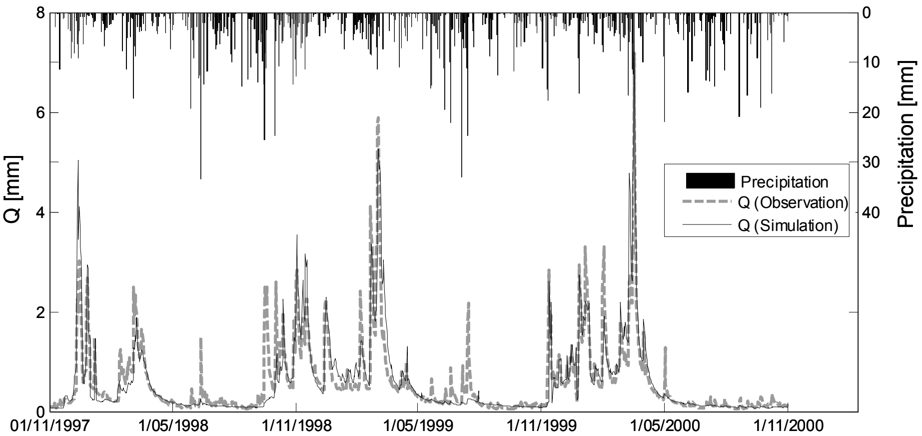

Fig. 1 Comparison of observed and simulated discharge for the calibration period.

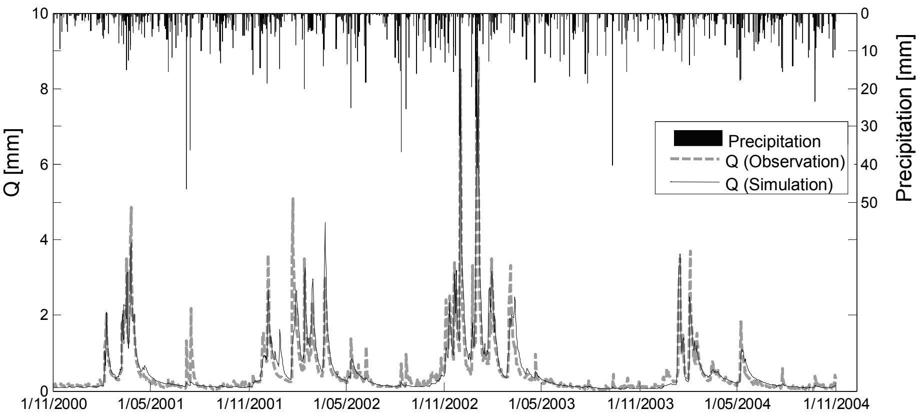

Fig. 2 Comparison of observed and simulated discharge for the test period.

Based on the results of the study, it can be seen that the model was able to simulate the general runoff dynamics adequately for both the calibration and test data sets. Most of the winter peaks were simulated reasonably well with a good match of the rising and recession limbs of the hydrographs. However, the model has problems in simulating some summer peaks, most of which were underestimated. These summer peaks occur after high intensity precipitation, which produces high surface runoff. Due to the averaging effect of the daily time step, the model cannot calculate high surface runoff and uses the excess precipitation to replenish the high saturation deficit present during summer.

The statistical performance of the model results indicates a good agreement between the simulated and observed discharges. The NSCE, R^2 and MAE performance were 0.746, 0.796 and 0.19 mm for the calibration periods and 0.856, 0.868 and 0.156 mm for the test periods, which are reasonable values. The statistical performance of the model was better for the test data sets in comparison with the calibration data sets, which indicates a better model fit for the test data sets. However, there is a better match for the calibration period in the water balance, with mean annual differences between the cumulative observed and simulated flows of 2.5 mm for the calibration period and 11.5 mm for the test period. Based on the overall performance, it can be said that the model was able to represent the catchment characteristics reasonably well.

ANN nutrient transport model

The results of the ANN models for the nitrate-N concentration and calculated daily nitrate-N load for training data sets are shown in Figs 3 and 4, respectively. Figures 5 and 6 show model results for nitrate-N for cross-validation and test data sets for concentration and calculated daily nitrate-N loads, respectively. The results show a good match between observations and simulations for the training data set, with regard to the dynamics of the stream nitrate-N concentrations, for both the ANNs with four and three inputs. The overall statistical performance of the ANN with four inputs in terms of NSCE, R^2 and MAE were 0.817, 0.819 and 1.056 mg/L, respectively, which is better than the ANN with three inputs which had NSCE, R^2 and MAE of 0.757, 0.773 and 1.188 mg/L, respectively.

The ANNs showed a reasonable performance for the cross-validation and test data sets. In this case too, the performance of the four input model was better than the three input model. For the cross-validation data sets, the four input ANN had a NSCE, R^2 and MAE of 0.629, 0.787 and 1.386 mg/L compared to 0.534, 0.733 and 1.496 mg/L

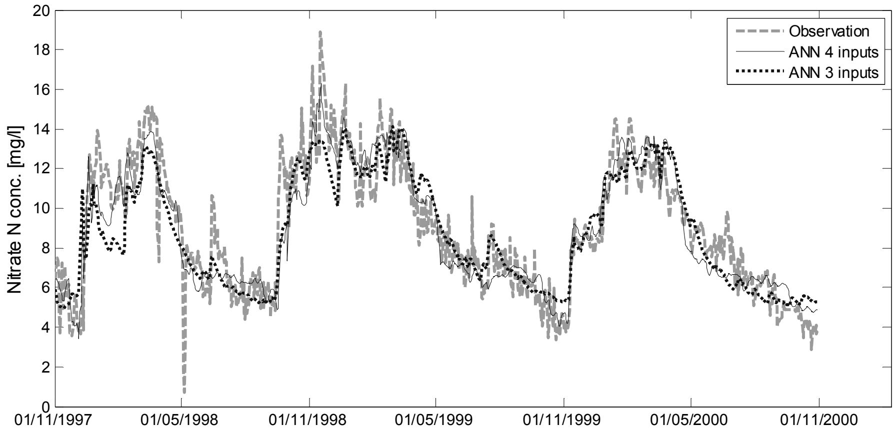

Fig. 3 Comparison of observed and ANN simulated nitrate-N concentration for the training period.

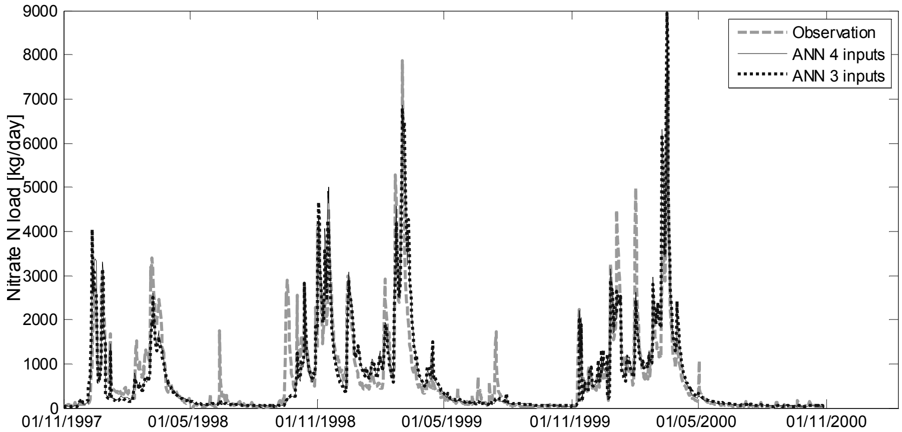

Fig. 4 Comparison of observed and ANN simulated daily nitrate-N load for the training period.

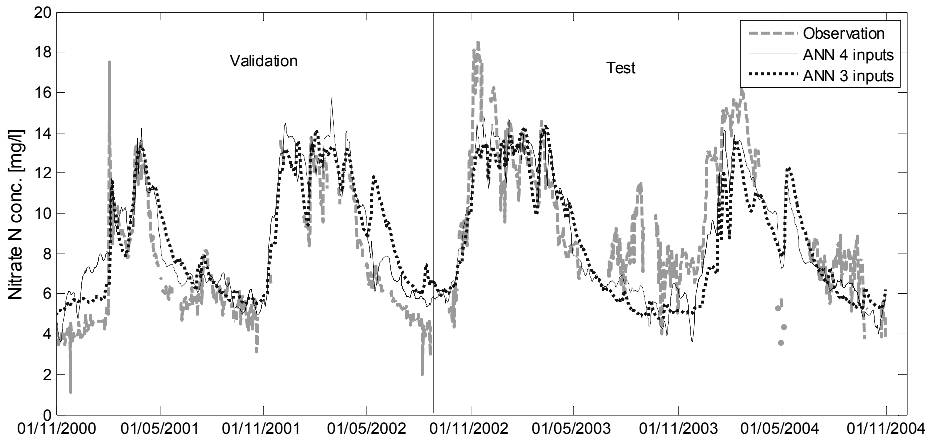

Fig. 5 Comparison of observed and ANN simulated nitrate-N concentration for the cross-validation and test periods.

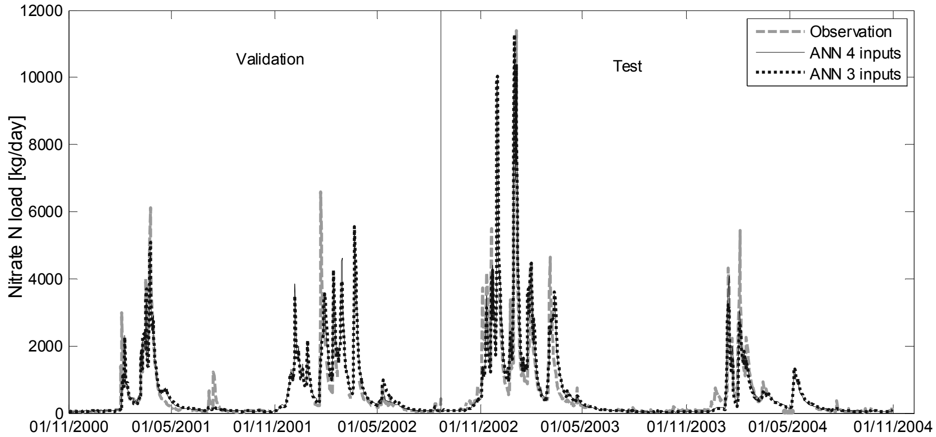

Fig. 6 Comparison of observed and ANN simulated daily nitrate-N for the cross-validation and test periods.

for the three input ANN. Similarly, for test data sets NSCE, R^2 and MAE values were 0.627, 0.719 and 1.622 mg/L for the four input ANN compared with 0.437, 0.582 and 2.044 mg/L for the three input ANN. The models had problems, mainly in correctly reproducing the extremes of low and high nitrate-N concentrations. Both the models slightly overestimated the 2002 summer nitrate-N concentration, while there was a considerable underestimation of the 2003 summer values. The period between May 2003 and November 2003 showed a co-occurrence of low flow, high temperature and relatively high nitrate-N concentration, which were not present in the training data sets and hence the model was not able to reproduce nitrate-N concentration in the period correctly. However, if the long-term data was used for the ANN training, it can be expected that the model would be able to reproduce such phenomenon more correctly.

The performance of the model for the simulation of nitrate-N load was good for all three data sets, which shows the ability of the model to simulate the nitrate-N balance, which is important for catchment-scale nutrient management. For the ANN with four inputs, the performance of the calculated daily stream nitrate-N load in terms of NSCE, and R^2 criteria was 0.748, and 0.808, respectively, which was poorer than the equivalent results for concentration. The NSCE and R^2 performance of the four input models for the cross-validation data sets were 0.771 and 0.771, and test data sets were 0.853 and 0.864, respectively, which was better than N-concentration for the same data sets. This was due to the better performance in the discharge simulation from the water balance model. The statistical performance for the load with the three inputs showed similar trends, but was inferior to the results from the four input ANN.

The superior performance of the ANN with the mean air temperature as an input indicates that the temperature is also an important driving variable. As the daily temperature is related to seasonal variability, the results indicate that the seasonal variability is also an important factor in the streamflow nitrate-N concentration in the Weida catchment.

CONCLUSIONS

The study has demonstrated an application of a combined deterministic – ANN model for the simulation of catchment-scale nitrate dynamics. A reliable water balance simulation from a process-based spatially distributed WaSiM-ETH model was combined with observations for the simulation of streamflow nitrate-N concentration and load. Four variables: subsurface flow, surface runoff, external inflow and mean daily temperature were considered as input variables for the FRBM. Two different ANN models were trained with and without temperature as an input variable. Both the ANNs produced acceptable performance with superior statistical parameters for the model that included temperature, which indicated the effect of seasonal variability. Based on the results of the study, it can be seen that the hybrid approach may be used to combine the advantage of both the process-based and data-driven models for an effective and efficient methodology for modelling nitrate transport from a catchment. The approach can also be easily adapted for modelling nitrate and other dissolved solute dynamics in different catchments.

REFERENCES

Beven, K. J. & Kirkby, M. J. (1979) A physically based variable contributing area model of basin hydrology. *Hydrol. Sci. Bull.* **24**(1), 43–69.

Creed, I. F., Band, L. E., Foster, N. W., Morrison, I. K., Nicolson, J. A., Semkin, R. S. & Jeffries, D. S. (1996) Regulation of nitrate-N release from temperate forests, a test of the N flushing hypothesis. *Water Resour. Res.* **32**, 3337–3354.

Dawson, C. W. & Wilby, R. B. (1999) A comparison of artificial neural networks for flow forecasting. *Hydrol. Earth System Sci.* **3**(4), 529–540.

Demuth, H., Beale, M. & Hagan, M. (2006) *Neural Network Toolbox for Use with MATLAB*, User's Guide, version 5. The MathWorks Inc. Online documentation: http://www.mathworks.com/access/helpdesk/help/toolbox/nnet/.

Doherty, J. (2004) *PEST Model Independent Parameter Estimation*, fifth edn of user manual. Watermark Numerical Computing, Brisbane, Australia.

Haykin, S. (1994) *Neural Networks a Comprehensive Foundation*, first edn. Macmillan College Publishing Company Inc., New York, USA.

Hornberger G., Bencala, K. & McKnight, D. (1994) Hydrological controls on dissolved organic carbon during snow melt in the Snake River near Montezuma, Colorado. *Biogeochemistry* **25**, 147–165.

Krasnopolsky V. M. & Fox-Rabinovitz M. S. (2006) A new synergetic paradigm in environmental numerical modeling: Hybrid models combining deterministic and machine learning components. *Ecol. Modelling* **191**(1), 5–18.

Lischeid, G. & Langusch, J. (2004) Comparative simulation of the nitrogen dynamics using the INCA model and a neural network analysis: implications for improved nitrogen modelling. *Hydrol. Earth System Sci.* **8**(4), 742–750.

Markus, M., Tsai, C. W. S. & Demissie M. (2003) Uncertainty of weekly nitrate-nitrogen forecasts using artificial neural networks. *J. Env. Engng. ASCE* **129**(3), 267–274.

Martin, C., Aquilina, L., Gascuel-Odoux, C., Gascuel-Odoux, C., Molenat, J., Faucheux, M & Ruiz, L. (2004) Seasonal and interannual variations of nitrate and chloride in stream waters related to spatial and temporal patterns of groundwater concentrations in agricultural catchments. *Hydrol. Processes* **18**(7), 1237–1254.

Schulla, J. & Jasper, K. (2001) Model Description WaSiM-ETH. Internal report, ETH-Zurich, Switzerland.

Shrestha, R. R., Theobald, S., & Nestmann, F. (2005) Simulation of flood flow in a river system using artificial neural networks. *Hydrol. Earth System Sci.* **9**(4), 313–321.

Srivastava P., McVair J. N. & Johnson T. E. (2006) Comparison of process-based and artificial neural network approaches for streamflow modeling in an agricultural watershed. *J. Am. Water Resour. Assoc.* **42**(3), 545–563.

Stieglitz, M., Shaman, J., McNamara, J., Engel, V., Shanley, J. & Kling, G. W. (2003) An approach to understanding hydrologic connectivity on the hill slope and the implications for nutrient transport. *Global Biogeoch. Cycles* **17**(4), 1105. 16-1–16-15.

Van Herpe Y. J. P. & Troch P. A. (2000) Spatial and temporal variations in surface water nitrate concentrations in a mixed land use catchment under humid temperate climatic conditions. *Hydrol. Processes* **14**(14), 2439–2455.

Yu C., Northcott W. J. & McIsaac G. F. (2004) Development of an artificial neural network for hydrologic and water quality modeling of agricultural watersheds. *Trans. Am. Soc. Agric. Engrs* **47**(1), 285–290.