

# Translating nitrogen emissions to loads in the aquatic ecosystem of the Scheldt basin

**P.J. Haest**<sup>a,b</sup>, **S. Broekx**<sup>b</sup>, **J. van der Kwast**<sup>b</sup>, and **P. Seuntjens**<sup>a,b,c</sup>

<sup>a</sup> *University of Antwerp, Antwerp, Belgium*

<sup>b</sup> *VITO NV, Boeretang 200, B-2400 Mol, Belgium*

<sup>c</sup> *Ghent University, Ghent, Belgium*

*Email: [pieterjan.haest@vito.be](mailto:pieterjan.haest@vito.be)*

**Abstract:** The European Union (EU) adopted the Water Framework Directive (WFD) in 2000 aiming to achieve a ‘good status’ in its aquatic ecosystems by 2015. In this respect, a numerical spatio-temporal model is developed to describe the nitrogen dynamics in large river basins with the Scheldt basin as pilot study area.

The uncertainty about the input data and the groundwater flow at this scale, and the intended application for scenario analyses prompted the use of an open model structure where the model developer can test different process descriptions to match the observed characteristics of water flow and water quality. Therefore, the fate model is programmed in a Python environment using the PCRaster library enabling an adequate balance between process descriptions and necessary input data. The PCRaster software expands the traditional GIS software packages with a time dimension and domain specific functions that operate on the defined GIS layers. The Python/NumPy environment enables a modeller with no advanced programming skills to add new or existing code to simulate specific environmental processes. The resulting GIS-based model uses a monthly time step and a spatial resolution of 1 km<sup>2</sup> to simulate time-dependent discharges and concomitant nitrogen dynamics in the Scheldt basin. It discerns 4 compartments in the basin: the soil, shallow subsurface, deep subsurface and the river network. Water flows along the steepest slope in all model compartments. The mass transport module uses these calculated flows to translate emissions to loads in the aquatic ecosystem. Diffuse emissions and direct inputs are calculated from administrative and statistical data. These emissions are geographically defined or are distributed over the domain according to land use and connectivity to the sewer system. Datasets provided at the EU level are used to construct the model when available.

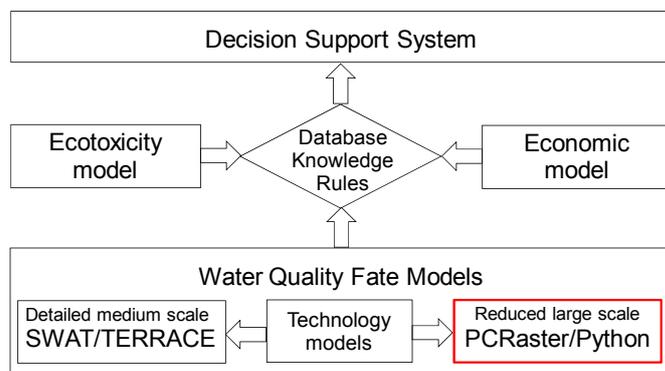
Different scenarios with conventional and innovative rehabilitation technologies are defined to estimate the effect of a measure on the reduction of the nitrogen load in the aquatic ecosystem. Conventional measures mainly reduce the future amount of nitrogen emitted in the basin, e.g. a reduction of livestock numbers, while innovative rehabilitation technologies such as wetlands reduce the amount of nitrogen that finally reaches the surface waters and, as such, are also effective in reducing nitrogen loads from historical emissions. The results of the scenario analysis will be used to derive knowledge rules translating nitrogen emissions and reductions thereof to loads in the aquatic ecosystem. This information will be used in a database driven Decision Support System.

The open model structure in Python/NumPy with the PCRaster library yielded a fast prototyping environment enabling scientists with no advanced programming skills to test different process descriptions to match the observed characteristics of water flow and water quality. The prototyping environment was an asset since the application of the conceptual model approach for scenario analysis in a fully distributed large river basin at a monthly time step proved difficult. The resulting GIS-based model will be used for a first evaluation of innovative rehabilitation technologies versus conventional measures at the river basin scale, taking into account the current limitations for predictions at the subbasin scale.

**Keywords:** *Catchment model, integrated model structure, nitrogen, diffuse pollution, water quality*

## 1. INTRODUCTION

The European Union (EU) adopted the Water Framework Directive (WFD) in 2000 which aims to ensure that all aquatic ecosystems meet ‘good status’ by 2015. However, the large population density in combination with agricultural and industrial activities in some European river basins pose challenges for river basin managers and prompts the need for innovative rehabilitation technologies. The EU financed AQUAREHAB project (FP7) examines different innovative rehabilitation technologies for soil,



**Figure 1:** The framework for model development in the AQUAREHAB project (EU FP7). The conceptual model in PCRaster/Python presented in this paper translates nitrogen emissions to loads in the aquatic ecosystem taking into account potential remediation measures.

groundwater and surface water to cope with a number of priority contaminants (nitrates, pesticides and chlorinated compounds). These innovative technologies will be compared to conventional measures in a generic river basin management tool that assesses the ecological and economical impact of rehabilitation technologies on the water system at both local and regional scales (Figure 1). For this purpose, a numerical spatio-temporal model is needed that translates emissions in the river basin to loads in the aquatic ecosystem taking into account potential remediation measures. The results of the scenario analysis will be used to derive knowledge rules translating nitrogen emissions and reductions thereof to loads in the aquatic ecosystem to readily verify the effect of changes in river basin management.

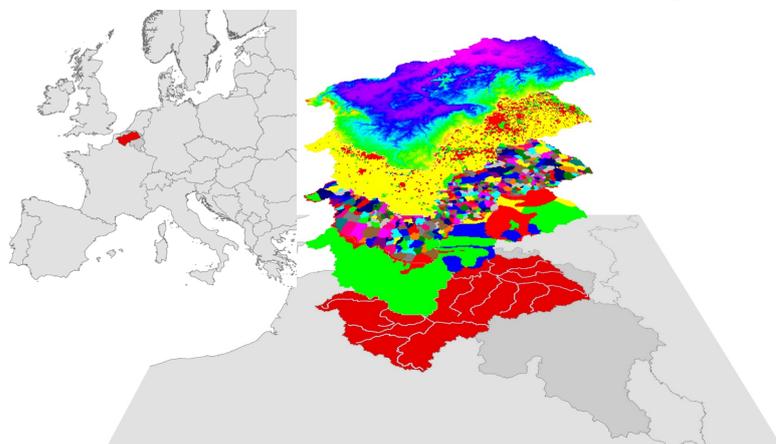
A variety of watershed models calculate chemical fate at the river basin scale by integrating fluxes of chemicals that enter the river system via multiple point and diffuse sources (Holvoet *et al.* 2007; Neitsch *et al.* 2005). However, most of these models do not explicitly describe the subsoil. A spatially and temporally explicit description of subsoil processes is useful to predict the long-term impact of groundwater contamination on the surface water quality. One option to achieve this goal is to combine models that describe the surface hydrology with models describing groundwater flow such as MODFLOW, *e.g.* Schmitz *et al.* 2009. Unfortunately, the coupling of these models and the application of MODFLOW to the large river basin scale is not straightforward. Another approach is to describe the river basin with a conceptual model that generalizes the large scale characteristics to such an extent that the data requirements of the model meet the available input data, *e.g.* minimum information requirement models (Quinn 2002). The authors favor this approach over the former mechanistic models since the conceptualization requires a careful consideration of the various processes in the river basin. In addition, such a data-driven model does not require a large number of parameters and often results in a reduction of the computing time which can prove useful for a scenario analysis of large sets of potential remediation measures in the river basin.

Data manipulation and process scripting can be a bottleneck for scientists with no advanced programming skills to apply the conceptual model approach to large river basins. A geographical information system (GIS) provides a useful environment for data manipulation and storage but most often does not provide dynamic modeling capabilities. The PCRaster modeling language provides the user with a set of analytical functions for manipulating raster GIS maps in time. It has already been used in a conceptual model simulating nitrogen dynamics in some European river catchments (de Wit *et al.* 2000; Mourad *et al.* 2006). PCRaster was recently made available in a Python environment (Derek Karszenberg *et al.* 2010) providing ample opportunities for the user to incorporate other existing process descriptions or define his own functions in the NumPy package for scientific computing with Python. This paper illustrates how the versatile Python/PCRaster model environment is used to build a conceptual model that translates emissions in the river basin to loads in the aquatic ecosystem using spatially explicit data when available and taking into account potential remediation measures.

## 2. MODEL DESCRIPTION

### 2.1. The Scheldt basin

The Scheldt basin covers almost 22000 km<sup>2</sup> and is located in Belgium, France and The Netherlands (Figure 2). It is a typical lowland catchment with a maximum altitude of 205 m a.s.l. The climate is temperate with an average monthly temperature ranging from 2 to 17 °C and an average precipitation of 845 mm/y (ECA&D 2008). The soil type varies from loam to sand in the northern direction and the subsoil mainly consists of consolidated sedimentary material in the southern part of the basin and of sands in the northern



**Figure 2:** The location of the Scheldt basin in Western Europe with an illustration of available data in a GIS environment, from top to bottom: the digital elevation, the landuse, the wastewater treatment zones, the aquifer texture and the location of the large rivers.

part. The Scheldt basin is 22% industrialized and urbanized with over 10 million inhabitants. In addition, 70% is classified as agricultural land and only 8% of the basin is classified as forest or nature reserve in the CORINE 2005 dataset (EEA 2008). The Scheldt River runs through the basin over a total length of 355 km. The Scheldt estuary in The Netherlands was not considered in this research due to the strong tidal influence. The studied catchment arbitrarily ends at the Belgian-Dutch border with an average non-tidal flow in the range of 100 m<sup>3</sup> s<sup>-1</sup> near Antwerp.

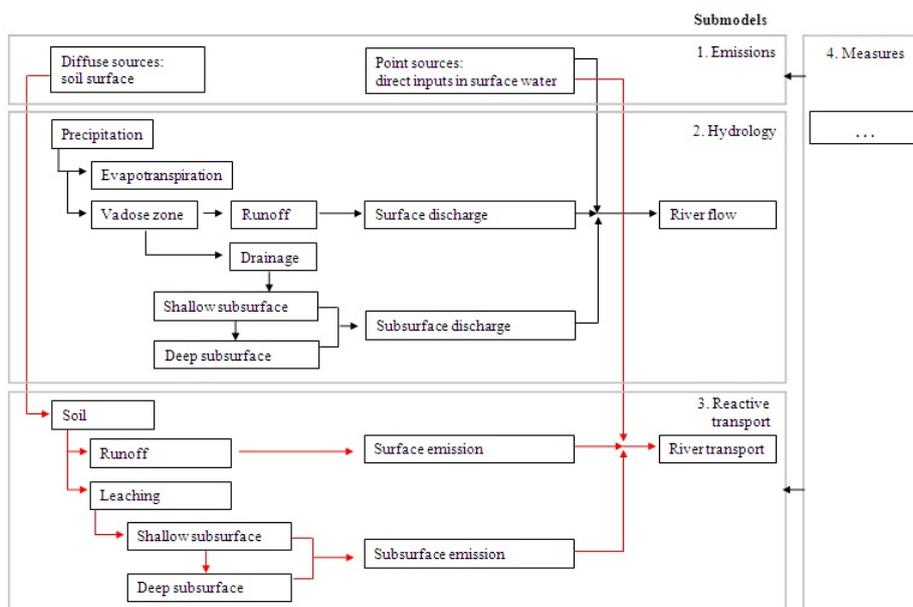
### 2.2. The conceptual model

The conceptual model is a compromise between model complexity, the available data and the applicability to scenario analysis. Since the concentrations of the selected pollutants are measured with a monthly time interval in the Scheldt basin, this interval was preferred for the conceptual model time step. The spatial discretization was set to 1 km<sup>2</sup> due to the large size of the catchment and the large scale characteristics inherent to the conceptual model. The model consists of 4 modules: (1) Emissions, (2) Hydrology, (3) Reactive transport and (4) Measures as indicated in Figure 3. The modules were implemented as separate classes in Python and the respective processes were defined as methods within each class. This setup yields a flexible model environment for the user who can easily select processes of interest.

#### **Module 1: emissions**

Module 1 describes the diffuse and point source emissions of nitrogen in the river basin. The data were gathered from European datasets (EMEP, 2008; Eurostat, 2008) or were supplied by local government agencies (VMM, DGRNE, Agence de l'eau Artois Picardie). As such, the data describe different administrative scales and were homogeneously distributed over the specific area (artificial surfaces or agricultural areas) (EEA, 2008) at the concomitant spatial scale.

Emissions from households and industrial facilities are emitted directly in the river network after an eventual treatment if the sewer system is connected to a waste water treatment plant (WWTP). Diffuse emissions from atmospheric deposition and the net agricultural inputs are defined on the soil surface. The net agricultural inputs are calculated from the difference of the gross inputs minus the amount of nitrogen that is taken up by the crops and the amount of nitrogen that is removed from the farmstead by manure processing.



**Figure 3:** A scheme of the conceptual model with 4 modules that contain the different modeled processes.

### **Module 2: hydrology**

The hydrological module defines 4 compartments that can be physically interpreted: the soil, the shallow subsurface, the deep subsurface and the river network. The water movement through these compartments is described by different processes given in Figure 3. The direction of water movement through the compartments is pivotal in the conceptual model and is determined by the topography. Water movement occurs in the direction of the steepest slope in all model compartments and is characterized by the local drain direction (LDD) (Wesseling & van Deursen 1997). This abstraction of flow paths in the subsurface can be considered acceptable at a large scale, e.g. km<sup>2</sup>.

The large model scale and the relative uncertainty about the input data with regard to land use, soil and subsoil texture, and vegetative cover prompted the use of a conceptual model description. Several approaches were considered for the various processes: a simulation of evapotranspiration using Hamon's formula (Stephens *et al.* 1963) or alternatively a user-defined function that explicitly takes into account the vegetative cover; a description of runoff using the SCS equation (Curve Number method) (USDA 1986) or alternatively a calculation of the water balance at the soil surface with a probability distribution function similar to the semi-distributed Xinanjiang model (Zhao *et al.* 1980); and a description of groundwater flow according to the head gradient or alternatively a conceptual representation of groundwater movement by routing influenced by the topography and by the conductivity and porosity of the aquifer. A selection between these processes was made during the model calibration and will be discussed below.

The river network is spatially defined using the CCM River and Catchment Database provided by the European Commission (Vogt *et al.* 2007). The lateral inputs from surface runoff and subsurface drainage are routed to the outlet of the model basin in one timestep since the residence time in the river network of the Scheldt basin before the Dutch border is smaller than the model timestep of one month. The amount of water that deviates to or originates from canals was taken into account by subtracting or adding that fraction in a specific river cell while accumulating the volumes of water through the river network.

### **Module 3: reactive transport**

The reactive transport module translates the emissions in module 1 to loads in the aquatic ecosystem of the river basin, i.e. loads in the subsurface and in the river network. The emissions are transported by the water flow calculated in module 2 with eventual storage and reaction in each model compartment.

The soil storage is defined by the balance of diffuse emissions, Michaelis-Menten degradation as a function of soil carbon content, runoff and leaching of nitrogen from the soil compartment if the nitrogen content exceeds the soil storage capacity as a function of C/N content. The leachate enters the subsurface where it is homogeneously distributed in the grid cell and transported out relative to the fraction of water that moves through the grid cell per timestep. A zero<sup>th</sup>-order reaction is defined in the subsurface. Nitrogen enters the

river network from point source emissions, surface runoff and groundwater discharge. The mass of nitrogen is routed through the river network in each time step taking into account a first order degradation. This yields loads in the aquatic ecosystem that can be translated to average monthly concentrations using the concomitant water flows determined in Module 2.

#### **Module 4: measures**

The European Commission allows for the Member States to motivate a derogation from the objectives for 2015 by the next reporting period in 2021 or finally reach the objectives in 2027. These reporting periods can be used to evaluate measures in a dynamic model.

Conventional measures included in river basin management plans typically include wastewater treatment for households (collective or individual) and a reduction of agricultural emissions by reducing livestock, fertilizer use or manure processing. These measures are usually stimulated by the government and could thus be homogeneously distributed over the agricultural area. Wetlands are considered as an innovative remediation technology in the SQUAREHAB project since it is also effective in reducing nitrogen loads from historical emissions that enter the river network from the subsurface. The implementation of wetlands will be spatially distributed to zones in the river basin that experience a large amount of nitrogen in the surface runoff or that are affected by historical loads that enter the river network from the subsurface.

### **2.3. Model calibration**

The conceptual model describes the nitrogen dynamics with a minimum of calibration parameters to prevent an eventual masking of conceptual errors by over-parameterization. The defined parameters were calibrated using a multi-objective optimization algorithm to evaluate the different responses of the various subbasins and, as such, verify if the spatial heterogeneity in the basin is adequately incorporated in the conceptual model. The AMALGAM model was selected for this purpose and the MATLAB code was kindly obtained from Dr. Vrugt (Vrugt & Robinson 2007). The model calibration is performed on a cluster with up to 200 computing nodes since the conceptual model takes about 25 seconds runtime per simulated year, the calibration requires a 50-year simulation, and a multi-objective optimization usually requires >1000 model simulations ( $\approx$  15 days on a desktop). The simulation time could be reduced by optimizing the model code but this would preclude the software's flexibility for users with no advanced programming skills to test different process descriptions.

## **3. RESULTS AND DISCUSSION**

### **3.1. Data processing**

The input data were obtained from different sources (see above) and, as such, often in different formats. The open source GDAL/OGR package ([www.gdal.org](http://www.gdal.org)) was used in this work to convert data to a common spatial reference and rasterize GIS data that were supplied in vector format. The PCRaster library in Python was used to automate the spatial distribution of numerical and statistical data in relation to landuse, soil texture, precipitation etc. However, it was found that the coupling of the available input data from different sources required special attention before automating the subsequent procedures. For example: the digital elevation model needed to be adjusted temporarily to force the local drain direction network towards the discerned rivers at the scale of a square km. Another issue was inconsistent georeferencing of data from different sources that referred to the same model object, e.g. rivers and monitoring stations.

### **3.2. The conceptual model**

A general description of evapotranspiration using Hamon's formula was inadequate to describe the different temporal responses at various locations in the catchment. It could not well describe large discharges during spring and low discharges in summer. Therefore, an alternative empirical function was developed based on Hamon's formula with a further spatial distribution based on the vegetative cover and the growing stage to simulate the effect of the agricultural practices that account for 70% of the catchment's surface.

The empirical SCS equation was derived for daily runoff and it could not be easily adapted to approximate runoff from monthly precipitation data. The probability distribution function in the Xinanjiang model uses the soil moisture to derive runoff from the amount of rainfall for a catchment and was better suited for this conceptual model. It was fully distributed by calculating the soil water balance in each grid cell taking into account the soil texture and percentage of covered surface.

The explicit description of groundwater flow is not straightforward in a conceptual model and its calculation by the head gradient was not compatible with the predefined direction of water movement

following the local drain direction. An alternative representation of groundwater was conceptualized: the recharge to the subsurface is fractionated between shallow and deep subsurface and is subsequently routed to the downstream river. The fractionation is influenced by the topography, and the conductivity and porosity of the aquifer. The water in the shallow subsurface is routed to the nearest waterway while the water in the deep subsurface moves to rivers with a Strahler number 2 or greater in this conceptual model. The fractionation and different routing paths simulate the fast response of a shallow aquifer to wet periods and the baseflow of the deep aquifer in the river basin.

The calibration of the reactive transport module required additional data on the current amount of nitrogen that is stored in the soil compartment. Preliminary results showed that the monitoring data were insufficient and that the optimization algorithm needed a constraint on the soil storage to obtain reliable parameters. Furthermore, the preliminary results suggested the inclusion of a temperature dependency in the degradation processes. A possible explanation for this observation is the nitrogen dynamics in the soil and the uptake of nitrogen by plants. As such, nitrogen mainly leaches from the soil during the colder winter months. The inclusion of a temperature dependency could be an alternative to a plant growth module. The adequacy of this conceptual representation will be evaluated in the ongoing calibration of the reactive transport module.

### 3.3. Model calibration

The calibration of module 2 of the spatially distributed conceptual model with a monthly timestep raised some questions with regard to the selection of the objectives and the evaluation of the results. The root mean square error of predicted versus observed flow was normalized to the standard deviation of the observed flow. This allows for an aggregation of the normalized root mean square errors (NRMSE) per subbasin with equal weight for each monitoring station. As such, 6 objective functions (OFs) were defined for the multi-objective optimization algorithm as the minimum of the aggregated NRMSEs per subbasin. This would allow for the optimization algorithm to neglect a subbasin with bad simulation results due to a conceptual error or data issues and would thus theoretically result in a better fit. However, a better outcome with a limited number of parameter evaluations was obtained by aggregating the NRMSE per flow regime distinguishing only 3 OFs. The selection of the main river as explicit objective (1 monitoring station) and the limitation of the objective space could have resulted in a faster convergence of the optimization algorithm to a parameter set that adequately simulates the overall processes in the river basin. This will be verified in further research.

The automatic calibration yielded a hydrologic model with a strong response of surface runoff to wet periods. This is attenuated at the larger scale but poorly simulates flow in the smaller rivers. The resulting Nash-Sutcliffe coefficient illustrates a bad fit of the medium and small-flow rivers due to the inherent emphasis on outliers (Table 1). The NRMSEs illustrate an overall error of 30 to 50% of simulated flows compared to the standard deviation of observed flows in the medium and small rivers. These results show that the selected objectives and the evaluation could be reconsidered to evaluate the hydrological module at the subbasin scale. For example by evaluating the relative contribution of each flow component (discharge, evaporation, runoff, and groundwater) in the selected objectives and by subsequently adapting the parameters of the dominant process. Alternatively, the non-linear interactions between the various processes in the conceptual model do not adequately simulate the observations. For example, the processes describing recharge and groundwater flow in the conceptual model could be adapted to lessen the dominant surface runoff. It can be concluded that module 2 adequately simulates the flow in the main river but should be reconsidered for predictions at the subbasin scale.

The observed limitations of module 2 could have an influence on the outcome of module 3, the reactive transport. But the uncertainty about the soil storage and the nitrogen application at the river basin scale due to the limited available datasets also adds to the uncertain outcome of module 3. The relative sensitivity of module 3 to these various sources of uncertainty needs to be assessed in further sensitivity analysis.

**Table 1:** The Nash-Sutcliffe and normalized root mean square error (NRMSE) values of module 2.

<i>Average flow [m<sup>3</sup> s<sup>-1</sup>]</i>	<i>Number of stations</i>	<i>Nash-Sutcliffe</i>	<i>NRMSE</i>
33.5	1	0.6	0.1
2.6	10	-1.9	0.3
0.6	7	-6.7	0.5

#### 4. CONCLUSION

The open-source model environment with Python/PCRaster enables users with no advanced GIS or programming skills to process heterogeneous data and to construct a conceptual model to describe water flow and water quality in a large river basin. However, the application of the conceptual model approach for scenario analysis in a fully distributed large river basin at a monthly time step proved difficult in this research. The model calibration showed the need for a good definition of the objectives and an eventual further evaluation of each objective to preferentially adapt the parameters of the dominant process. Alternatively, the non-linear interactions between the various processes in the conceptual model require a careful consideration of each process description. The open model structure allows for a continuous adaptation of the conceptual model during this evaluation. The resulting GIS-based model will be used for a first evaluation of innovative rehabilitation technologies versus conventional measures at the river basin scale, taking into account the current limitations for predictions at the subbasin scale.

#### ACKNOWLEDGEMENTS

This work was co-funded by the EU-project SQUAREHAB (FP7 – Grant Agreement Nr. 226565). The authors gratefully acknowledge the use of the various datasets in this work and the help of Stijn Van Looy (VITO, Belgium) with the cluster computing.

#### REFERENCES

- de Wit, M.J.M. et al., 2000. Modelling water fluxes for the analysis of diffuse pollution at the river basin scale. *Hydrological Processes*, 14(10), pp.1707-1723.
- ECA&D, 2008. European Climate Assessment & Dataset, Royal Netherlands Meteorological Institute. Available at: <http://eca.knmi.nl/>.
- EEA, 2008. CORINE Land Cover. *Coordination of information on the environment*. Available at: <http://reports.eea.europa.eu/COR0-landcover/en>.
- EMEP, 2008. European Monitoring and Evaluation Programme (EMEP) of the Convention on Long-range Transboundary Air Pollution. Available at: <http://www.emep.int>.
- Eurostat, 2008. Geographic Information System GISCO, Statistical Office of the European Communities, Luxembourg. Available at: <http://epp.eurostat.ec.europa.eu>.
- Holvoet, K., Seuntjens, P. & Vanrolleghem, P.A., 2007. Monitoring and modeling pesticide fate in surface waters at the catchment scale. *Ecological Modelling*, 209(1), pp.53-64.
- Karssenberg, Derek et al., 2010. A software framework for construction of process-based stochastic spatio-temporal models and data assimilation. *Environmental Modelling & Software*, 25(4), pp.489-502.
- Mourad, D.S.J., Van Der Perk, M. & Piirimäe, K., 2006. Changes in nutrient emissions, fluxes and retention in a north-eastern European lowland drainage basin. *Environmental monitoring and assessment*, 120(1-3), pp.415-48.
- Neitsch, S.L. et al., 2005. Soil and Water Assessment Tool theoretical documentation. Report 02-01. , p.494.
- Quinn, P., 2002. Models and monitoring: scaling-up cause-and-effect relationships in nutrient pollution to the catchment scale. In *Agricultural effects on ground and surface waters: Research at the edge of science and society*. IAHS Publ. no. 273, pp 397-403.
- Schmitz, O. et al., 2009. Linking external components to a spatio-temporal modelling framework: Coupling MODFLOW and PCraster. *Environmental Modelling & Software*, 24(9), pp.1088-1099.
- Stephens, J.C. & Stewart E.H. 1963. A comparison of procedures for computing evaporation and evapotranspiration. *Proceedings of the General Assembly of Berkeley*. IAHS Publ. no. 62, pp 123-133.
- USDA, 1986. Urban Hydrology for Small Watersheds. *TR-55, USDA, USA*, p.164.
- Vogt, J. et al., 2007. *A pan-European River and Catchment Database*, European Commission - JRC, Luxembourg. Available at: <http://agrienv.jrc.it/activities/catchments>.
- Vrugt, J.A. & Robinson, B.A., 2007. Improved evolutionary optimization from genetically adaptive multimethod search. *Proceedings of the National Academy of Sciences of the United States of America*, 104(3), pp.708-11.
- Wesseling, Cees G & van Deursen, W.P.A., 1997. Rhine Basin. In S. Hodgson, M. Rumor, & J. J. Harts, eds. *Proceedings Third Joint European Conference Exhibition on Geographical Information*. Vienna, Austria, pp. 487-496.
- Zhao, R.-jun et al., 1980. The Xinanjiang model. In *Hydrological forecasting (Proceedings of the Oxford symposium)*. Oxford: IAHS Publ. no. 129, pp. 351-356.