PROPAGATION OF UNCERTAINTY IN DIFFUSE POLLUTION INTO WATER QUALITY PREDICTIONS: APPLICATION TO THE RIVER DENDER IN FLANDERS, BELGIUM.

Vandenberghe Veronique*, Ann van Griensven**(***), Willy Bauwens** and Peter A. Vanrolleghem*

** Ghent University, Department of Applied Mathematics, Biometrics and Process Control (BIOMATH)*
Coupure Links 653, B-9000 Ghent, Belgium (E-mail: veronique.vandenberghe@biomath.rug.ac.be)
*** University of California Riverside, Dept. Environmental Sciences (Geology building), Riverside CA92521, USA*
**** Free University of Brussels, Laboratory of Hydrology and Hydraulic Engineering, Pleinlaan 2, B-1050 Brussels, Belgium*

ABSTRACT
In this paper the uncertainty of water quality predictions caused by uncertainty on the inputs related to emissions of diffuse pollution is detected and analysed. An uncertainty analysis on the effects of diffuse pollution is essential to compare the cost and benefits of measures to lower those emissions. In this research we focus on the effects of diffuse pollution due to fertiliser use on nitrate in the river water. With the use of an efficient Monte Carlo method based on Latin Hypercube sampling, the contribution to the overall uncertainty by each of the inputs is calculated. The modelling environment used is ESWAT, an extension of SWAT, the Soil and Water Assessment Tool, which is developed at the Blackland Research Center, Texas. ESWAT was developed to allow for an integral modelling of the water quantity and quality processes in river basins. The diffuse pollution sources are assessed by considering crop and soil processes. The crop simulations include growth and growth limitations, uptake of water and nutrients and several land management practices. The in-stream water quality model is based on QUAL2E. The spatial variability of the terrain strongly affects the non-point source pollution processes. GIS (Geographical Information System(s)) is used to account for the spatial variability. The methodology is applied to the Dender basin in Flanders, Belgium. A sensitivity analysis showed eight inputs that have significant influence on the time that the nitrate content in the river is higher than 3 mg/l. It is the start date of growth of pasture, the amount of fertiliser applied on pasture, on farming land and on corn depending on the subbasin. The uncertainty analysis indicated wide uncertainty bounds (95% percentile bounds differs up to +/- 50 % from the average NO_3 predictions). It was seen that in periods of dry weather no uncertainty is propagated into the results of nitrate in the river water. Consequently those periods are useful calibration periods for the point pollution inputs.

Keywords: Diffuse pollution, integrated river basin modelling, river water quality, uncertainty analysis

INTRODUCTION
Recently the European Union has approved the EU Water Framework Directive (EU-WFD). This directive claims that by the end of 2015 a “good status of surface water “ and a “good status of groundwater” should be achieved (EU, 2000). To that end, the EU-WFD provides several guidelines for monitoring the water bodies, leaving the practical implementation to the local governments. To make sure that the new water policy will succeed, a profound analysis of the actual and future state of the water is necessary. In this context, the evaluation of emissions into river water will be important. The effects of a pollution load into the river can be evaluated using models. However, for several reasons those model outputs are uncertain (Beck, 1987). Model outcome uncertainties can become very large due to:

- input uncertainty,
- model uncertainty,
- uncertainty in the estimated model parameter values and
- mathematical uncertainty.

Pollutants discharged into the river at certain points can be measured by monitoring the discharge and the water quality. Diffuse pollution sources on the other hand are distributed along the river and are very difficult to monitor. Therefore, estimating and calculating the diffuse pollution to a river can be subject to large input uncertainties. To optimally allocate the efforts necessary to reduce those input uncertainties, it is useful to evaluate the sensitivity of the outputs to the different inputs.

In this study we focus on the diffuse pollution of nitrate in the water due to fertiliser use. With the use of an efficient Monte Carlo method based on Latin Hypercube sampling (McKay, 1988), the contribution to the uncertainty by each of the inputs is calculated. The methodology is applied to the Dender basin in Flanders, Belgium.
This paper is structured as follows: first we discuss the model environment and give a description of the river basin for the case study. Then the applied methodology, which consists of a sensitivity and uncertainty analysis, is explained. The next section gives and discusses the results and finally conclusions are formulated.

THE MODELLING ENVIRONMENT ESWAT
ESWAT is an extension of SWAT (van Griensven and Bauwens, 2000), the Soil and Water Assessment Tool developed by the USDA (Arnold et al., 1998). ESWAT was developed to allow for an integral modelling of the water quantity and quality processes in river basins. The diffuse pollution sources are assessed by considering crop and soil processes. The

crop simulations include growth and growth limitations, uptake of water and nutrients and several land management practices. The in-stream water quality model is based on QUAL2E (Brown and Barnwell, 1987). The spatial variability of the terrain strongly affects the non-point source pollution processes. GIS (Geographical Information System(s)) is used to account for the spatial variability. Based on soil type and land use a number of Hydrological Response Units (HRU's) can be defined. For each HRU, the ESWAT model simulates the processes involved in the land phase of the hydrological cycle, and computes runoff, sediment and chemical loading. Based on the areas of each HRU, the results are then summed for each subbasin.

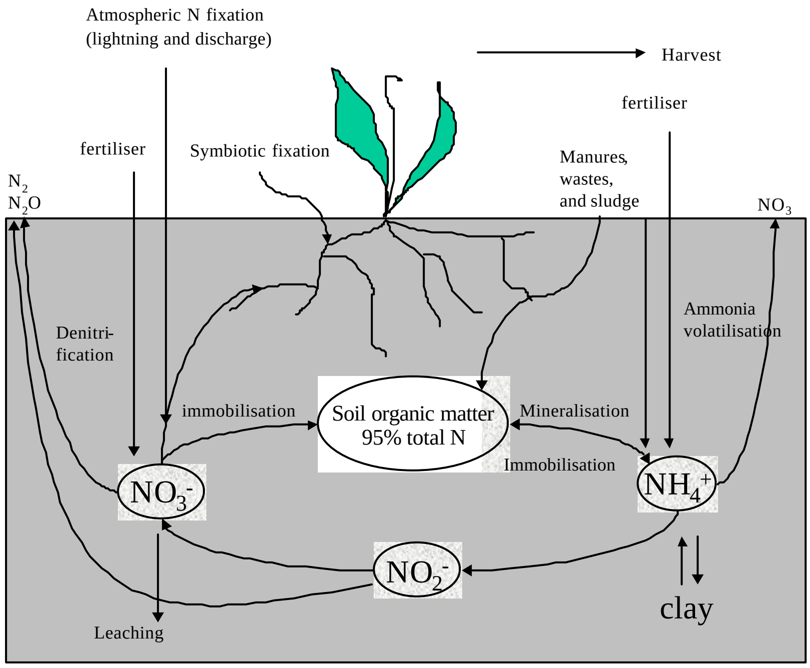


Figure 1. The nitrogen cycle in ESWAT

In terms of the nitrogen cycle, the three major forms in mineral soils are organic nitrogen associated with humus, mineral forms of nitrogen held by soil colloids and mineral forms of nitrogen in solution. Nitrogen may be added to the soil by fertiliser, manure or residue application, fixation by symbiotic or nonsymbiotic bacteria and rain. Nitrogen is removed from the soil by plant uptake, leaching, volatilization, denitrification and erosion. Figure 1 shows the major components of the nitrogen cycle.
Nitrate is an anion and is not attracted to or sorbed by soil particles. Because retention of nitrate by soils is minimal, nitrate is very susceptible to leaching. In SWAT the algorithms to calculate nitrate leaching simultaneously solve for loss of nitrate in surface runoff and lateral flow. Finally nitrate ends in the river.

THE DENDER IN FLANDERS, BELGIUM

The catchment of the river Dender has a total area of 1384 km^2 and has an average discharge of 10 m^3/s at its mouth. Figure 2 shows how the Dender basin is situated in the Flanders. As about 90% of the flow results from storm runoff and the point sources make very little contribution, the flow of the river is very irregular with high peak discharges during intensive rain events and very low flows during dry periods (Bervoets et al., 1989). The river Dender is heavily polluted. Part comes from point-pollution (e.g. industry) but also from diffuse sources of pollution originating mainly from agricultural activity. Although there is an unmistakable relation between intensive agricultural activity and the occurrence of high nutrient concentrations in the environment, few precise data are available about the contribution of agricultural activity to the total nutrient concentrations. The data needed for the model implemented in ESWAT were also very sparse and conversions had to be made to make the data useful for the model (Smets, 1999).

To build the model, the total catchment was subdivided into 16 subbasins (Figure 3). Input information for each subbasin is grouped into categories for unique areas of landcover, soil and management within the subbasins. Main soil classes are sand, loamsand, silty loam and impervious areas. For landuse, 5 classes are important: impervious areas, forests, pasture, corn (maize and corn) and land for common agricultural use (crop culture, not corn). About 30% of the landuse is pasture, while crop farming represents ca. 50% of the landuse.

METHODOLOGY

A previous study for the same kind of nitrogen leaching model (implemented in SWIM) from arable land in large river basins (Krysanova and Haberlandt, 2001) showed that the relative importance of natural and antropogenic factors affecting nitrogen leaching in the Saale river basin was as follows: (1) soil, (2) climate (3) fertilisation rate and (4) crop rotation. Reducing the uncertainty on inputs for soil and climate depends on better equipment to measure the different variables and proper use of sophisticated mathematical techniques to interpolate for places that are not measured. A lot of studies on that subject already exist (Sevruk, 1986). Until recently reducing the input uncertainty relating to fertilisation rate was not studied. In Flanders new legislation concerning fertilisation application was made in the late nineties. Campaigns to list the fertiliser use were then started and it is known that still a large amount of information is wrong or missing. A lot of effort is still needed to complete the information. The evaluation and quantification of the impacts of land management practices on nitrogen wash-off to surface water is therefore very important.

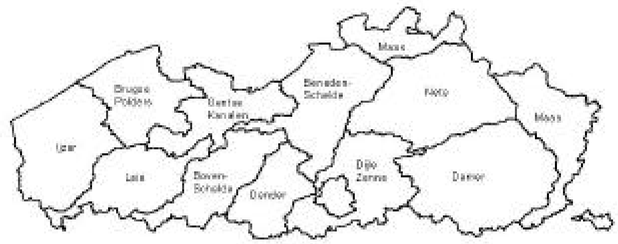


Figure 2. The river Dender basin in Flanders

*Figure 3. The river Dender Basin, Denderbelle (**o**) and the subbasins)*

In this study we focus on fertilisation rate and time of fertiliser application on the most important crops for the Dender river basin. For the application of fertiliser for the different land uses three application dates were assumed; 1st of March, 1st of April and 1st of May. Also operations as planting and harvesting dates can be defined. Day and months were used to specify the planting and harvesting dates. Of course, those dates depend on climate, crop and farmer. So assumptions had to be made concerning those dates. The output that was focussed upon was the time that nitrate concentrations in the river Dender at Denderbelle (near the mouth) were higher than 3 mg/l.

Data on fertiliser and manure use are provided by The Flemish Institute for Land Use (Vlaamse Landmaatschappij,VLM). They provided data on the nutrient use and production for each municipality in Flanders. In SWAT, one has to specify for each subbasin the total amount of fertiliser and the detailed composition of the fertiliser. Some conversions of the supplied data had to be made so that they could be used in ESWAT. It consisted of recalculations of the application rates for each municipality to application rates per subbasin (Smets, 1999). Further the same amount of fertiliser on all crops was assumed for this model. This is clearly different from practice, but at this stage, insufficient details are available to specify this more realistically.

First, a global sensitivity analysis is used to show what the most important factors are. Then the influence of uncertain data on the river NO_3 concentrations time series is evaluated with an uncertainty analysis. For both the same Monte Carlo sampled inputs could be used. The sensitivity analysis focuses on the inputs, to rank their importance while the uncertainty analysis assumes uncertain inputs and only considers and evaluates the outputs.

Sensitivity analysis

The sensitivity analysis (SA) technique used here is based on a multi-linear regression of the inputs on a specific output. A Monte Carlo technique, Latin Hypercube sampling makes sure that the total range of inputs is covered. When the number of samples equals 4/3 times the number of inputs such a sampling is sufficient to perform a reliable SA (McKay, 1988). For the sampling of the inputs and the analysis of the outputs, a program was written to couple UNCSAM, the program used for the SA (Janssen, 1992) with the management input files of ESWAT. We used the standardised regression coefficient (SRC) as an indication of the relative importance of the different inputs.

$SRC_i = \dfrac{\Delta y / S_y}{\Delta x_i / S_{x_i}}$ with $\Delta y / \Delta x_i$ = change in output due to a change in an input factor and S_y, S_{x_i} the standard deviation of respectively the output and the input. The input standard deviation S_{x_i} is specified by the user.

The ordering of importance of the input factors based on that statistic is as good as the associated model coefficient of determination R^2 of the whole multi-linear regression. The closer R^2 is to 1, the better the results.

When the input variables are linearly related, the application of a linear regression can lead to an accuracy problem, the collinearity problem (Hocking, 1983). The Variance Inflation Factor (VIF) is considered to be a useful measure to detect collinearities. The VIF is defined as:

$VIF_i = [C_x]_{ii} = (1 - R_i^2)^{-1}$ where R_i^2 = the R^2 value that results from regressing y on only x_i.

For every subbasin the total amount of fertiliser and the time of planting and harvesting of crops have to be given to the model. A sensitivity analysis can now be performed to evaluate the influence of those inputs on the model results for NO_3 in the river water. We evaluate the sensitivity of the model on the following result: the time that NO_3 is higher than 3 mg/l. The fractions of mineral N, organic N, and NH_3-N in the fertilisers are considered to be known and fixed. Hence, we only analyse the total amount of fertiliser used (Table 1). As there are a lot of differences in management practices between the different farmers and the time of planting and harvesting is different from year to year, the plant date and harvest date for the crops are also considered in the analysis. For a global sensitivity analysis we take the uniform distribution with standard deviation S_{x_i}. The ranges of the uniform distributions are given in Table 2. We assumed no correlation. To supply the information on those ranges a few farmers living in Maarkedal (situated in the Dender basin) were interviewed about their land management practices.

Table 1. Composition of the manure as input in SWAT

Chemical	**Percentage of total fertiliser (100*kg/kg)**
HNO3	28.5%
Mineral P	7.5 %
Organic N	28%
Organic P	7.5%
Ammonia	28.5 %

Table 2. Ranges for global sensitivity analysis of management practice inputs for Nitrogen

Input	**Uncertainty**
Plant date for the crops	+/- 1 month
Harvest date of the crops	+/- 1 month
Amount of fertiliser applied per subbasin and per crop (kg/ha)	+/-25%

Uncertainty analysis

The Flemish Institute for Land Use provided input data for the model. Due to unregistrated manure and fertiliser use it is very likely that those data are underestimated. The amount of fertiliser was the same for all crops which is unrealistic and it is very likely that the composition of the fertiliser used is not always the same as was assumed here. The uncertainty of the latter two generalisations is caught by the uncertainty in the amount of fertiliser used on the crops. Such an uncertainty is propagated through the model and gives a final uncertainty on the model results of the water quality near the mouth of the river. An uncertainty analysis in which all of the uncertain sources are varied at the same time is performed to see the effects of the uncertainty of the inputs. For this analysis we calculate the uncertainty bands i.e. the 5% and 95% percentiles for the results of the time series obtained after sampling the inputs as also done in the sensitivity analysis.

RESULTS AND DISCUSSION

Sensitivity analysis

As the used sensitivity analysis technique is based on linear regression, two measures are calculated to see whether a linear regression is acceptable. The first measure is the Regression Coefficient (RC) which was 0.845 for the whole multilinear regression. Because this value is close to 1 and the F-statistic showed that the regression is significant, a linear regression is adequate. The value of 0.845 means that there is a fraction of the output variance, 15.5%, that is left unaccounted for.

The second measure is the Variance Inflation Factor (VIF). The largest VIF for this analysis was 1.35. A VIF smaller than 5 means that the correlation between the inputs is small enough to allow application of a linear regression (Janssen et al, 1992). The standardised regression coefficient is significant on the 10% level for eight parameters. In table 3 the parameters are ranked

Table 3. The eight most important inputs, their SRC and their sensitivity ranking.

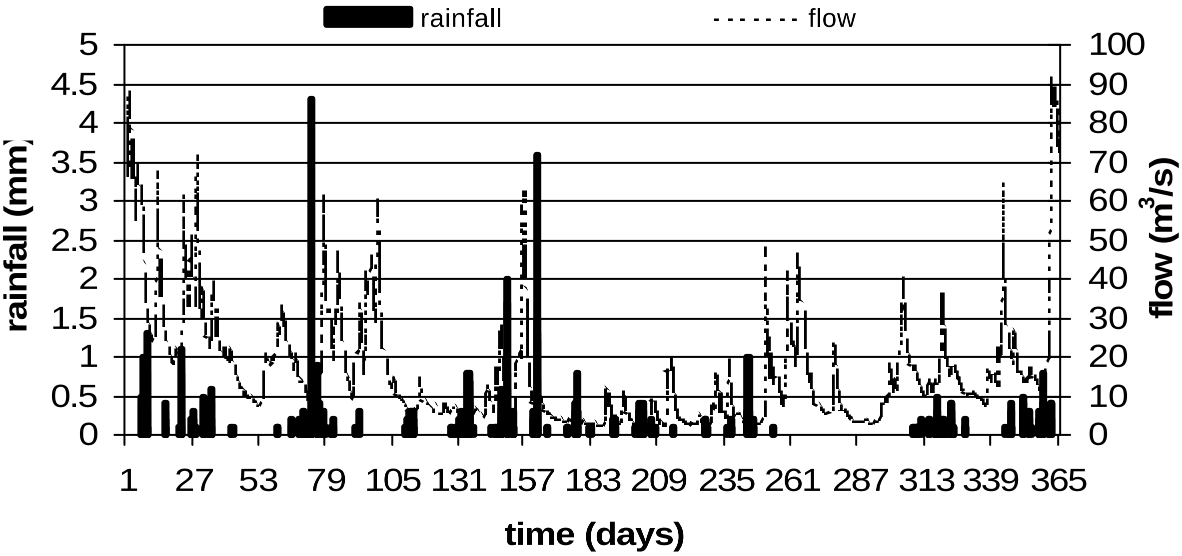
Input	SRC	Rank
Amount of fertilisation on pasture in subbasin 16	-0.303	1
Amount of fertilisation on farming land in subbasin 4	0.226	2
growth date of pasture	-0.183	3
Plant date on farming land	0.171	4
Amount of fertilisation on corn in subbasin 5	-0.169	5
Amount of fertilisation on corn in subbasin 15	-0.165	6
Amount of fertilisation on pasture in subbasin 12	0.159	7
Amount of fertilisation on corn in subbasin 11	0.155	8

For river nitrate concentrations the amount of fertiliser used in the subbasins that are laying upstream are especially important. This SA also shows that it is more important to focus on the amount of fertiliser than on the management practices.

Uncertainty analysis

Figure 4 illustrates the contribution of an uncertainty up to 25% to the amount of fertiliser applied on the crops and an uncertainty of one month for the plant and harvest dates. The 95 % bound shows much higher peaks than the mean concentrations time series. This means that some peak values of nitrate in the river water at Denderbelle may not be predicted properly due to an underestimation of the amount of fertiliser used. Those peaks (eg. day 156 and 260) are above levels of nitrate concentrations for basic water quality.

Only a few measurements were available for the calibration of the model. As can be seen, some of the measuring points are available at time instants were the uncertainty on the inputs is not propagated. Fortunately, the model calibration was successful. The data around day 340 are within the uncertainty bound, but the ones on day 240 are clearly not predicted well.



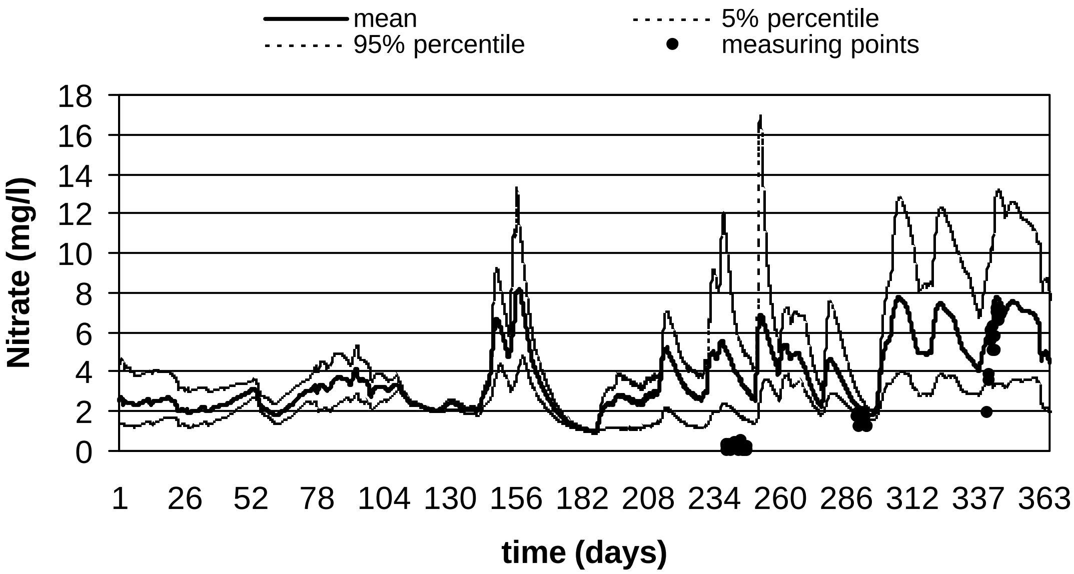


Figure 4. The simulation of nitrate and flow in the Dender river with measurements of nitrate (bottom) and rainfall (top) at Denderbelle and the 5 and 95% confidence intervals.

The inputs we considered uncertain in this study are only a fraction of all the inputs that cause uncertainty, but it is shown that the inputs related to fertiliser use already give a large uncertainty in certain periods of the year (95% percentile bounds differs up to +/- 50 % from the average NO_3 predictions). Those are periods with rainfall and high flows. Knowing that there are more sources of uncertainty related to diffuse pollution, the input uncertainty is expected to be even higher. Therefor it is clear that care has to be taken during the gathering of data needed to run a dynamic process-based model for the prediction of effects of diffuse pollution.

This uncertainty analysis shows also some important results for future measurement campaigns. This study learns that we could obtain a better calibration for the diffuse pollution part of the model with data that were taken during periods with rainfall and high flows, because the model output nitrate is more sensitive towards inputs of diffuse pollution in those periods. If you want to focus on calibrating the in-stream behaviour and point pollution then measurements during dry periods are needed as the model is then not sensitive towards input of diffuse pollution.

CONCLUSIONS

The model output nitrate in the river water is sensitive towards eight inputs related to fertiliser use. It are the start date of the growth of pasture, the amount of fertiliser applied on pasture in subbasin 16 and 12, on farming land in subbasin 4 and on corn in subbasin 5, 11 and 12. The dates of planting and harvesting the crops appear not to be important. To reduce the output uncertainty it is best to focus on data about fertiliser amounts applied on crops.

In the uncertainty analysis, the effects of uncertain input of fertiliser use and land management on nitrate concentrations in the river are shown. In certain periods of the year the uncertainty bounds are very wide. Because there are more sources of uncertainty, not considered in this study, it becomes clear that it is very important to gather accurate data to run a dynamic process-based model for the prediction of effects of diffuse pollution.

The uncertainty analysis is also of great use for experimental design. Measurements during dry periods can be used to better calibrate the model for point source pollution because the inputs of diffuse pollution are not important then. On the other hand, periods with rainfall and high flows are needed for the calibration of the model with diffuse pollution because the model output nitrate is then very sensitive towards the inputs related to farmer's practices.

More detailed studies are needed to see the exact contribution of the input uncertainties to the output uncertainty but we can already conclude on the basis of this study that uncertainty analysis is an essential part of diffuse pollution modelling to evaluate and draw conclusions from model predictions.

ACKNOWLEDGEMENTS

The authors gives special thanks to the financial support of the EU Harmoni-Ca project (EVKI-CT-2002-20003).

LITERATURE

Arnold J.G., Williams J.R., Srivnivasan R. and King K.W. (1996). *SWAT manual*. USDA, Agricultural Research Service and Blackland Research Center, Texas.

Beck, M.B. (1987). Uncertainty in water quality models. *Wat. Res. Res.* , **23**(8), 1393.

Bervoets L., Schneiders A. and Verheyen R.F. (1989). *Onderzoek naar de verspreiding en de typologie van ecologisch waardevolle waterlopen in het Vlaams gewest*. Deel 1 - Het Denderbekken, Universitaire Instelling Antwerpen, 76p.

Brown, L.C. and Barnwell T.O. (1987). *The Enhanced Stream Water Quality Models QUAL2E and QUAL2E-UNCAS*: Documentation and User Model, EPA/600/3-87/007, USA.

Janssen P.H.M., Heuberger, P.S.C. and Sanders S. (1992). *Manual Uncsam 1.1., a software package for sensitivity and uncertainty analysis*. Bilthoven, The Netherlands.

Krysanova V. and Haberlandt U. (2001). Assessment of nitrogen leaching from arable land in large river basins. Part I. Simulation experiments using a process-based model. *Ecol. Modelling*,**150**, 255-275.

McKay M.D. (1988). *Sensitivity and uncertainty analysis using a statistical sample of input values*. In: Uncertainty analysis, Y. Ronen, ed., CRC Press, Inc., Boca Raton, Florida, 145-186.

Sevruk , B. (1986). Correction of precipitation measurements. *Proceedings ETH, IAHS International Workshop on the Correction of Precipitation Measurements*, 1-3 April 1985, ETH Zürich, Züricher Geographische Schriften , 23.

Smets S. (1999). *Modelling of nutrient losses in the Dender catchment using SWAT.* Masters dissertation. Katholieke Uuniversiteit Leuven – Vrije Universiteit Brussel, Belgium

Vandenberghe V., van Griensven A., Bauwens W. (2002). Detection of the most optimal measuring points for water quality variables: Application to the river water quality model of the river Dender in ESWAT. *Wat.Sci.Tech.*, **46**(3).

Van Griensven A. and Bauwens W. (2000). *Integral modelling of catchments. Wat.Sci.Tech,*, **43**(7), 321-328.

Willems P.(2000). *Probabilistic immission modelling of receiving surface waters*. PhD thesis, Katholieke Universiteit Leuven, Faculty of Applied Sciences, Leuven, Belgium