Sensitivity and uncertainty analysis for river quality modelling

Mona Radwan, Patrick Willems and Jean Berlamont

ABSTRACT

Uncertainty analysis for model simulation is of growing importance in the field of water quality management. The importance of this concern is provided by recent public awareness over health risks from improper disposal of toxic wastes as well as by the continuing emphasis on risk assessment. The first step in the chain of risk assessment is the quantification of the error in predicting water quality.

In each mathematical modelling application, different uncertainties are involved. The uncertainty sources can be classified into different categories (in this study, as model-input uncertainty, model-structure uncertainty, model-parameter uncertainty and measurement errors). These different types of uncertainty sources determine collectively the total uncertainty in the model results. In this paper, the relative contributions of uncertainties associated with each source are studied for the physico-chemical water quality modelling of a river in Belgium. This provides information as to where available modelling resources should be focused.

Key words | mathematical modelling, physico-chemical water quality, sensitivity analysis, uncertainty analysis

Mona Radwan
Patrick Willems (corresponding author)
Jean Berlamont
Hydraulics Laboratory,
Department of Civil Engineering,
Katholieke Universiteit Leuven,
Kasteelpark Arenberg 40,
3001 Heverlee,
Belgium
Tel.: +32 1632 1658;
Fax: +32 1632 1989;
E-mails: *mona_radwan2@yahoo.com;*
Patrick.Willems@bwk.kuleuven.ac.be;
Jean.Berlamont@bwk.kuleuven.ac.be

INTRODUCTION

Mechanistic modelling of physical systems is often complicated by the presence of uncertainties. River water quality modelling, for example, entails uncertainties in the estimates of pollutant emissions, transformation and transport parameters, etc., that impact on estimates of the water quality status of the river and of the related risks. The implications of these uncertainties are particularly important in the assessment of several potential regulatory options, for example, with respect to the selection of a strategy for the control of pollution levels. Even though significant effort may be needed to incorporate uncertainties into the modelling process, this could potentially result in providing useful information that can aid in decision-making.

A systematic uncertainty analysis provides insight into the level of confidence in model estimates and can aid in assessing how various possible model estimates should be weighed. Further, it can lead to the identification of the key sources of uncertainty (such as data gaps) which merit further research, as well as the sources of uncertainty that are not important with respect to a given response.

The purpose of quantitative uncertainty analysis is to use currently available information in order to quantify the degree of confidence in the existing data and models. The purpose is not to somehow 'reduce' uncertainty. Reduction in uncertainty can only come from gathering additional information and filling 'data gaps'. Even though the applicability of a model is limited by the model assumptions, the input data and the uncertainties in the evaluation data, understanding the judgements associated with the modelling process is more valuable than side-stepping the uncertainty analysis. In fact, it is precisely for problems where data are limited and where simplifying assumptions have been used that a quantitative uncertainty analysis can provide an illuminating role, to help identify how robust the conclusions about model results

are, and to help target data gathering efforts (Frey 1992). For each mathematical modelling application, different uncertainties are involved. In this study, the uncertainty sources have been classified into the following categories:

1. uncertainties of the model input variables: 'input uncertainties';
2. uncertainties of the model parameter values: 'parameter uncertainties';
3. uncertainties originating from the imperfect description of the physical reality by a limited number of mathematical relations: 'model–structure uncertainties';
4. uncertainties of the measurements to which the model results are compared: 'measurement uncertainty'.

The primary objective of the sensitivity and uncertainty analysis presented in this paper is to study the uncertainties associated with each of the above-mentioned uncertainty sources. Further, this can be used to identify the relative contributions of uncertainties associated with each source. This provides information as to where available resources should be focused, for example, increasing the amount of input data to calibrate the model parameters, filling data gaps, through more detailed measurement campaigns or by model-structure refinement.

THEORETICAL CONSIDERATIONS ON SENSITIVITY AND UNCERTAINTY ANALYSIS

The aim of *sensitivity analysis* is to estimate the rate of change in the output of a model with respect to changes in the model inputs or the model parameters. Such knowledge is important for (a) evaluating the applicability of the model, (b) determining parameters for which it is important to have more accurate values (by identifying the most sensitive parameters) and (c) understanding the behaviour of the system being modeled.

Uncertainty analysis aims to quantify the level of confidence the modeller has in the inputs, the parameters and the system representation by model equations (the model structure). As a result, the level of confidence in the model results can also be calculated. In this way, the

uncertainty analysis is the first step in the chain of risk assessment, by quantification of the error in predicting water quality.

Different uncertainty types

In each mathematical modelling application, different uncertainties are involved. The uncertainty sources can be classified into different categories, as mentioned before. The different types of uncertainty sources determine collectively the total uncertainty in the model output variables (and furthermore in the model results). Willems (2000) stated that this classification into different types is based on the different physical nature of the uncertainties and the model entities to which the uncertainties are related. Model input, model parameters and model structures indeed have a very different physical nature. Model input consists of directly measured data or is estimated. For water systems, they are essentially variable in time. This is in contrast to model parameters, which are considered as constant properties, and for which the values are measured, estimated or calibrated. The model structure is set up on the basis of scientific knowledge (from previous research) or is built on the basis of data available for the model-output variables (in data-based model-structure identification applications). Also the mathematical relations of the model structure are considered fixed and do not vary in time.

According to Willems (2000), model-input uncertainties have to do with measurement errors (if the model input is directly measured) or estimation errors (if the input is estimated). Also, for the model parameter values that are measured or estimated (on the basis of experience), the parameter uncertainties consist of measurement or estimation errors. For the parameters that are calibrated on the basis of measurements for the model-output variables, parameter uncertainties exist whenever erroneous and/or limited time series of measurements are used for calibration or whenever the calibration procedure is not optimal. An optimal parameter calibration is, however, only possible in a modelling application with an optimal model structure. For such a model structure, the parameter uncertainties decrease whenever a longer time

series of measurements is used for the calibration. Detailed physically based models are, however, often overparametrized. By this reason, it is often difficult to derive a unique or optimal set of parameter values.

Model-structure uncertainties have to do with the modeller's limitations to describe the physical reality perfectly. They can be considered as the remaining uncertainties after use of error-free input and measurements and after an optimal calibration. In the same way as the input and parameter uncertainties have to be considered for each input variable and each parameter (together with their correlations), the model-structure uncertainties can also be described for the different submodels (different mathematical relations or subsets of relations). The model-structure uncertainties can also be represented in a lumped way by describing the total model-output uncertainty explained by the model-structure uncertainties in a single model. However, the separate description for the different submodels (or the model-structure 'uncertainty decomposition') is only possible for models with a transparent model structure and calibration procedure. By uncertainty decomposition, a transparent uncertainty structure is derived. This allows the modeller to separately quantify and compare the different uncertainty-source contributions to the total uncertainty in the model output (see also Willems & Berlamont, 1999, 2002). In particular, the comparison between the uncertainties resulting from the data and the uncertainties resulting from the model structure is interesting.

Calculation of model output uncertainties

In the probabilistic analysis, the uncertainties associated with model inputs, model parameters and model structure have to be characterized and described by probability distributions. The objective is then to estimate the output probability distributions. This process comprises of two stages.

Determination and quantification of the contributing uncertainty sources

This process involves the determination of the errors (uncertainty levels) of the inputs and parameters. This is accomplished by using either statistical techniques or expert judgments. Statistical techniques involve estimating the errors in the model parameters after model calibration and the errors in the estimation of the inputs from available data or by collection of a large number of representative samples. In most detailed analyses, the full probability distributions of the errors and their correlations are calculated. In cases where limited data are available, an expert judgment provides the information about the error probability distribution. For example, a normal distribution is typically used to describe unbiased measurement errors.

Transfer to model-output uncertainties

In the second stage, the input probability distributions have to be transferred to output probability distributions. At the same time, they have to be combined with the model-structure and parameter uncertainties. For the latter application, a probabilistic modelling using stochastic terms can be used, as described in the next section.

The random variation due to natural variability (variation in time of the meteorology, the emissions, etc.) is considered by the simulation of long-time series in the model. By the statistical processing of these time series, probability distributions of the output concentrations can be calculated and used on the basis of decision-making in river water management. More details about the simulation of these time series and the statistical processing can be found in Radwan (2002). This paper focuses rather on the uncertainties involved in the modelling of the river water quality (at particular time moments).

Representation of uncertainties by stochastic terms

In a probabilistic model, the different uncertainty sources can be represented by so-called 'stochastic terms'. These terms take the form of random variables E, which are random in magnitude and time. The randomness in the magnitude is described by probability distributions, while the randomness in time is represented by autocorrelations of the time series. To transform the deterministic mathematical model into a probabilistic one, the stochastic
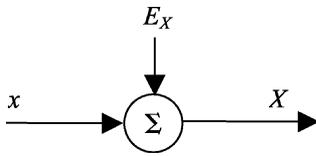
Figure 1 │ Schematic representation showing how the addition of a stochastic term E_X transforms a deterministic model variable x into a stochastic one.

Figure 2 │ Representation of the three different types of uncertainties in a probabilistic model (Willems 2000).

terms E_X are added to the model variables x to which the probability distributions are related (see also Figure 1):

$$X = x + E_X$$

In this way, the stochastic term represents an absolute error for the variable X. To represent a relative error, the stochastic term has to be applied to the logarithmically transformed variable $\ln(x)$.

A random or stochastic series X is derived by this operation. In this paper, capitals are used for random variables, while lowercase letters or symbols denote specific (measured or calculated) values of these variables.

Whenever the probability distribution of the magnitude equals the normal distribution N:

$$E_X \sim N(\mu(E_X), \sigma(E_X))$$

the two parameters, the expected value $\mu(E_X)$ and the standard deviation $\sigma(E_X)$ of the normal distribution, correspond to the systematic error (or 'bias' in the description of X) and the random absolute error. The absolute error $\sigma(E_X)$ may be constant or dependent on the value of X. In the latter case, a transformation of X can often be determined to reach a constant absolute error (error independent on X). Most flexible is the Box–Cox (BC) transformation (Box & Cox 1964):

$$BC(x) = \frac{x^\lambda - 1}{\lambda}$$

After application of this transformation to the model variables, the uncertainty can be represented by one single value: $\sigma(E_X)$. The BC transformation also has an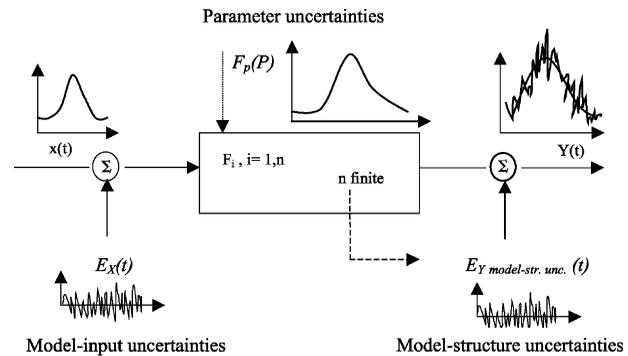 advantage for model parameter optimization. Using the mean squared error as an objective function to be minimized, equal weight will be given to both the lower and the higher values.

Referring to Figures 1 and 2, input uncertainties are represented by stochastic terms E_X added to the input variables x. The output of the stochastic term are time series of coloured noise that are added to the time series of the model input variables. In the same way, model-structure uncertainties are represented by the stochastic terms $E_{Y \mid \text{model-struc.unc.}}$ and added to the model output variables. Whenever the model-structure uncertainty is strongly decomposed (for different submodels), the stochastic terms may also be added to other (internal) state variables of the model. These state variables can be considered as output variables of submodels, but not of the complete model.

In comparison with the input uncertainties and the model-structure uncertainties, the stochastic terms of the parameter uncertainties have a totally different time-correlation structure. Because model parameters are considered as fixed values that do not change in time, maximum time correlation has to be assumed for their stochastic terms. This means that, whenever random simulations are performed by the probabilistic model, the random value at the output of a stochastic term representing parameter uncertainty is taken as constant during each simulation. The parameter stochastic terms thus have a full description on the basis of only the probability distribution $f_p(p)$ for the different model parameters P.

When modelling water systems, the variability in time of the internal state and model output variables (and, as a consequence, also the autocorrelation of the model-structure stochastic terms) is most dominantly determined by the rainfall input. This rainfall input consists of different successive rainfall events. The autocorrelation of the rainfall input stochastic terms and, as a consequence also of the model output errors, are often strong within one event, while it is rather weak between two rainfall events. In this case, the time series can be divided into a number of events or subperiods and inside each subperiod maximum correlation can be assumed. The stochastic terms can then be considered as discrete random variables. This means that the output of the stochastic terms is considered constant during each subperiod. The output only changes randomly between two subperiods according to the probability distribution of their magnitude and according to their time correlation structure (which can be neglected in some cases). In comparison with the more detailed continuous time description of the stochastic terms, this discrete description is a simplification of the uncertainty structure representation. As an implication of the simplification, the time-correlation structure of the stochastic terms is only considered for time lags larger than the time between succeeding subperiods. This may be different for different periods. In this way, subperiods can be considered as periods that contain runoff hydrographs with significant magnitude and which are separated by sufficiently long dry weather periods. For hydrological modelling, the dry weather periods are considered 'sufficiently large' if the high-frequency subflow discharges (surface runoff and interflow discharges) are nearly zero at the end of the periods. As the rainfall during these subperiods often corresponds with rain storms, these subperiods are also called 'storm periods'.

APPLICATION TO MOLENBEEK CASE STUDY

Introduction

The probabilistic methodology has been applied to the physico-chemical water quality (WQ) modelling of the Molenbeek brook in Belgium. The Molenbeek brook is one of the main tributaries of the River Dender and is located in the Flemish region of Belgium to the west of Brussels. It has an area of $57.44\,\text{km}^2$ and one liminigraphic station (hourly water level measurements and rating curve) is available at Mere with an upstream area of $40.5\,\text{km}^2$ (see Figure 3). It is a narrow catchment with relatively steep slopes. The upstream part is rural, while the downstream part is more urbanized (villages of Mere, Erpe and Hofstade). For this river, a Mike 11 model (DHI 2002) has been implemented for the conceptual hydrological modelling, hydrodynamic modelling and physico-chemical water quality (WQ) modelling.

Along the Molenbeek brook, the major sources of pollution are agricultural drainage, industrial and domestic wastewater. To estimate the agricultural nitrate leaching, a DRAINMOD-N model (Brevé *et al.* 1997) is used. A detailed description of each functional relationship and a model application in Flanders is given by El-Sadek *et al.* (2002). Its application for the Molenbeek brook is given by Radwan (2002). Along the Molenbeek brook there are 21 sewerage outlet pipes which receive domestic wastes. For each point the total population discharging into it is calculated and then the concentration of rough effluent was evaluated by assuming concentrations of BOD and ammonia equal to 54 and 10 g/(capita d). For industrial wastes, there is one factory located in the study area. The pollution load for different water quality parameters is estimated on the basis of measurements of the Flemish Environmental Agency (VMM) and assumed constant in time. More details about the WQ modelling (boundaries, model parameter calibration and results) can be found in Radwan (2002). This paper focuses on the sensitivity and uncertainty analysis for the WQ modelling application.

The final aim of the uncertainty analysis is to separate the total uncertainty in terms of the variance into the following terms:

$$\sigma^2_{Total} = \sigma^2_{HY+HD} + \sigma^2_{WQ}$$

$$\sigma^2_{WQ} = \sigma^2_{\text{model-input-unc.}} + \sigma^2_{\text{model-parameters-unc.}} + \sigma^2_{\text{model-structure-unc.}} + \sigma^2_{\text{measurements-errors}}$$
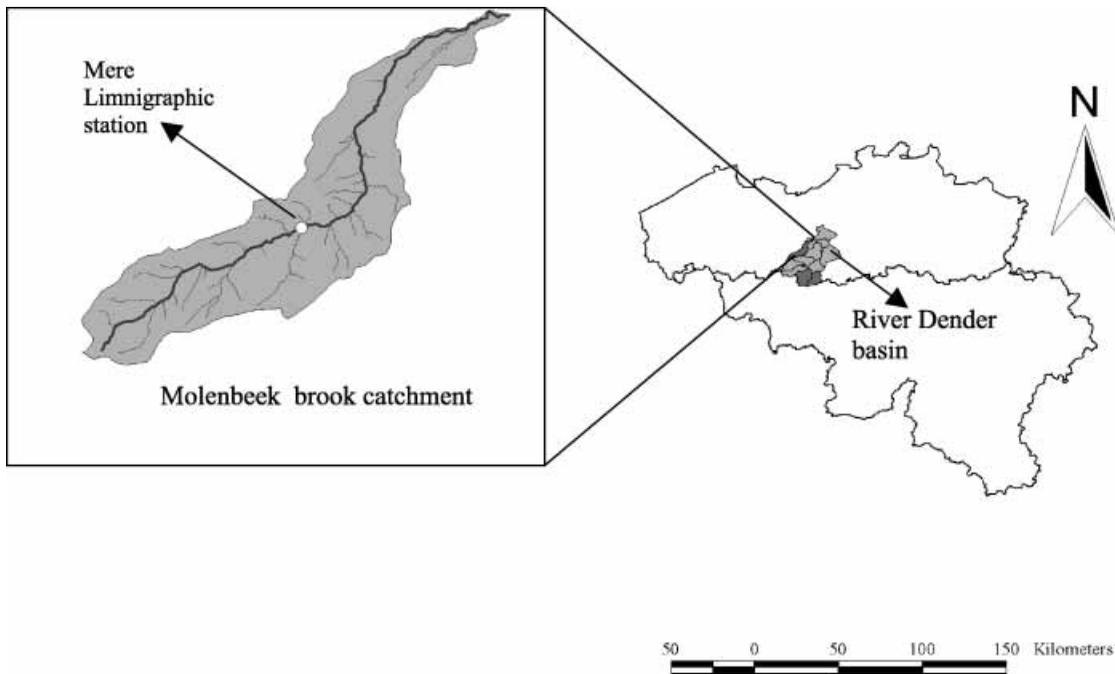
Figure 3 | Plan view of Belgium together with the location of the river Dender basin and the Molenbeek subcatchment.

where

σ^2_{Total} = total uncertainty,

σ^2_{HY+HD} = hydrological (HY) and hydrodynamic (HD) modelling uncertainty,

σ^2_{WQ} = water quality (WQ) modelling uncertainty,

$\sigma^2_{model\text{-}input\text{-}unc}$ = WQ model input uncertainty,

$\sigma^2_{model\text{-}parameters\text{-}unc}$ = WQ model parameter uncertainty,

$\sigma^2_{model\text{-}structure\text{-}unc}$ = WQ model structure uncertainty,

$\sigma^2_{measurements\text{-}errors}$ = WQ measurement errors.

It is clear that using this expression it is assumed that the errors/uncertainties from the different uncertainty sources are uncorrelated. Of course, this assumption cannot be made for the variables considered in the model input and the model output. However, for the errors on these variables, the uncorrelation assumption will be largely valid. The over- or underestimations made for the different inputs and parameters, and by the model structure, have indeed different—and in most of the cases uncorrelated—causes. The error in the hydrological sub-model results, for instance, is mainly caused by rainfall estimation errors, while the error on the domestic BOD

input concentrations is mainly caused by the temporal resolution in the BOD emission measurements.

In the study, the total uncertainty is quantified in the first stage, whereas the other uncertainty types are estimated as subcomponents of this total uncertainty.

Total uncertainty

The total uncertainty is estimated after comparison of the Mike 11 simulation results, as presented in Radwan (2002), with the WQ measurements as follows:

$$\sigma^2_{total} = \frac{\sum_{i=l}^{n}(x_{Mike11,i} - x_{measured,i})^2}{n-1}$$

where $x_{Mike11,i}$ are the Mike 11 results for the time moments, for which WQ measurements $x_{measured}$ are available. In total, n measurements are used ($i = 1, \ldots, n$).

The results of these σ calculations are derived for the different WQ variables. Because the standard deviation is
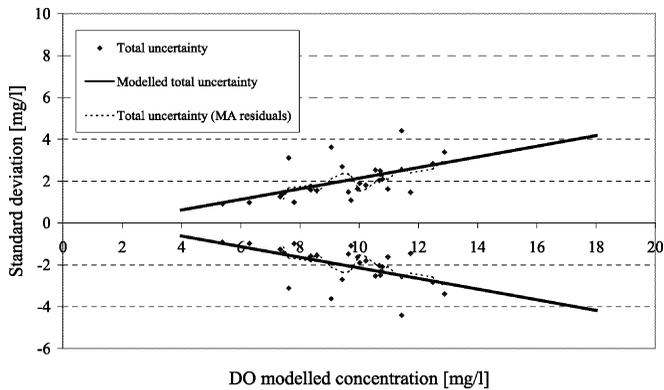
Figure 4 │ Modelled total uncertainty for DO.

Figure 5 │ Total uncertainty (in terms of 68% two-sided confidence limits) for BOD concentrations.

dependent on the WQ concentration, the calculation of σ is carried out for different ranges of the concentration (using a moving window procedure). Afterwards, the mean relation between the standard deviation and the concentration x is calibrated (see Figure 4). This is done by applying first a moving average operation to the plot of the standard deviation results versus the concentration, and then by calibrating a regression curve to these results. The following relations are derived:

$$\sigma^2_{Total(\text{BOD})} = (0.046\text{BOD} + 0.7459)^2$$

$$\sigma^2_{Total(\text{DO})} = (0.254\text{DO} - 0.3967)^2$$

$$\sigma^2_{Total(\text{NH}_4\text{-N})} = (0.0932\text{NH}_4\text{-N} + 0.7502)^2$$

$$\sigma^2_{Total(\text{NH}_3\text{-N})} = (0.3202\text{NO}_3\text{-N} - 0.1166)^2$$

As an example, the DO results are presented in Figure 4. From these results, it can be seen that the standard deviation σ is linearly dependent on the concentration x.

Then random (Monte Carlo) simulations are carried out using random variations of E_{BOD}, E_{DO}, $E_{\text{NH}_4\text{-N}}$ and $E_{\text{NO}_3\text{-N}}$. The probabilistic simulation results derived from the Monte Carlo simulations for BOD, DO, NH$_4$-N and NO$_3$-N are presented in Figures 5–8. The total uncertainty in the predicted concentrations is presented by the grey lines in terms of the 68% two-sided confidence limits (confidence limits according to one times the standard
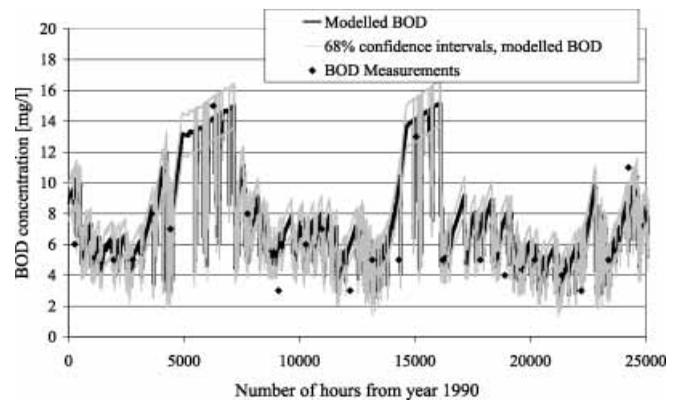
deviation for a normal distribution). For clarity reasons, the results are only shown in these figures for a limited time span.

Uncertainties from hydrological and hydrodynamic modelling

The results from the hydrological (HY) simulation are used as input for the hydrodynamic (HD) simulations. Finally the HD results are used as input for the water quality simulations. To determine the contribution of the hydrological and the hydrodynamic uncertainties to the total uncertainty, the standard deviation of the difference between the HY and HD simulation results and the river discharge measurements is calculated. Then a random (Monte Carlo) simulation is performed with a stochastic term $E_{\text{HY + HD}}$. Using a moving average technique, the relation of the standard deviation to x (see Figures 9–12) can be quantified. This relation can be described using the following relations for the average contribution of the HY + HD modelling uncertainty to the total uncertainty:

$$\sigma^2_{\text{HY + HD(BOD)}} = (0.0096\text{BOD} + 0.5)^2$$

$$\sigma^2_{\text{HY + HD(DO)}} = (0.1)^2$$

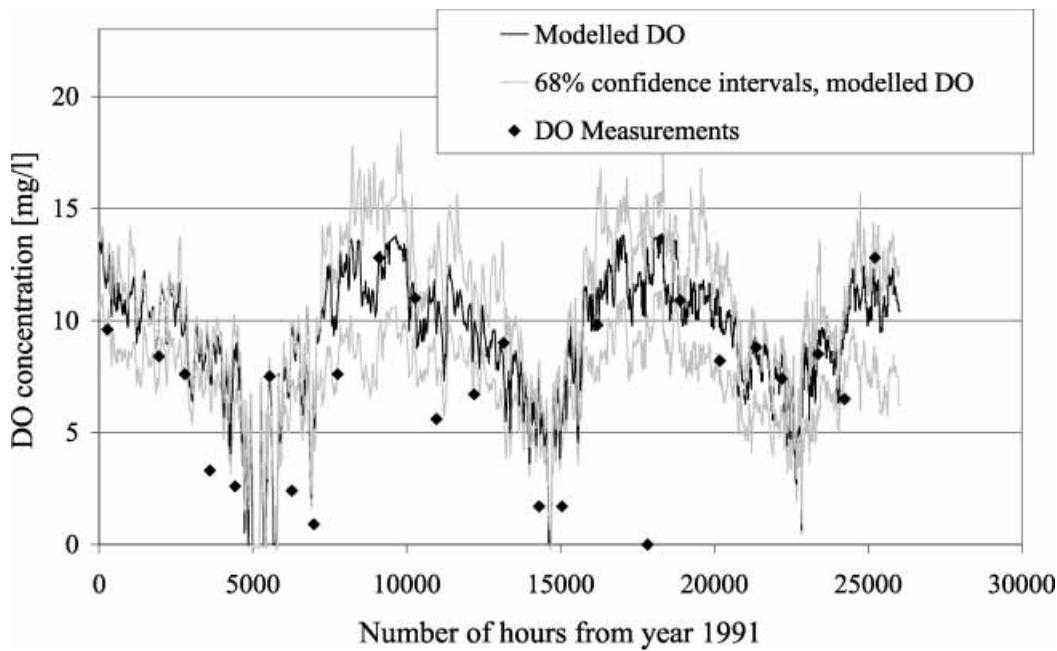$$\sigma^2_{\text{HY + HD(NH}_4\text{-N)}} = (-0.0002\text{NH}_4\text{-N} + 0.8)^2$$

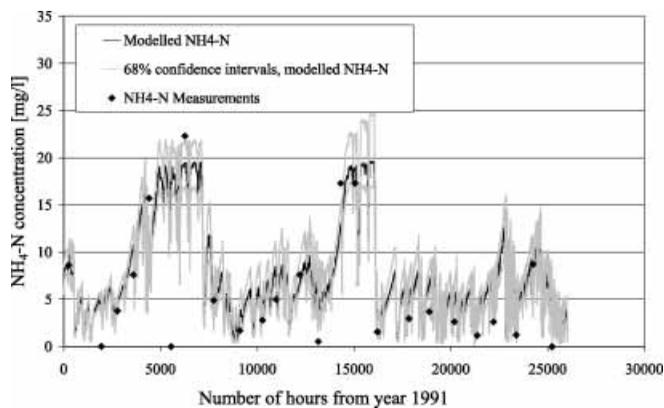**Figure 6** │ Total uncertainty (in terms of 68% two-sided confidence limits) for DO concentrations.

Figure 7 │ Total uncertainty (in terms of 68% two sided confidence limits) for NH$_4$-N concentrations.

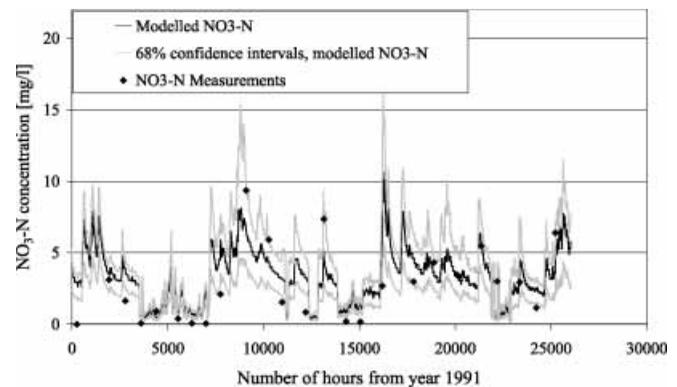**Figure 8** │ Total uncertainty (in terms of 68% two-sided confidence limits) for NO$_3$-N concentrations.

$$\sigma^2_{\text{HY + HD(NO}_3\text{-N)}} = (0.05)^2$$

After computation of the HY + HD uncertainty contribution to the water quality model residuals, the remaining range is explained by the water quality modelling:

$$\sigma^2_{WQ(\text{BOD})} = (0.046\text{BOD} + 0.7459)^2 - (-0.0096\text{BOD} + 0.5)^2$$

$$\sigma^2_{WQ(\text{DO})} = (0.254\text{DO} + 0.3967)^2 - (0.1)^2$$

$$\sigma^2_{WQ(\text{NH}_4\text{-N})} = (0.0932\text{NH}_4\text{-N} + 0.7502)^2 - (-0.0002\text{NH}_4\text{-N} + 0.8)^2$$

$$\sigma^2_{WQ(\text{NO}_3\text{-N})} = (0.3202\text{NO}_3\text{-N} - 0.1166)^2 - (0.05)^2$$

The uncertainty due to water quality modelling can be split into: (1) uncertainties related to the model structure,
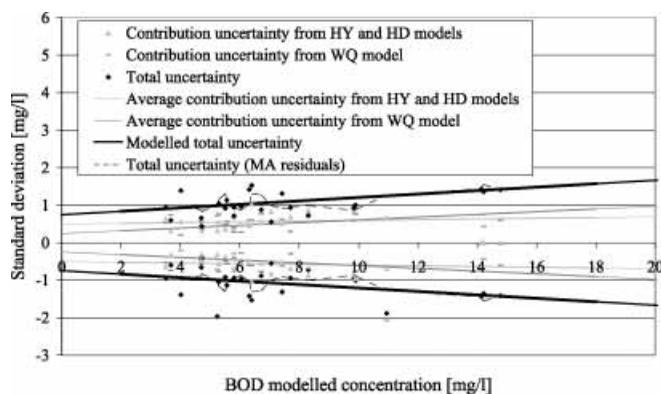
Figure 9 │ Modelled total uncertainty with the contribution from the HY+HD modelling uncertainty and total WQ modelling uncertainty for BOD.
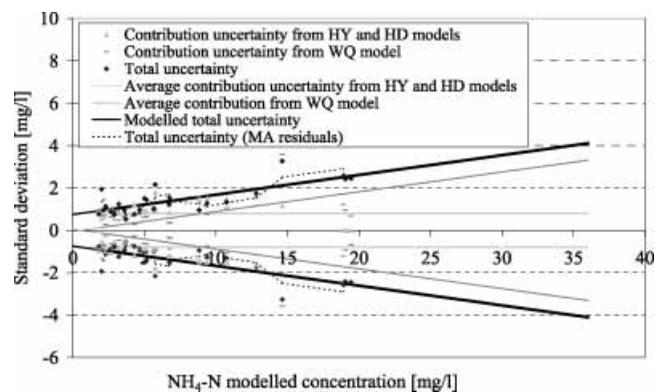

Figure 11 │ Modelled total uncertainty with the contribution from the HY+HD modelling uncertainty and total WQ modelling uncertainty for NH₄-N.
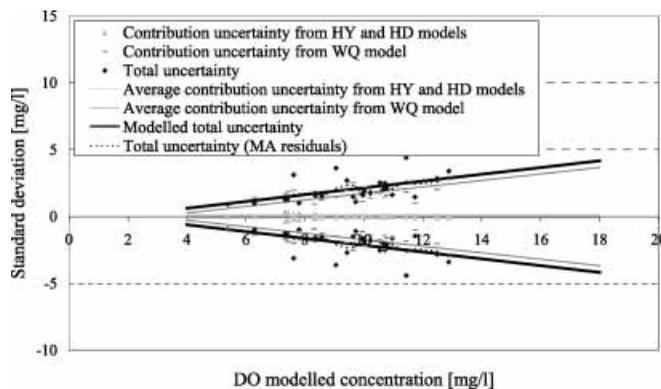

Figure 10 │ Modelled total uncertainty with the contribution from the HY+HD modelling uncertainty and total WQ modelling uncertainty for DO.
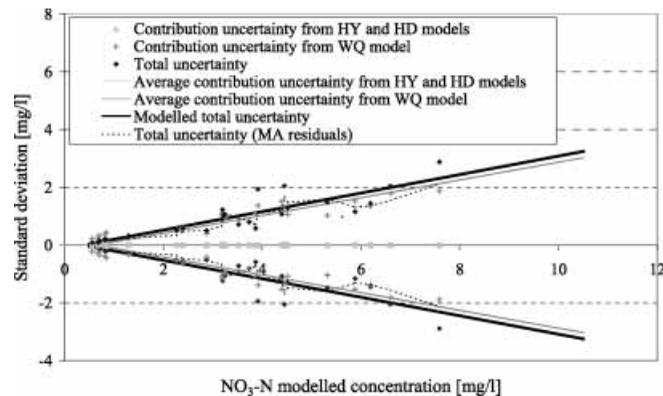

Figure 12 │ Modelled total uncertainty with the contribution from the HY+HD modelling uncertainty and total WQ modelling uncertainty for NO₃-N.

(2) uncertainties by measurements errors, (3) uncertainties in the model parameters and (4) uncertainties in the model input.

In the following subsections, detailed calculations of each of the above-mentioned subgroups are presented.

Water quality model-structure uncertainty

For different water quality models, the water quality process equations are describing the changes in the constituent concentrations due to biological, chemical, biochemical and physical processes. To be able to quantify the uncertainties due to the model structure, different model structures have to be compared. In this study, the

Mike 11 water quality module (DHI 2002) is compared with the Qual2E water quality model (Brown & Barnwell 1987).

Two types of model structures

The Mike 11 water quality model has been developed by DHI Water & Environment. The WQ module deals with the basic aspects of river water quality in areas influenced by human activities, such as oxygen depletion and ammonia levels as a result of organic matter loading. The WQ module is coupled to the advection–dispersion (AD) module, which means that the WQ module deals with the transforming processes of compounds in the river and the AD module is used to simulate the simultaneous transport

process. The WQ module solves the system-coupled differential equations describing the most important physical, chemical and biological interactions in the river. The river water quality can be dealt with at different levels of detail (model levels from 1 to 6). In this study, model level 4 has been chosen. At this level, concentrations of dissolved oxygen (DO), biological oxygen demand (BOD), ammonium (NH_4-N) and nitrate (NO_3-N) are computed in Mike 11. The following processes are taken into consideration as a minimum: reaeration, BOD degradation, respiration of plants and animals and photosynthesis. All these processes consider the effect on the oxygen and nitrogen cycles. Level 4 also means that BOD-DO relationships include exchange with the river bed (e.g. sediment oxygen demand) and nitrification and denitrification.

Qual2E is a similar stream water quality model. Both models indeed assume that the major transport mechanisms, advection and dispersion, are significant only along the main direction of flow (the longitudinal axis of the stream or canal). This means that it is assumed that the streams are well mixed. Also the process formulations of both river water quality models are similar and they allow both for multiple waste discharges and withdrawals.

Comparison of the Qual2E and Mike 11 models

In spite of this general similarity between the Mike 11 and Qual2E water quality models, there are some remarkable differences.

Nitrate nitrogen

- A difference between Mike 11 and Qual2E is the simplified treatment of nitrification in Mike 11 that ignores nitrite as an intermediate product.
- Inhibition of nitrification at low dissolved oxygen concentration is considered only in the Qual2E model.
- The fraction of algal uptake from the nitrogen pool is considered only in the Qual2E model.
- Denitrification is considered in the Mike 11 model but not in the Qual2E model.

Ammonia nitrogen

- In Mike 11, a source of NH_4-N is the degradation of BOD, but for Qual2E, organic nitrogen decay is the source of NH_4-N.

Dissolved oxygen

- Algae is not considered in the DO cycle in Mike 11.

Biochemical oxygen demand

- Monod terms (reduction in BOD degradation at low oxygen levels) are considered only in the Mike 11 model.
- Ultimate BOD concentration is used in the Qual2E model, while BOD_5 is used in Mike 11.

The difference between the two models is calculated as the variance between the simulation results of the two models. The difference between the concentration results for the two models show no dependence with concentration, so the model structure uncertainty can be written as follows:

$$\sigma^2_{\text{model-structure-unc.(BOD)}} = (0.206)^2$$

$$\sigma^2_{\text{model-structure-unc.(DO)}} = (0.148)^2$$

$$\sigma^2_{\text{model-structure-unc.(NH}_4\text{-N)}} = (0.036)^2$$

$$\sigma^2_{\text{model-structure-unc.(NO}_3\text{-N)}} = (0.02)^2$$

It is clear that, in this analysis, the comparison between the Mike 11 and Qual2E results only provides information about the difference in outcomes between the two models and not the difference between the model and the physical reality. The difference between the models provides, however, information on the order of magnitude of the real model-structure uncertainty. Because the Mike 11 and Qual2E model structures are similar, this uncertainty might have been underestimated. Additional comparisons with other (less similar) models will be needed for a more accurate estimation.

Water quality measurement uncertainty

The measurement errors found in the literature (Ahyerre *et al.* 1998) are 15–20% for most water quality variables

and 30–40% for BOD. In this study, it is assumed that, for these values, half of the error can be explained by sampling and the other half by the analysis technique used in the laboratory. For BOD, the laboratory analysis has a larger contribution.

Input and model parameter uncertainty

The simplest method of analyzing parameter uncertainty is a sensitivity analysis. Sensitivity analysis is used to identify the parameters that have the greatest effect on the model output. The effort involved in gathering data to characterize the uncertainty in each parameter is considerable so the sensitivity analysis helps us to focus on those parameters most important for model calibration.

In an attempt to separate the range of uncertainties due to model parameters and model input, the sensitivity of each model parameter and model input change to the model output response is studied. For each modelled WQ variable, the different model input variables and model parameters are identified as follows:

- For DO: nitrification process rate, BOD degradation rate, BOD agricultural input, BOD input from domestic sources and NH_4-N from domestic sources.
- For NO_3-N: nitrification process rate, denitrification process rate, NH_4-N from domestic sources and NO_3-N input from agricultural activities.
- For NH_4-N: BOD degradation rate, nitrification process rate, BOD input from agricultural sources and runoff, BOD input from domestic sources and NH_4-N input from domestic sources.
- For BOD: BOD degradation rate, BOD input from domestic sources and BOD input from agricultural and runoff sources.

In terms of variance, the total model input and model parameter uncertainty for each modelled WQ variable is written as follows:

$$\sigma^2_{model\_input + model\_parameter(DO)} = \sigma^2_{k_{nitr}} + \sigma^2_{k_{degrad}} + \sigma^2_{BOD\_agr\_input} + \sigma^2_{BOD\_domestic\_input} + \sigma^2_{NH_4\text{-}N\_domestic\_input}$$

$$\sigma^2_{model\_input + model\_parameter(NO_3\text{-}N)} = \sigma^2_{k_{nitr}} + \sigma^2_{k_{denitr}} + \sigma^2_{NH_4\text{-}N\_domestic\_input} + \sigma^2_{NO_3\text{-}N\_agr\_input}$$

$$\sigma^2_{model\_input + model\_parameter(NH_4\text{-}N)} = \sigma^2_{k_{degrad}} + \sigma^2_{k_{nitr}} + \sigma^2_{NH_4\text{-}N\_domestic\_input} + \sigma^2_{BOD\_agr\_input} + \sigma^2_{BOD\_domestic\_input}$$

$$\sigma^2_{model\_input + model\_parameter(BOD)} = \sigma^2_{k_{degrad}} + \sigma^2_{BOD\_agr\_input} + \sigma^2_{BOD\_domestic\_input}$$

Again, the assumption is made that the individual errors on the different model inputs and parameters are uncorrelated. It is clear that this assumption is not valid for the inputs itself. However, for the errors on these inputs the independence assumption is much more valid.

The contribution of each term in the variance decomposition equation is a percentage of the total variance. For example, for DO the variance contribution of the nitrification process rate is calculated as follows:

$$\frac{\sigma^2_{k_{nitr}}}{\sigma^2_{model\_input + model\_parameter(DO)}}$$

Using a first-order uncertainty analysis, each variance contribution can be calculated based on the sensitivity coefficients and the variances of the parameters or input variables. The results of the sensitivity analysis are presented in Table 1. In this table, the relative comparison of the different model input and model parameter sensitivities is presented.

The results in Table 1 show that, for all modelled variables, model input is more sensitive than model parameters. For DO, BOD, NH_4-N and NO_3-N, the percentage of the model input sensitivity to the sum of the model input and model parameter sensitivity is 58%, 93%, 94.5% and 79%, respectively.

After having an idea about the sensitivity of each of the model inputs and parameters, a random error is estimated according to the experience of the author in combination with the available literature. For example, the ranges (standard deviations) for the nitrification, denitrification and BOD decay process rates, as found in the literature, are presented in Table 2. These values are adapted from Van der Perk (1996).

Estimated ranges (random errors) for the different model inputs and model parameters are presented in Table 3.

The dependence in time of the random errors modelled by the stochastic terms is represented by an

Table 1 | Relative sensitivity of different model inputs and model parameters

		DO (%)	BOD (%)	NH$_4$-N (%)	NO$_3$-N (%)
Model input	BOD runoff from agricultural sources	8	54	3.5	—
	BOD from domestic and industrial sources	8	39	1	—
	Ammonium from domestic and industrial sources	42	—	90	14
	Nitrate leaching from agricultural sources		—	—	65
Model parameters	BOD degradation parameter	8	7	3.5	—
	Ammonium degradation parameter (nitrification process)	34	—	2	15
	Nitrate degradation parameter	—	—	—	6

Table 2 | Rate constants for nitrification, denitrification and BOD degradation reported in the literature

River	Reference	k_{nitr}	k_{denitr}	k_{BOD}
Biebrza River, Poland	Van der Perk (1996)	3.0–4.0	1.75	
'Normal range'	Veldkamp & Van Mazijk (1989)	0.5–0.8		
'Normal range'	Thomann (1972)	0.1–0.6	0.2–0.7	
'Smaller streams'	Thomann & Muller (1987)	>1		
River Trent, UK	Garland (1978)	0.5–3.0		
Willamette River, Oregon, USA	Rickert (1982)	0.7		
Regge River, The Netherlands	Van den Boomen *et al.* (1995)	0.5	0.5	
River Rhine, The Netherlands	Admiraal & Botermans (1989)	0.0–1.9		
Grindsted River, Denmark	Bach *et al.* (1989)	3.0	1.75	
'Normal range'	ICIM (1992)	0.1	0.0–0.1	
Bedford Ouse River, UK	Whitehead *et al.*		0.1	
River Thames, UK	Whitehead & Williams (1982)		0.05	
'Normal range'	Mike 11 manual (DHI 2002)	0.01–1.54	0–1	0.5–1.5

Adapted from Van der Perk (1996).

Table 3 │ Estimated random errors for the different model inputs and model parameters

	The model inputs and most sensitive parameters	Range
Model input	BOD input from domestic sources	$C \pm 25\%$
	BOD input from agricultural and runoff sources	$C \pm 40\%$
	NH_4-N input from domestic sources	$C \pm 25\%$
	NO_3-N input from agricultural sources	$C \pm 50\%$
Model parameters	Nitrification process rate	0.1–1
	Denitrification process rate	0.1–0.5
	BOD decay rate	0.3–0.7

C is the input concentration used in the simulation.

autocorrelation model. The most flexible model structure for this autocorrelation model is the ARMA structure (Willems 2000). By applying such a model, an error series with a certain autocorrelation structure ($e(t)$) is generated from the uncorrelated series ($u(t)$) using a moving average (MA) operation applied to the uncorrelated series, combined with an autoregression (AR) operation applied to the correlated series:

$$e(t) = a_0 u(t) + a_1 u(t-1) + \ldots + a_p u(t-p) + \text{(MA part)}$$

$$b_1 e(t-1) + b_2 e(t-2) + \ldots + b_q e(t-q) \ \text{(AR part)}$$

In this way, the ARMA model structure can also be seen as a linear transfer function model with an uncorrelated noise series as model input (also called 'white noise') to generate an autocorrelated error series as output ('coloured noise'). Using the model orders p and q, the ARMA model is denoted by 'ARMA (p,q)'. When a normal error distribution is assumed, the $u(t)$ white noise series can be randomly generated on the basis of only one parameter (the white noise standard deviation σ_u):

$$E_U \sim N(0, \sigma_U)$$

This description of U, together with the ARMA model, then forms the mathematical formulation of the 'stochas-

tic process' $e(t)$. The parameters of the ARMA model (a_0, ..., a_p, b_1, ..., b_q) can be calibrated and the orders identified on the basis of the autocorrelations and partial autocorrelations calculated from an observed realization of the $e(t)$ process. Such a realization can be derived whenever continuous time series of measurements are available to calculate the errors on the variable considered (Willems 2000). After calibration, the stochastic ARMA model can be used to generate long-term time series of random errors. The problem is, however, that continuous time series of measurements are often not available. This is also the case for the Molenbeek catchment. Therefore, the autocorrelation is roughly estimated. A random number generator is used for the generation of the white noise series. After running the random (Monte Carlo) simulation by taking into account the random errors as in Table 3 and the correlation as mentioned above, the results are as shown in Figures 13–16.

Biochemical oxygen demand

The results indicate that, for the modelling of the BOD concentration, the BOD input to the model from the two different sources (domestic and agricultural/runoff) is the cause of 45% of the total uncertainty in the results of the
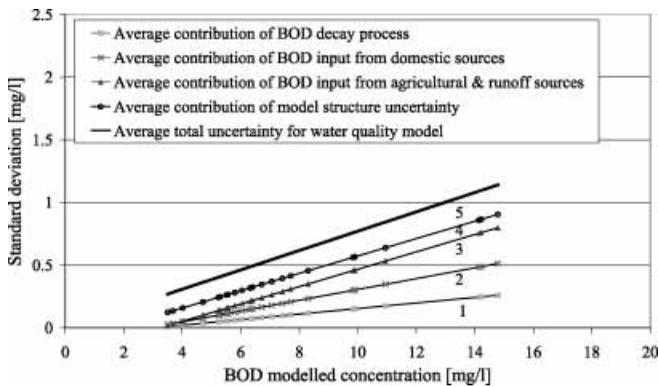
Figure 13 │ Average total uncertainty for the BOD model, together with the average uncertainty contribution of (1) BOD decay process, (2) BOD input from domestic sources, (3) BOD input from agricultural and runoff sources, (4) model structure and (5) measurements.
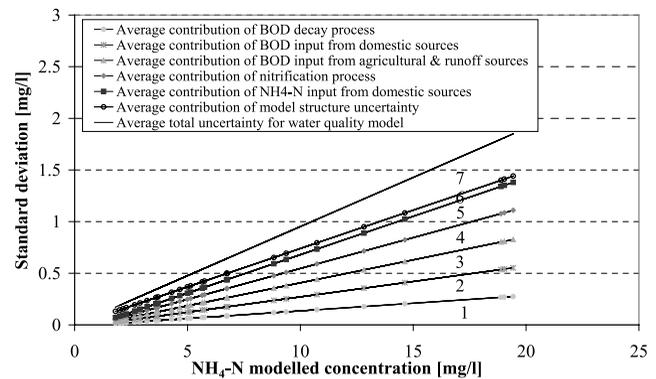

Figure 15 │ Average total uncertainty for the NH_4-N model, together with the average uncertainty contribution of (1) BOD decay process, (2) BOD input from domestic sources, (3) BOD input from agricultural and runoff sources, (4) nitrification process, (5) NH_4-N input from domestic sources, (6) model structure and (7) measurements.
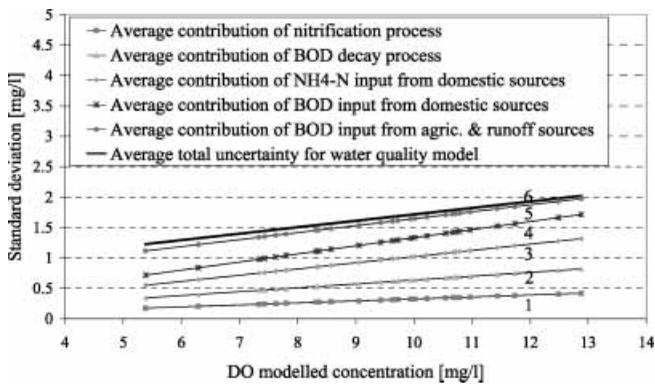

Figure 14 │ Average total uncertainty for the DO model, together with the average uncertainty contribution of (1) nitrification process, (2) BOD decay process, (3) NH_4-N input from domestic sources, (4) BOD input from domestic sources, (5) BOD input from agricultural and runoff sources and (6) model structure and measurements.
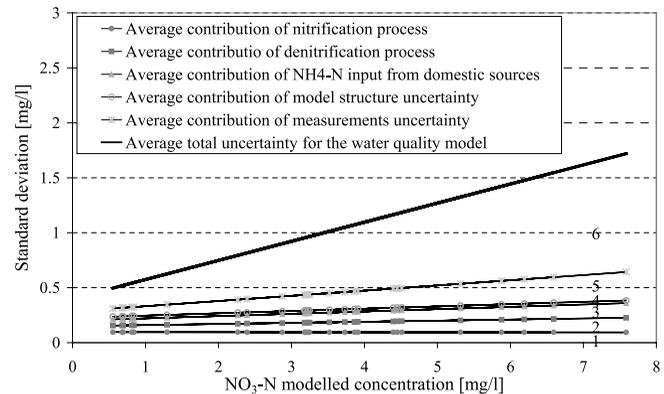

Figure 16 │ Average total uncertainty for the NO_3-N model, together with the average uncertainty contribution of (1) nitrification process, (2) denitrification process, (3) NH_4-N input from domestic sources, (4) model structure, (5) measurements and (6) Drainmod model (NO_3-N leaching from agricultural sources).

WQ submodel. This value is an average for the range of concentration values considered in the figure.

The BOD decay, which is the most important model parameter affecting the BOD calculation, can explain 25% of the total WQ uncertainty. The smallest contribution is from the model structure.

Dissolved oxygen

The results indicate that, for DO, the model input has the highest contribution (60%) to the total WQ uncertainty. The second contribution is from the model parameters of the BOD decay and nitrification processes which are the cause of 36% of the total WQ uncertainty.

Ammonia-nitrogen

For NH_4-N, when the different uncertainty sources are ranked and grouped, model input has the highest contribution, followed by model parameters, measurement errors and finally the model structure.

Table 4 | Summary of the average uncertainty contribution for the different uncertainty sources for the different water quality variables

Different uncertainty sources		DO (%)	BOD (%)	NH$_4$-N (%)	NO$_3$-N (%)
Model input	BOD runoff from agricultural sources	20	25	18	—
	BOD from domestic and industrial sources	18	31	18	—
	Ammonium from domestic and industrial sources	23	—	20	11
	Nitrate leaching from agricultural sources	—	—	—	61
Sum model input		**61**	**56**	**56**	**72**
Model parameters	BOD degradation parameter	18	31	18	—
	Ammonium degradation parameter (nitrification process)	19	—	19	13
	Nitrate degradation parameter	—	—	—	11
Sum model parameters		**37**	**31**	**37**	**24**
Model structure		**2**	**13**	**7**	**4**
Total		**100**	**100**	**100**	**100**

Nitrate-nitrogen

Results of the NO$_3$-N uncertainty analysis show that the model input, which is mainly the result of the nitrate leaching model, has the highest contribution to the total WQ uncertainty. The second highest contribution to the total WQ uncertainty is from the model parameters, followed by measurement errors and finally the model structure.

Summary of the results

It is clear that, for all the water quality variables, the uncertainty contributions can be ranked in the same way: the model input is most important, followed by model parameters and then the model structures. Table 4 gives a summary of these contributions for the four variables considered. Measurement errors were eliminated for this table as they do not contribute to the uncertainty in the model results.

Uncertainty from nitrate leaching model

As the nitrate leaching model has the highest contribution to the NO$_3$-N modelling uncertainty, it is needed to do further analysis for this model. To analyze the different uncertainty sources for the nitrate leaching model, a similar procedure as presented here for the river water quality model can be applied. This means that the total nitrate leaching model uncertainty can be split up into model input, model parameters, model structure and measurement uncertainty.

A previous study (Haan 2000) dealing with model parameter uncertainty results showed that the bulk density, the coefficient of denitrification and the coefficient of mineralization are the most uncertain parameters for the Drainmod model.

CONCLUSIONS

Uncertainty analysis for model simulations is assuming a growing importance in the field of water quality management. The importance of this concern is provided by recent public awareness over health risks from the improper disposal of pollutants as well as by the continuing emphasis on risk assessment. The first step in the chain of risk assessment is quantification of the error in predicting water quality.

In water quality models for water resources management and planning, the model output depends on the modelled processes (model structure), the input data and a number of parameters, which are essentially not known perfectly. The task of the uncertainty analysis is to estimate the different uncertainty contributions to the model output in terms of the uncertainties in the input, the model parameters and the model structure. Also, the measurement errors have to be taken into account. To be able to do this analysis for the water quality model results, the uncertainty due to the hydrological and hydrodynamic modelling has to be excluded in a first step. Then the remaining uncertainty can be explained by the water quality modelling. To split this uncertainty into the different components, the model structure uncertainty was first calculated by comparing two different model structures (Mike 11 and Qual2E). Secondly, the measurement uncertainty was estimated based on the literature. In the third step, the remaining uncertainty was split into uncertainty components due to different model input variables and model parameters. To have an idea about the relative importance of the different model inputs and parameters, a sensitivity analysis was carried out for the Molenbeek brook catchment at the village of Erpe-Mere in Belgium. The results of this analysis show that, respectively, for DO, BOD, NH_4-N and NO_3-N, the percentage of the model input sensitivity to the combined sensitivity of model inputs and model parameters equals 58%, 93%, 94.5% and 79%. These results indicate that model input is more sensitive than model parameters, especially for BOD and NH_4-N.

The absolute values for the random errors are derived for the different model inputs and model parameters, based on the experience of the author and the available literature. Based on these values, Monte Carlo simulations

are carried out. For the case study, the results of the Monte Carlo simulation show that the percentage of model input contribution to the total uncertainty is 61% for DO, 56% for BOD, 56% NH_4-N and 72% for NO_3-N. For the percentage of the contribution of model parameters to the total uncertainty, values of 37% for DO, 31% for BOD, 37% for NH_4-N and 24% for NO_3-N are derived. Finally, the model structure uncertainty contribution to the total uncertainty is 2% for DO, 13% for BOD, 7% for NH_4-N and 4% for NO_3-N. When the water quality model results are compared with measurements, the residuals are also partly influenced by measurement errors. This contribution equals, for the case study, 2% for DO, 20% for BOD, 17% for NH_4-N and 15% for NO_3-N. From the previous results, it is clear that the model input needs the greatest attention when one is attempting a more accurate modelling of water quality. Model parameters and model calibration have the second priority, followed by the measurement accuracy. Finally, the model structure seems to need less attention as it contributes only a small percentage to the total uncertainty. In this study, however, the model structure uncertainty is analyzed only by comparing the Mike 11 and Qual2E model structures. Both models are quite similar. Therefore, the recommendation can be made to look at additional comparisons with other WQ model structures, such as the submodels based on the Activated Sludge Modelling concept (Rauch 1998).

REFERENCES

Admiraal, W. & Botermans, Y. J. H. 1989 Comparison of nitrification rates in three branches of the lower River Rhine, *Biogeochem.* **8**, 135–151.

Ahyerre, M., Chebbo, G., Tassin, B. & Gaume, E. 1998 Storm water quality modelling, an ambitious objective? *Wat. Sci. Technol.* **37**(1), 205–213.

Bach, H. K., Brink, H., Olesen, K. W. & Havno, K. 1989 *Application of PC-based Models in River Water Quality Modeling*. Water Quality Institute, Hørsholm, Denmark.

Brevé, M., Skaggs, R., Kandil, H., Parsons, J. & Gilliam, J. 1997 DRAINMOD-N, a nitrogen model for artificially drained soils. *Trans. ASAE* **40**(4), 1067–1075.

Brown, L. C. & Barnwell, T. O. 1987 The enhanced stream water quality models Qual2E and Qual2E-uncas: documentation and user manual. Report EPA/600/3-87/007, USEPA, Athens, GA.

Box, G. E. P. & Cox, D. R. 1964 An analysis of transformations. *J. R. Statist. Soc.* 211–243, discussion 244–252.

DHI. 2002 *MIKE11, A Microcomputer-based Modelling System for Rivers and Channels. User and Reference Manual*, DHI Water and Environment, Hørsholm, Denmark.

El-Sadek, A., Feyen, J. & Ragab, R. 2002 Simulation of nitrogen balance of maize field under different drainage strategies using DRAINMOD-N model. *Irrigation and Drainage Journal* **51**(1), 61–75

Frey, H. C. 1992 Quantitative analysis of uncertainty and variability in environmental policy making. Directorate for Science and Policy Programs, American Association for the Advancement of Science, Washington, DC.

Garland, J. H. N. 1987 Nitrification in the River Trent, Chapter 7. In *Mathematical Models in Water Pollution Control* (ed. A. James), pp. 167–191. Wiley, Chichester.

Haan, P. 2000 The effect of parameter uncertainty on Drainmod predictions for hydrology, yield and water quality. *PhD dissertation* North Carolina State University, Raleigh, NC.

ICIM 1992 *DUFLOW, A Micro-computer Package for the Simulation of One-dimensional Unsteady Flow and Water Quality in Open Channel System*, p. 107. Bureau ICIM, Rijswijk.

Radwan, M. 2002 River water quality modelling as water resources management tool at catchment scale. *PhD thesis* Faculty of Engineering, KU Leuven, Belgium.

Radwan, M., Willems, P. & Berlamont, J. 2000 Assessment of surface water quantity and quality. *International Conference on Wadi Hydrology, Sharm El-Sheikh, Egypt, 21–23 November* CD ROM.

Rauch, W., Henze, M., Koncsos, L., Reichert, P., Shanahan, P., Somlyody, L. & Vanrolleghem, P. 1998 River water quality modelling: I. State of the art. *Wat. Sci. Technol.* **38**(11), 237–244.

Rickert, D. A. 1982 *Use of Dissolved Oxygen Modeling Results in the Management of River Quality; Case History of the Wittamette River, Oregon*, pp. 13–22. IAHS publication 139.

Thomann, R. V. 1972 *Systems Analysis and Water Quality Management*. McGraw-Hill, New York.

Thomann, R. V. & Muller, J. A. 1987 *Principles of Surface Water Quality Modeling and Control*. Harper Collins, New York.

Van Den Boomen, R. M., Salverda, A. P., Uunk, E. J. B. & Roos, C. 1995 Modellering van het Reggesysteem met DUFLOW (in Dutch with English summary). *H2O* **28**, 107–111.

Van der Perk, M. 1996 Uncertainty issues in modelling the influence of bed sediments on water composition. *PhD thesis* Koninklijk Nederlands Aardrijkskundig Genootschap, Utrecht, Faculteit Ruimtelijk Wetenschappen Universiteit Utrecht.

Veldkamp, R. G. & Van Mazijk, A. 1989 *Waterkwaliteitsmodellering oppervlaktewater* (in Dutch). Document n11, Sanitary Engineering and water management group, Delft University of Technology, the Netherlands.

Willems, P. 2000 Probabilistic immission modelling of receiving surface waters. *PhD thesis* Faculty of Engineering, KU Leuven, Belgium.

Willems, P. & Berlamont, J. 1999 Probabilistic modelling of sewer system overflow emissions. *Wat. Sci. Technol.* **39**(9), 47–54.

Willems P. & Berlamont J. 2002 Probabilistic emission and immission modelling: case study of the combined sewer–WWTP–receiving water system at Dessel (Belgium). *Wat. Sci. Technol.* **45**(3), 117–124.

Whitehead, P. G., Beck, M. B. & O'Connel, E. 1981 A systems model of streamflow and water quality in the Bedford Ouse River system–II Water quality modeling. *Water Research* **15**, 1157–1171.

Whitehead, P. G. & Williams, R. J. 1982 *A Dynamic Nitrogen Balance Model for River Systems*, pp. 89–99. IAHS Publication 139.