Project Number 017841

NitroEurope IP

The nitrogen cycle and its influence on the European greenhouse gas balance.

Sixth Framework Programme

Priority 6.3

Global Change and Ecosystems

D6.3.3

An updated protocol detailing agreed procedures for UQ/UA.

Due date of deliverable: **01/01/02008** Actual submission date: **00/00/0000**

Start Date of Project: **01/02/2006** Duration: **60 months**

Organisation name of lead contractor for this deliverable :
NERC

Revision: **final**

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)		
Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	☐
RE	Restricted to a group specified by the consortium (including the Commission Services)	☐
CO	Confidential, only for members of the consortium (including the Commission Services)	☐

Operational methodology for spatial uncertainty quantification and spatial uncertainty analysis of regional scale NEU models

Gerard Heuvelink[1], Marcel van Oijen[2], Wim De Vries[1] and Hans Kros[1]

[1] Alterra, Wageningen University and Research Centre
[2] Centre for Ecology and Hydrology, Edinburgh

September 2008

1 Introduction

1.1 Background

The five-year Integrated Project NitroEurope (NEU, www.nitroeurope.eu) started in February 2006 to address the prime issues of European N budgets in relation to C cycling and greenhouse gas exchange. Climate change policy requires integrated assessment of net greenhouse gas (GHG) exchange rather than just CO_2. Within the context of NitroEurope both detailed ecosystem models, such as DNDC (Li et al. 1992) and DAYCENT (Parton et al. 1998), and a simplified process-based integrated multi-sector, multi-component model called INTEGRATOR are used to quantify the net N and GHG exchange at the European scale.

Uncertainty quantification (UQ) and uncertainty analysis (UA) are important subjects of NitroEurope. UQ and UA consider the case in which (part of) the model input data and the model as such are uncertain, and where an analysis is made of how these uncertainties propagate to the model outputs. The purpose of UQ is to analyse how uncertainties in model input and model structure propagate to the model output, whereas UA determines how much individual sources of uncertainty contribute to the output uncertainty. Modellers within NEU are committed to carry out an UQ and UA as specified in the UQ-UA protocol (Van Oijen 2007). The UQ-UA protocol is a concise document that provides definitions and describes in general terms how the UQ and UA are done. It applies to both plot-scale modelling (Component 3, Activity 3.1) and to regional scale modelling (Component 6, Activity 6.3). However, in order to apply the protocol in practice it is necessary to work out the methodology in more detail.

1.2 Objectives

The purpose of this work is to focus on UQ/UA for regional scale modelling and propose an operational methodology for NEU modellers and model users involved in Activity 6.3.

Although this text provides more detail than the UQ-UA protocol and aims to provide a ‘recipe cookbook’, it is inevitable that at places readers will have to consult the references for further details or must choose themselves between multiple approaches in specific cases. In an attempt to provide guidelines it is also inevitable that some of the generality of the methodology is lost. In specific applications, one may therefore decide to divert from the instructions provided here. For instance, in this text it is assumed that uncertainty in model inputs is characterised by probability distribution functions, whereas in fact other representations can also be used, such as ranked and unranked scenarios, rough sets and fuzzy sets (Brown et al. 2005). Also, it is assumed that UQ and UA are always done using a Monte Carlo simulation approach, whereas alternative methods, such as Taylor series approximation, can also be used in specific cases (Heuvelink 1998). The restrictions are made with NEU regional scale applications in mind and appear most useful for the type of analyses to be done within the NEU project. The models used are process-based regional-scale models where the environmental inputs are mainly constants, time series and spatially distributed variables characterising soil and atmospheric conditions, and with some model parameters having all kinds of domains of applicability (from completely generic to ecosystem-specific to species-specific). In many cases the outputs need to be scaled up to the regional scale.

Literature on UQ/UA and the associated operational methodology is abundant and no attempt is made here to give a thorough review (see the references in the UQ-UA protocol, also). Van Der Sluijs et al. (2007) provide a diagnostic diagram to synthesize results of quantitative and qualitative UQ/UA. Janssen et al. (2005) also extend UQ/UA beyond the quantitative assessment of uncertainties in model results and address the communication of uncertainty. Brown and Heuvelink (2007) present a software tool that helps assessing and storing uncertainties in environmental data and generates realisations of uncertain data for use in UQ/UA. Saltelli et al. (2004) provide a guide to UA. De Vries et al. (2003) and Winiwarter (2007) apply some of the methodologies to assess the uncertainty in nitrogen fluxes and national GHG inventories.

The UQ/UA methodology presented here draws on existing work and merges it into an operational methodology appropriate for regional scale environmental modelling. It is also aimed at keeping things as simple as possible. More elaborate approaches are not discussed but merely hinted at.

Section 2 of this text presents the ten steps of the UQ/UA chain. Section 3 draws conclusions and proposes next steps for UQ/UA applications within Activity 6.3.

2 Operational methodology for UQ/UA in ten steps

The UQ/UA methodology consists of ten steps that are sequentially executed:

1. Define the model, its inputs and outputs.
2. Select the model output(s) for which the uncertainty must be assessed.
3. Decide which sources of uncertainty are included: input uncertainty and/or model structure uncertainty. Input uncertainty can be further subdivided into uncertainty in environmental constants and variables, initial conditions and model parameters.
4. When model inputs are considered uncertain, decide which of these are treated as uncertain and derive their joint probability distribution.
5. When model structure uncertainty is considered, decide how it is expressed and characterise it with a probability distribution.
6. Generate many possible realities of all uncertainty sources by repeated sampling from their probability distributions.
7. Run the model for the simulated realities and store the results.
8. Scale up the simulated model outputs to the desired spatial and temporal support.
9. Apply steps 6 to 8 in a modified way when the uncertainty contributions of individual or grouped uncertainty sources are required (i.e. Uncertainty Analysis).
10. Communicate the outcomes of the UQ/UA.

The ten steps are now described. Actions required within each step are represented by bullets. Jointly, the bulleted list contains all actions required in the UQ/UA.

2.1 Define the model, its inputs and outputs

A model may be defined as a simplified representation of a part of reality. It communicates with the outside world (i.e., the other parts of reality) through inputs and outputs (Figure 1). All arrows entering the system boundaries are model inputs, arrows leaving the system are outputs. Some model inputs such as initial conditions and model parameters may not be represented as arrows but enter the model implementation at some stage and are therefore inputs as well. Thus:

- List all inputs and outputs of the model, where inputs include initial conditions, environmental constants and variables, and model parameters.

It is important that the boundaries of the modelled system and the inputs and outputs are clearly defined because the results of the UQ/UA depend on it. The following is therefore also needed:

- Define the spatial and/or temporal *extent* of the model. It must be known over which area and for what time interval the input and output are required. For example, is it the atmospheric deposition for the whole of Europe, for all natural areas within Europe, is it for the year 2000 or the period from 1970 to 2010? Vertical extent (depth or height) must also be defined when relevant.
- Define the spatial and/or temporal *resolution* of the model. This refers to the density in space and time that inputs are needed and output is produced by the model. For instance, soil pH may be needed on a 1 km grid, at 10 cm depth intervals, with annual frequency.
- Define the sectors and/or compartments (i.e. the boundary of the model system) that are represented by the model. It must be clear which ecosystems/compartments are considered. For instance, this may be the (semi) natural terrestrial system, agro-system, atmospheric system and/or aquatic system and the industry and/or agro-forestry sector.

All inputs and outputs must be listed, and for each item the following actions are required:

- Unambiguous definition of the phenomenon that it describes. For example, the description ‘soil pH’ is insufficient because it matters how pH is defined, e.g. is it pH($CaCl_2$), pH(H_2O) or pH(KCl).
- Measurement scale and measurement units. Are the inputs and outputs continuous numerical, discrete numerical (i.e. counts), or categorical (i.e. ordinal or nominal)? In case of numerical variables, what are the SI units in which the variable is expressed? In case of categorical variables, how many classes are distinguished and what do these mean? For instance, soil type may be classified using the FAO World Reference Soil system, the European Soil Data Base or a local, national classification system.
- Spatial and/or temporal *support*. This refers to the spatial area or volume and temporal interval over which the variable is aggregated. For instance, is it the average soil pH over the 1 km^2 grid cell, averaged over the top 10 cm and the whole year that is required, or is it the soil pH of a 25 m^2 plot at 10 cm depth at a

specific day of the year? Note that this is not a trivial issue, because the average of point-support pH values over some area is not the same as the pH of a bulk sample from the area (because pH is the logarithm of a concentration). Clarity about the support of model inputs and outputs is crucially important for UQ/UA, because the uncertainty about variables is affected by their support. In general, the larger the support, the smaller the uncertainty. Thus, modellers should think carefully about the required support of the model inputs. Possibly, all inputs must have the same spatio-temporal support. This may require that some of the inputs are scaled up or down before these are submitted to the model, which can be done using the methodologies reviewed in step 8. It should be noted that support is not the same as resolution. For instance, a spatial database may contain the forest biomass measured at 1 hectare plot support on a 10 km grid resolution. Note that for inputs and outputs that are constant in space or time there is no need to specify the spatial or temporal support and resolution.

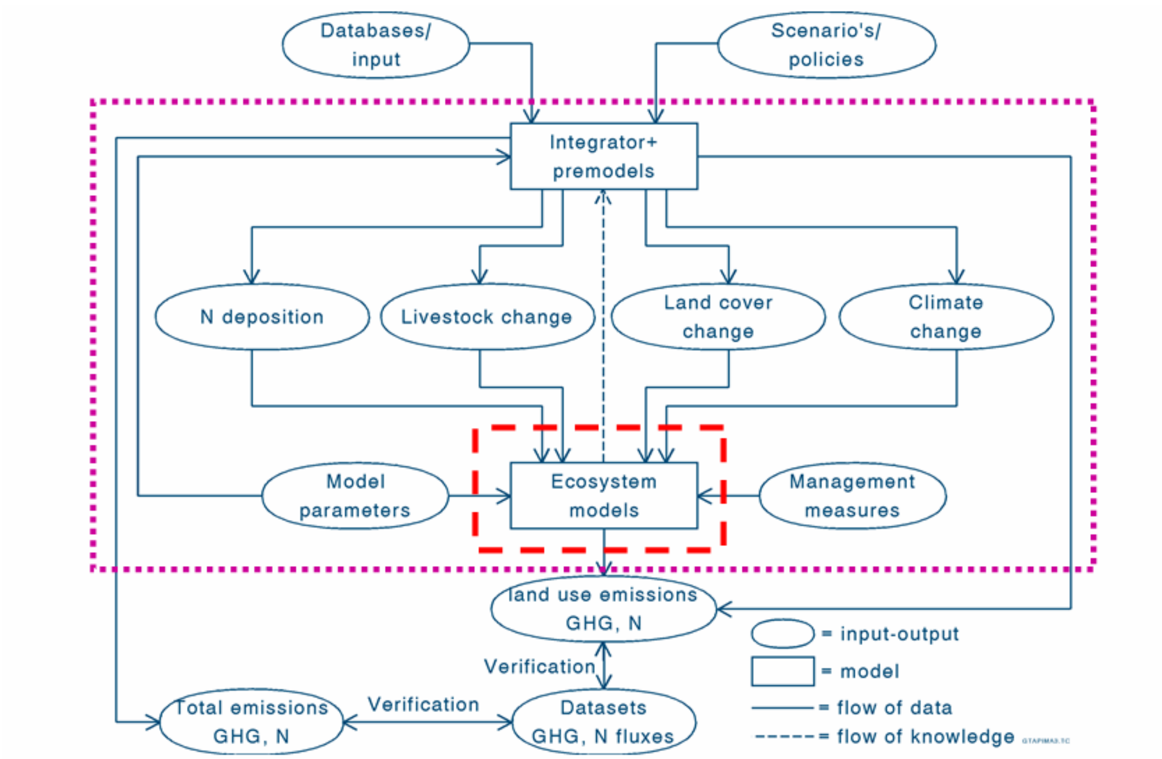

Figure 1. Example boundaries (dashed lines) of an ecosystem model used within NEU. The choice of model boundary will affect the results of the UQ/UA and must be clearly defined.

2.2 Select the model output(s) for which the uncertainty must be assessed

A model usually has multiple outputs. The uncertainty about one model output may be different from that of another output. Moreover, the contribution of individual uncertainty sources to the model output will differ between outputs. A UQ/UA usually focuses only on those uncertainty sources that have the largest contribution to the output uncertainty. The following actions are required:

- Define the targeted model output of the UQ/UA. If multiple outputs are considered, uncertainty sources must be included that importantly affect one or more of these outputs (see next step).
- Define the spatio-temporal output support. This is important because the magnitude of uncertainty and the contribution of uncertainty sources to output uncertainty generally depend on the spatio-temporal support of the output. Typically, the targeted output support is greater than the support at which model calculations are done. For instance, soil emission processes may be represented in the model at the scale of the soil column taking daily and seasonal effects into account while the interest is in the annual average emission over a large region (see also step 8 below).

2.3 Decide which uncertainty sources are included in the analysis

Model output uncertainty is affected by model input uncertainty and model structural uncertainty, where model input uncertainty can be further subdivided into (i) uncertainty in initial conditions, (ii) environmental variables and constants, and (iii) parameters, yielding four categories of uncertainty sources in total. Treating all four categories uncertain from the beginning would make a complex analysis. It is advisable to increase the complexity step by step:

- Decide which of the four uncertainty categories are included in the UQ/UA and rank these in order of treatment. Begin the UQ/UA by considering only the first category. Once this is completed, increase complexity by also considering the second uncertainty category and so on.

2.4 When input uncertainty is considered, decide which inputs are treated uncertain and derive their joint probability distribution

This step is the largest and most complex step of the UQ/UA chain. First the uncertain inputs must be selected. Next, the uncertainty must be quantified. Here, distinction must be made between continuously distributed and discretely distributed inputs and between uncertainty characterised by probability distributions and by samples. Also spatial, temporal and cross-correlations must be addressed and the methods available to estimate parameters of the probability distributions must be reviewed.

Models can have hundreds or even thousands of inputs. It is neither necessary nor possible to treat all inputs as uncertain. Only those inputs should be considered uncertain that have a meaningful contribution to the uncertainty in model output. Two factors determine whether an input has a meaningful contribution. These are:

1. the magnitude of uncertainty about the model input;
2. the sensitivity of the model to changes in the input.

The difficulty is in deciding which are the main uncertain inputs, because this is the objective of UA. The decision which inputs to include must therefore often be based on expert judgement and simplified analyses. Model experts usually have a fair idea what the sensitive model inputs are, and crude sensitivity analyses that compute changes in model output caused by fixed changes in inputs (say ±10%) in a ‘representative’ situation can support these ideas. The UQ-UA protocol notes that “this so-called screening step can be done by one-at-a-time sensitivity analysis", but one must be careful not to exclude inputs that would have become important for different settings of the other inputs than were used in screening. For example, if the screening is done by simulations for a cold site, one may delete from further UQ some soil turnover rate parameters that would not have been deleted if the screening had been done for a warmer site. For more details on screening see Saltelli et al. (2000).

Once model sensitivity has been analysed, next the uncertainty about the model inputs must be crudely assessed. Again, experts usually have a fair idea about the magnitude of the difference between the true value of an input and the estimated value stored in the database. Often, data are available to support the assessment. For instance, a climatologist may use time series of past annual precipitations and additional information to estimate the future annual precipitation in a given region and the associated uncertainty.

The proposed strategy is therefore:

- List the inputs that are treated as uncertain. Include only those inputs that have a large uncertainty and for which the model output is sensitive (Figure 2). Limit the total number of uncertain inputs to two to five in an initial phase, which may later be extended to ten to thirty. A practical approach is to list all model inputs in a table and score them on sensitivity and uncertainty. Only inputs that get high scores on both accounts are included in the UQ/UA.

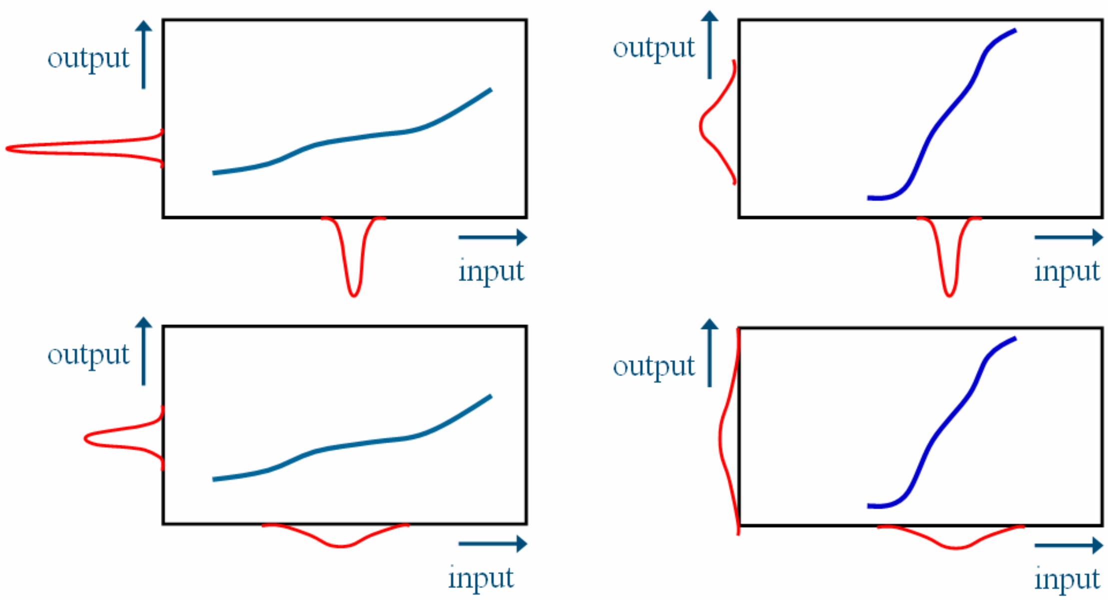

Figure 2. Four cases that illustrate how the magnitude of output uncertainty depends on input uncertainty and model sensitivity. Input uncertainty can be large (wide probability distribution functions (pdfs) in bottom panels) or small (narrow pdfs in top panels). Model sensitivity can be small (gentle slopes in left panels) or large (steep slopes in right panels). Output uncertainty is large when input uncertainty and model sensitivity are both large (wide pdfs in bottom right panels).

Once it is decided which inputs are treated uncertain, the next step is to quantify their uncertainty. This can be done in two different ways:

- For each uncertain input that is considered, decide whether the uncertainty is represented by a probability distribution or by a (large) sample of possible realities (i.e. random draws from the probability distribution).

The choice depends on the information available. For instance, when the uncertain input is the output of another Monte Carlo UQ/UA analysis (i.e., a chain of Monte Carlo UQ/UA analyses) or when it is obtained using Bayesian calibration (see later), then uncertainty representation by a sample is the prevailing approach. This approach may also be used when a random sample from the input variable is available (e.g. the SPADE/WISE datasets of soil properties). Representation by probability distributions is the preferred option because it is more comprehensive but it may not always be feasible. We discuss it first.

- When input uncertainty is described by a probability distribution function, start with defining the marginal (univariate) distribution and next extend it with spatial and/or temporal correlation.

Different distributions are used depending on whether the input is measured on a continuous numerical scale, on a discrete numerical scale or on a categorical scale. It also matters whether the input is constant in space and time or varies in space and/or time.

An uncertain continuous numerical constant C is characterised by its cumulative distribution function (cdf):

$$F_C(c) = P(C \le c) \tag{1}$$

The cdf is a continuous, non-decreasing, function on the real numbers, whose limit values are $F_C(-\infty)=0$ and $F_C(+\infty)=1$. The probability density function is the derivative of F and must be non-negative everywhere, with a surface area below the curve that equals one (Figure 3).

The probability distribution function (pdf) for a discrete numerical or categorical constant is:

$$P(c_i) = P(C = c_i) \tag{2}$$

where the c_i (i=1,..,m) are numbers or categories, respectively. Each of the $P(c_i)$ should be non-negative and their sum should equal one (Figure 3).

An uncertain continuous numerical variable V that varies in space and/or time is characterised by its (cumulative) joint pdf:

$$F_V(v_1, s_1 \ldots, v_n, s_n) = P(V(s_1) \le v_1, \ldots, V(s_n) \le v_n) \tag{3}$$

where the s_i are coordinates (i.e., s_i may comprise x_i, y_i, z_i and t_i) and n may assume any integer value. F_V must be known for each and every combination of the v_i and s_i. The corresponding pdf for a discrete numerical or categorical variable is:

$$F_V(v_1, s_1, \ldots, v_n, s_n) = P(V(s_1) = v_1, \ldots, V(s_n) = v_n) \tag{4}$$

where the v_i are numbers or categories, respectively.

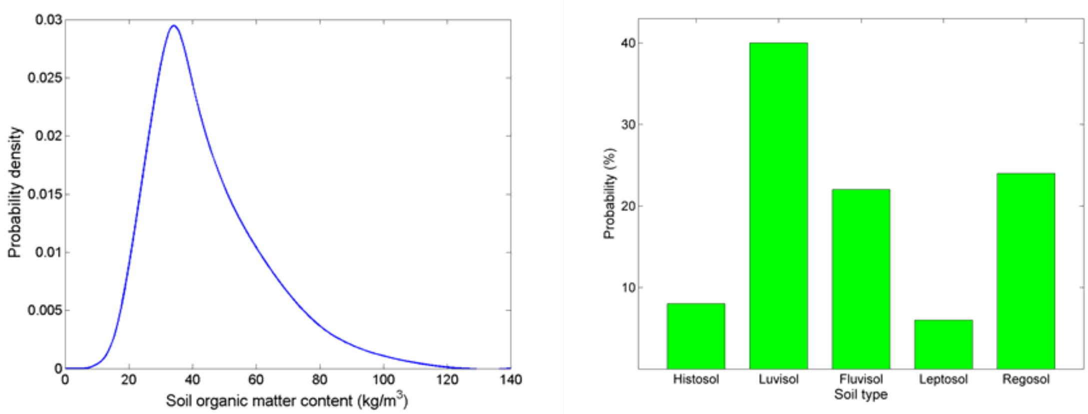

Figure 3 Example continuous probability density for soil organic matter content (left) and discrete probability distribution for soil type (right).

The equations above present the general case, where the only restriction on the functions is that they are valid cdf's or pdf's. How are these functions obtained in practice? Several approaches can be used, depending on the amount and type of information available, but in all cases assumptions and simplifications will have to be made to obtain reliable estimates of the pdf.

- Parameterise the cdf of an uncertain continuous numerical variable to a known shape, usually the normal distribution. Estimate the parameters of the cdf (e.g. mean, variance, skewness) from available data or derive them from expert knowledge.

Non-normal distributions may be used, but perhaps easier is to assume that some transformation of the input is normally distributed and define the uncertainty in terms of the normally distributed transformed variable. Common transformations are square root, logarithm and the Box-Cox transform. Hard minima and maxima can be imposed by using truncated normal distribution (discard unrealistic simulated values afterwards, see step 7 below). In fact, for uncertain inputs that are spatially and/or temporally correlated, the normal distribution is imposed here because without it the UQ/UA analysis would become too complex:

- Characterise (transformed) uncertain continuous numerical inputs that are correlated in space and/or time with the normal distribution. Assume second-order stationarity, possibly after trend removal and/or stratification. Derive the parameters (i.e. mean, variance, semivariogram and/or correlogram) from literature, observations or with expert judgement.

Second-order stationarity (Goovaerts 1997, Diggle and Ribeiro 2007) is needed to limit the number of parameters that characterise the cdf. For example, a digital terrain model may be constructed from a limited sample of elevation measurements. Using expert judgement, the elevation uncertainty may be assumed 'second-order stationary', indicating that the associated cdf has a variance that is spatially invariant (i.e. constant) and for which the spatial correlation between the errors at two locations depends only on the distance between the locations. In that case, the correlations may be estimated from a simple function (variogram), which can be fitted directly to the sample data. The same approach may be used to quantify temporal correlation, although in time series analysis it is customary to quantify temporal correlation directly by means of a correlogram or indirectly through Auto Regressive Moving Average models (e.g., Box et al., 1994).

- For discrete numerical and categorical uncertain inputs characterised by a pdf, list all possible outcomes and assign probabilities to each of these. When the input is dynamic or spatially distributed, either assume spatial/temporal statistical independence (i.e. zero correlation) or assume perfect dependence, possibly within subregions (i.e. constant input within subregions and independence between subregions).

The assumption of complete dependence or complete independence can be relaxed but substantially increases complexity. Possible approaches in the space domain are indicator geostatistics (Goovaerts 1997, Finke et al. 1999, Kyriakidis and Dungan 2001), Bayesian Maximum Entropy (Bogaert 2002, Brus et al. 2008) and Markov random fields (Norberg et al. 2002, Banerjee et al. 2004). These approaches may also be used in the time domain, in which case approaches that make use of the Markov property (i.e. Markov chains that use transition probability matrices) become more feasible and attractive (Davis 2003).

For most categorical variables, a parametric shape may not be available. In that case, each possible outcome and its associated probability must be listed in a ‘non-parametric’ pdf. For example, the pdf of the land use at some location will not satisfy a common shape, implying that the probability of each possible outcome must be specified. These probabilities may be estimated with a ‘people-driven’ approach (i.e., expert elicitation (Cooke 1991, Ayyub 2001) or ‘data-driven’ approaches (e.g., through a confusion matrix computed on a comparison of observed and predicted land uses). The latter assumes stationarity, i.e. it is assumed that given the estimated land use, the pdf is independent of location).

- Generate and store a sample of possible realities for those uncertain inputs that are characterised by a sample (and not by a pdf or cdf). Make sure that the sample is large enough (see step 8 below).

In practice this is usually not difficult because samples are automatically generated as output of Monte Carlo UQ/UA and Bayesian calibration. Uncertainty characterisation by samples will most often be restricted to inputs that are the result of this type of analysis. The sample approach may also be used when a real-world sample is available. For instance, the uncertainty about the soil hydraulic properties at any location within a soil type unit may be represented by the variability within a sample of observations taken within that unit (Heuvelink et al. 2008). However, it is important to be aware that such an approach may be hampered by small sample size or non-random selection of sampling sites (in which case the sample may not be an accurate representation of the population) and lack of taking spatial/temporal correlation into account.

Many model calibration techniques not only provide estimates of the model parameters but also quantify the associated parameter uncertainty. Traditional approaches use the variance-covariance matrix derived from the Jacobian, but in recent years Bayesian calibration has become increasingly popular. In Bayesian calibration, prior distributions of the model parameters are updated with observations to compute a posterior distribution (Van Oijen et al. 2005, Reinds et al. 2008). Typically, numerical solution approaches based on Markov Chain Monte Carlo are used, which have the pleasant property that simulated values (i.e., a sample) of the uncertain parameters are automatically generated. Therefore:

- Use Bayesian calibration for uncertain model parameters that are calibrated.

The Bayesian calibration methodology is theoretically sound and has proven merit in practice, but some fundamental choices have to be made for which further guidance needs to be developed. For instance, how is decided which parameters are treated uncertain and which not? When a parameter is treated as uncertain, is it then assumed to be constant over space and time or can it vary in space and/or time? If it is allowed to vary, what spatio-temporal regularity assumption (e.g. smooth behaviour, constant value within sub-regions) is imposed to prevent that the number of degrees of freedom becomes too large?

When multiple inputs are uncertain, their cross-correlations must also be addressed. For instance, tree properties such as root biomass, stem diameter and height are correlated and so will the uncertainty about these properties. For normally distributed variables, these dependencies are fully characterised with a correlation matrix. For spatially and/or temporally distributed variables, invoking a stationarity assumption leads to a cross-correlation function, which can be estimated from paired observations using geostatistical methodologies as used in cokriging (Goovaerts 1997). Cross-dependencies between uncertain categorical variables are addressed by treating all combinations of categories as separate classes. This increases the number of classes dramatically, so that generalisation is often needed in practice and the total number of combined classes is reduced to ten or fewer. Finke et al. (1999) and Kros et al. (1999) used this approach to handle dependencies between land use and soil type.

- Represent statistical dependence between uncertain categorical inputs by defining probabilities for combinations in cross-tables. Restrict statistical dependence between continuous inputs to the normal distribution and use correlation matrices. Use cross-variograms for spatially distributed uncertain variables. Ignore correlation between uncertain inputs represented by samples unless the correlation was included in the sample generation. Ignore correlations between inputs represented by a cdf and inputs represented by a sample.

Finally, a list of approaches is given that may be used to estimate the parameters of probability distributions of uncertain inputs.

- Use one or several of the following sources of information to estimate the parameters of the cdf's and pdf's of uncertain inputs:
    1. Specification of instrument accuracy and precision (when inputs are measured with instruments).
    2. Sampling error using sampling theory from statistics (e.g. standard error of the mean, confidence intervals).
    3. Kriging prediction variance (when inputs are interpolated from point observations using geostatistical interpolation).
    4. Classification error (when categorical inputs are obtained with multivariate statistics methods such as maximum likelihood classification).

5. Use ground truth verification data to compute parameters of the cdf (e.g. use of Mean Error and Root Mean Squared error of DEM, soils data bases providing variability of soil properties within soil mapping units).
6. Goodness-of-fit measures such as the coefficient of determination, R^2 (when inputs are derived from other variables through regression and $(1-R^2)$ measures the variance reduction
7. Change of support effects using (geo)statistical upscaling and downscaling (see also step 8 below).
8. Use expert elicitation / educated guesses.

Despite the diversity of approaches, the identification and estimation of probability models is, in practice, a difficult task. Useful references with nitrogen modelling applications are De Vries et al. (2003), Kroeze et al. (2003), Oenema et al. (2003) and Miehle et al. (2006).

2.5 When model structure uncertainty is considered, decide how it is expressed and characterise it with a probability distribution

Model structural uncertainty can be quantified in two different ways. One approach uses Bayesian model comparison (Kass and Raftery 1995). In this case one needs multiple models that all describe the same phenomenon. Each model is assigned a prior probability of being the true model. These probabilities are updated using a Bayesian calibration approach. The approach is described in detail in Van Oijen (2008).

When there is only a single model, Bayesian model comparison cannot be used. In such a case a practical approach is to represent the model structural error by an additive stochastic residual. A typical example is linear regression:

$$Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_n X_n + \varepsilon \quad (5)$$

Here the residual ε represents the deviations of the linear model from the real relationship between the dependent Y (model output) and predictors X_i (model inputs). A similar approach is used in state-space models such as used in Kalman filtering (Kalman 1960), where the system noise is part of the so-called state equation and represents the deviation of the model structure from the true process. In this case the residual is dynamic. It is typically represented by a zero-mean normal distribution, which is assumed uncorrelated over time. Extension to space-time can be done (e.g. Heuvelink et al. 2006), but identification of the space-time correlations is difficult and stationarity or independence are often assumed.

In summary:

- Represent model structural uncertainty either by introducing an additive stochastic residual or by using Bayesian model comparison. The latter requires that multiple sufficiently independent models are available, the former is a simpler approach that makes various assumptions that may not always be very realistic.

2.6 *Generate many possible realities of all uncertainty sources by repeated sampling from their probability distributions*

The Monte Carlo UQ/UA requires that realisations from the probability distribution of uncertain model inputs, and possibly the model residual, are generated. When uncertainty is represented as samples (see step 4 above), then these realisations are already available. When uncertainty is represented by cdfs or pdfs, then realisations must be drawn from them using a pseudo-random number generator, such as the ‘Mersenne Twister’ algorithm (Matsumoto and Nishimura 1998). It is easy to draw random numbers from marginal pdfs (i.e. a pdf of a single input or parameter). Most statistical software tools and programming languages can draw random numbers from univariate pdfs. Joint pdfs usually have correlations between the components that must be preserved (one cannot sample each component independently). This can be taken care of fairly easily in the case of a joint normal distribution that has no more than a few hundred components, by applying a Cholesky decomposition of the covariance matrix (Brown and Heuvelink 2007). For larger joint normal pdfs one can rely on sequential stochastic simulation (Goovaerts 1997). Here, variables are visited sequentially, and for each variable the conditional probability distribution is computed, using the already simulated variables as conditioning data. A realisation is then simulated from the conditional pdf, the simulated value is added to the conditioning data set and the process is repeated for the next variable until all variables are simulated.

- Use a pseudo-random number generator to sample from pdfs and cdfs of uncertain inputs and stochastic model residuals. Use Cholesky decomposition or conditional simulation to take cross-correlation and space-time correlation into account.

For Monte Carlo UQ/UA multiple realisations of the uncertain inputs and parameters are needed. The most common sampling method is simple random sampling (De Gruijter et al. 2007), in which case each sample is generated independently, each time simulating from the same pdf. In case the input is high-dimensional, straightforward random drawing from the input pdf may not yield a sample that is representative across the whole range of allowable values for each input, unless the sample size is extremely large. It may then be useful to use Latin Hypercube Sampling (LHS), which is a form of stratified sampling where values are selected for each input from each of a pre-specified number of equal-probability intervals (Iman and Conover 1980). The latter can also be done in a spatial context (Pebesma and Heuvelink 1999). However, LHS slightly distorts the correlation structure and estimates of estimation accuracy are not easily obtained.

The Monte Carlo sample size must be large enough to guarantee that the outcome of the UQ/UA is sufficiently accurate. The required sample size cannot be calculated beforehand because it is derived from the Monte Carlo output. For instance, when the purpose of the UQ is to compute the variance of the model output, then this is estimated from the Monte Carlo sample of outputs. In case of simple random sampling, the accuracy (i.e. variance) of the estimation error can easily be computed as (Heuvelink 1998):

$$\mathrm{Var}(S^2 - \sigma^2) \cong \frac{\mu_4 - \sigma^4}{N} \tag{6}$$

where σ^2 is the output variance, S^2 its estimate, μ_4 the fourth central moment of the output and N the Monte Carlo sample size. Eq. (6) shows that the variance is inversely proportional to the sample size, and so are the variances of estimated percentiles of the output distribution. Thus, to halve the standard error, one must quadruple the number of Monte Carlo runs.

- Use simple random sampling to generate the Monte Carlo sample unless the gain of Latin hypercube sampling or other sampling algorithms justify their use. Compute the required sample size by computing standard errors of the output (using step 7) for a first batch of runs and update these estimated standard errors as new batches come available.

2.7 Run the model for the simulated realities and store the results

This step may be computationally demanding, particularly when the model is complex and requires much computing time. The method is very suited for parallel computing and for grid computing technology (Foster and Kesselman 2003). In this way, computation times can be dramatically reduced provided the equipment is available in large enough numbers.

- Make an assessment of the amount of computing time and storage capacity required to run the model multiple times and store results. Identify bottlenecks and resolve these by extending computing resources and/or storage capacity. If this is not feasible, consider reducing complexity by limiting the spatio-temporal extent or resolution, or by using a simplified model (e.g. by taking default outputs for submodules of the model).

This step also requires efficient setup and organisation of the procedure:

- Streamline the repeated running of the model with different inputs such that the amount of work remains manageable. Run the Monte Carlo analysis in batch mode with efficient and transparent storage and management of the many input and output files that are involved. Start with a small subset to check the entire procedure before running the full analysis. Make use of existing uncertainty propagation tools (e.g. Karssenberg and De Jong 2005, Brown and Heuvelink 2007, Tarantola 2008) if useful.

2.8 Scale up the model outputs to the desired spatial and temporal support

In many practical situations, users and decision makers are interested in averages over space and time rather than in point values or values that are defined over a small support. Although upscaling is a complex issue that has been addressed for many decades without providing a final solution that serves all goals (Bierkens et al. 2000, Stehfest and Bouwman 2006, Bouma et al. 2008), it is not difficult in the case of a Monte Carlo UQ/UA. This is because the analysis produces realisations of the small support outputs that can easily be aggregated to the larger support, simply by computing the average of the small support values contained in the large support:

$$\overline{V} = \frac{1}{N}\sum_{i=1}^{N} V_i \tag{7}$$

where $\overline{V}$ is the desired output at the large support, N is the number of small supports contained within the large support, and where the V_i are the small support model outputs. Note that N need not be chosen equal to the ratio of the large and small supports because this might yield an extremely large number which would lead to prohibitive calculation times, particularly with UQ/UA of complex process-based models. In fact, the sampling error invoked by using a subset of points instead of an exhaustive sample is usually small, even when the number of points (i.e., the sample size) is limited to a few hundred or even less (e.g. Kros et al. 1999 and Heuvelink et al. 2008). Note also that non-linear aggregation, such as computation of the median or the proportion of points within the large support that exceed a threshold, is not difficult. Examples are provided in Heuvelink and Pebesma (1999) and Kros et al. (2002).

- Aggregate the small-support output for the separate Monte Carlo runs over space and/or time before computing summary statistics and graphical illustrations used in step 10. Limit the number of small-support runs within the large support to a maximum of a few hundred when computational burden is a problem.

Disaggregation (downscaling) is much more difficult, but rarely needed within the context of an UQ/UA for regional scale environmental modelling.

2.9 Apply steps 6 to 8 in a modified way when the uncertainty contributions of individual or grouped uncertainty sources are required (i.e. Uncertainty Analysis)

The purpose of Uncertainty Quantification is to analyse how uncertainties in model input and model structure propagate to the model output. However, in many cases a second objective is to determine how individual sources of uncertainty contribute to the output uncertainty. This is an Uncertainty Analysis and provides valuable insight into how the accuracy of model output can best be improved.

The simplest approach to UA is to redo the uncertainty analysis with only those sources made uncertain whose contribution must be estimated. The remaining uncertainty sources are assumed certain and fixed on their reference value. Comparison of the output uncertainties (e.g. ratio of variances) provides a measure of the relative contribution. The disadvantages of this approach are that the uncertainties are only evaluated with the remaining sources fixed on their ‘reference’ value, which need not be the true value, that correlations and dependencies between the uncertainties of the free and fixed sources are ignored, and that the number of Monte Carlo runs is multiplied with the number of uncertainty sources considered (unless Latin hypercube sampling is used, but this will decrease the accuracy of obtained results). Although not all of the disadvantages can be completely avoided, several techniques have been developed to address the problem of uncertainty contributions with more rigour and efficiency. Good sources are Jansen et al. (1994), Saltelli et al. (2000), Crosetto and Tarantola (2001) and Jansen (2005).

- Redo UQ with all uncertainty sources assumed certain except those for which the uncertainty contribution must be assessed. Compute the variance reduction achieved and derive the uncertainty contribution from this. Consult UA literature to use more advanced stochastic sensitivity techniques .

2.10 Communicate the outcomes of the UQ/UA

The Monte Carlo UQ/UA produces a large set of model outputs, which may be interpreted as a random sample from the model output probability distribution. This means that estimates of parameters of the model output pdf, such as the mean, variance, median, interquartile range, percentiles and probabilities of exceeding thresholds can be estimated from the sample using statistical sampling theory. It is not only important that these properties are communicated to users in an efficient way, but also that they are understandable to users that have little background in statistics. Decision makers need help to make use of the uncertainty information and incorporate it in decision making (i.e. risk analysis). Useful references that can be used as starting points are Ehlschlaeger et al. (1997), Janssen et al. (2005) and Van Der Sluijs (2007). For transparency, it is important that the choices made at every preceding step in the 10-step procedure are listed and archived.

- Compute summary statistics (e.g. mean, percentiles, limits of confidence intervals, relative contributions of uncertainty sources) of the Monte Carlo output and present these to users in tables, graphs and maps. Provide ample explanation of what the results mean and how these may be used to assess the accuracy of model results, determine the uncertainty contributions and improve model accuracy.

3 Conclusions

The aim of this document was to provide an operational methodology for Monte Carlo UQ/UA for regional model applications within the NEU project. Although it is impossible to provide full guidance to researchers that wish to apply an UQ/UA within NEU, it lists and describes the ten main steps and summarizes the many actions that make up the UQ/UA in a bulleted list.

The next step is to apply the methodology to selected cases and use these cases as benchmarks for subsequent UQ/UA analyses. Each of the example studies may focus on specific elements of an UQ/UA, such as the identification of input uncertainty, the propagation of input uncertainty, or Bayesian calibration. Jointly, the operational methodology and benchmark examples should provide the necessary background information for researchers who want to apply an UQ/UA within NEU.

References

Ayyub, B.M. (2001), *Elicitation of Expert Opinions for Uncertainty and Risks*. CRS Press, Florida.

Banerjee, S. B.P. Carlin, and A.E. Gelfand (2004), *Hierarchical Modeling and Analysis for Spatial Data*. Chapman and Hall/CRC Press, Boca Raton.

Bierkens, M. F. P., Finke, P. A., and De Willigen, P. (2000), *Upscaling and Downscaling Methods for Environmental Research*. Kluwer, Dordrecht.

Bogaert, P. (2002), Spatial prediction of categorical variables: the Bayesian maximum entropy approach. *Stochastic Environmental Research and Risk Assessment* 16, 425–448.

Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. (1994), *Time Series Analysis: Forecasting and Control*. Prentice–Hall, Englewood Cliffs, NJ.

Bouma, J., J.A. de Vos, M.P.W. Sonneveld, G.B.M. Heuvelink and J.J. Stoorvogel (2008), The role of scientists in multiscale land use analysis: lessons learned from Dutch communities of practice. *Advances in Agronomy* 97, 175–237.

Brown, J.D. and G.B.M. Heuvelink (2007), The Data Uncertainty Engine (DUE): a software tool for assessing and simulating uncertain environmental variables. *Computers & Geosciences* 33, 172–190.

Brown, J.D., G.B.M. Heuvelink and J.C. Refsgaard (2005), An integrated framework for recording uncertainties about environmental data. *Water Science and Technology* 52, 153–160.

Brus, D.J., P. Bogaert and G.B.M. Heuvelink (2008), Bayesian Maximum Entropy prediction of soil categories using a traditional soil map as soft information. *European Journal of Soil Science* 59, 166–177.

Cooke, R.M. (1991), *Experts in Uncertainty: Opinion and Subjective Probability in Science*. Oxford University Press, Oxford.

Crosetto, M. and S. Tarantola (2001), Uncertainty and sensitivity analysis: tools for GIS-based model implementation. *International Journal of Geographical Information Science* 15, 415–437.

Davis, J.C. (2003), *Statistics and Data Analysis in Geology, 3rd edition*. New York: Wiley.

De Gruijter, J.J., D. Brus, M.F.P. Bierkens and M. Knotters (2007), *Sampling for Natural Resource Monitoring*. Berlin: Springer.

De Vries, W. J. Kros, O. Oenema and J. De Klein (2003), Uncertainties in the fate of nitrogen II: A quantitative assessment of the uncertainties in major nitrogen fluxes in the Netherlands. *Nutrient Cycling in Agroecosystems* 66, 71–102.

Diggle, P.J. and P.J. Ribeiro, 2007. *Model-based Geostatistics*. Springer, New York,

Ehlschlaeger, C.R., A.M. Shortridge and M.F. Goodchild (1997), Visualizing spatial data uncertainty using animation. *Computers & Geosciences* 23, 387–395.

Finke, P.A., Wladis, D., Kros, J., Pebesma, E.J., Reinds, G.J. (1999), Quantification and simulation of errors in categorical data for uncertainty analysis of soil acidification modelling. *Geoderma* 93, 177–194.

Foster I. and C. Kesselman (Eds.) (2003), *The Grid 2. Blueprint for a New Computing Infrastructure*. Elsevier.

Goovaerts, P. (1997), *Geostatistics for Natural Resources Evaluation*. New York: Oxford University Press.

Heuvelink, G.B.M. (1998), *Error Propagation in Environmental Modelling with GIS*. London: Taylor & Francis.

Heuvelink, G.B.M. and E.J. Pebesma (1999), Spatial aggregation and soil process modelling. *Geoderma* 89, 47–65.

Heuvelink, G.B.M., J.D. Brown and E.E. Van Loon (2007), Representing and simulating uncertain environmental variables in GIS. *International Journal of GIS* 21, 497–513.

Heuvelink, G.B.M., J.M. Schoorl, A. Veldkamp and D.J. Pennock (2006), Space-time Kalman filtering of soil redistribution. *Geoderma* 133, 124–137.

Heuvelink, G.B.M., F. Van Den Berg, S.L.G.E. Burgers and A. Tiktak (2008), Uncertainty and stochastic sensitivity analysis of the GeoPEARL pesticide leaching model. *Geoderma* (in review).

Iman, R. L., and Conover, W. J. (1980), A distribution free approach to inducing rank correlation among input variables. *Communications in Statistics, Part B-Simulation and Computation* 11, 311–333.

Jansen, M.J.W. (2005), *ASSA: Algorithms for Stochastic Sensitivity Analysis.* Biometris report, Wageningen.

Jansen, M.J.W., W.A.H. Rossing and R.A. Daamen (1994), Monte Carlo estimation of uncertainty contributions from several independent multivariate sources. In: Grasman, J. and G. Van Straten (Eds), *Predictability and Nonlinear Modelling in Natural Sciences and Economics,* Kluwer, Dordrecht: 334-343.

Janssen, P.H.M., A.C. Petersen, J.P. Van Der Sluijs, J. Risbey and J.R. Ravetz (2005), A guidance for assessing and communicating uncertainties. *Water Science and Technology*, 52 125–131.

Kalman, R.E. (1960), A new approach to linear filtering and prediction problems. *Transactions of the American Society of Mechanical Engineers: Journal of Basic Engineering* 82D, 35–45.

Karssenberg, D. and K. De Jong (2005), Dynamic environmental modelling in GIS. 2: Modelling error propagation. *International Journal of Geographical Information Science* 19, 623–637.

Kass, R.E. and A.E. Raftery (1995), Bayes factors. *J.Am. Stat. Assoc.* 90, 773-795.

Kroeze,C., R. Aerts, N. Van Breemen, D. Van Dam, K. Van Der Hoek, P. Hofschreuder, M. Hoosbeek, J. De Klein, H. Kros, H. Van Oene, O. Oenema, A. Tietema, R. Van Der Veeren and W. De Vries (2003), Uncertainties in the fate of nitrogen I: An overview of sources of uncertainty illustrated with a Dutch case study. *Nutrient Cycling in Agroecosystems* 66, 43–69.

Kros, J., E.J. Pebesma, G.J. Reinds and P.F. Finke (1999), Uncertainty assessment in modelling soil acidification at the European scale: a case study. *Journal of Environmental Quality* 28, 366–377.

Kros, J., J.P. Mol-Dijkstra and E.J. Pebesma (2002), Assessment of the prediction error in a large-scale application of a dynamic soil acidification model. *Stochastic Environmental Research and Risk Assessment* 16, 279–306.

Kyriakidis, P.C. and J.L. Dungan (2001), A geostatistical approach for mapping thematic classification accuracy and evaluating the impact of inaccurate spatial data on ecological model predictions. *Environmental and Ecological Statistics* 8, 311–330.

Li, C., S. Frolking S and T.A. Frolking (1992), A model of nitrous oxide evolution from soil driven by rainfall events: 1. Model structure and sensitivity. *J Geophys Res* 97, 9759–9776.

Matsumoto, M. and T. Nishimura (1998), A 623-dimensionally equidistributed uniform pseudorandom number generator. *ACM Transactions on Modeling and Computer Simulation* 8, 3–30.

Miehle, P., S.J. Livesley, C.S. Li, P.M. Feikema, M.A. Adams and S.K. Arndt (2006), Quantifying uncertainty from large-scale model predictions of forest carbon dynamics. *Global Change Biology* 12, 1421–1434.

Norberg, T., L. Rosen, A. Baron En S. Baron (2002), Modeling discrete geological structure as Markov random fields. *Mathematical Geology* 34, 63–77.

Oenema, O., H. Kros and W. De Vries (2003), Approaches and uncertainties in nutrient budgets: implications for nutrient management and environmental policies. European *Journal of Agronomy* 20, 3–16.

Parton, W.J., M. Hartman, D. Ojima and D. Schimel (1998), DAYCENT and its land surface submodel: description and testing. *Global Planet Change* 19, 35–48.

Pebesma, E.J. and G.B.M. Heuvelink (1999), Latin hypercube sampling of Gaussian random fields. *Technometrics* 41, 303–312.

Reinds, G.J., M. Van Oijen, G.B.M. Heuvelink and H. Kros (2008), Bayesian calibration of the VSD soil acidification model using European forest monitoring data. *Geoderma* 146 (3-4), 475-488.

Saltelli, A., K. Chan and E.M. Scott (Eds) (2000), *Sensitivity Analysis*. Wiley, Chichester.

Saltelli, A., S. Tarantola, F. Campolongo and M. Ratto (2004), *Sensitivity Analysis in Practice, A Guide to Assessing Scientific Models.* Wiley, New York.

Stehfest, E. and Bouwman, L. (2006), N2O and NO emission from agricultural fields and soils under natural vegetation: summarizing available measurement data and modeling of global annual emissions. *Nutrient Cycling in Agroecosystems* 74, 207-228.

Tarantola S., (2008), SimLab 2.2 Reference Manual. European Commission – IPSC, http://simlab.jrc.ec.europa.eu/.

Van Der Sluijs, J.P. (2007), Uncertainty and Precaution in Environmental Management: Insights from the UPEM conference. *Environmental Modelling and Software* 22, 590–598.

Van Oijen, M. (2007), *“Protocol-UQ/UA”: A Protocol for Uncertainty Quantification and Uncertainty Analysis of Models used in NEU*. NEU internal report.

Van Oijen, M. (2008), *Bayesian Calibration (BC) and Bayesian Model Comparison (BMC) of process-based models in NitroEurope: Theory, implementation and guidelines*. NEU internal report.

Van Oijen, M., Rougier, J. & Smith, R. (2005), On Bayesian calibration and evaluation of process-based forest models: bridging the gap between models and data. *Tree Physiology* 25, 915–927.

Winiwarter, W. (2007), National greenhouse gas inventories: understanding uncertainties versus potential for improving reliability. *Water Air Soil Pollution* 7, 443–450.