NitroEurope IP

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

D3.1.4

An updated written protocol detailing agreed procedures for UQ/UA.

Due date of deliverable: **01/01/2008**

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)	☐

“Protocol-UQ/UA”:
A protocol for uncertainty quantification and uncertainty analysis of models used in NEU

M. van Oijen and NitroEurope modellers, 14 Feb 2008

Status of this protocol
The previous version of this protocol (31 Oct 2006) was accepted by the modelling community in NitroEurope for the first phase of the project. Subsequent recommendations for change were made at various NitroEurope meetings and electronic discussions.
This version of the protocol will be discussed at the NitroEurope IP 3rd Annual Meeting and General Assembly, February 2008.

1. Introduction

1.1 Uncertainty in NEU

This document is the protocol for Uncertainty Quantification and Uncertainty Analysis (UQ/UA) followed by the modellers in NEU, and will be referred to as “Protocol-UQ/UA”. Uncertainty quantification and uncertainty analysis are essential parts of good modelling practice, for which an overarching protocol exists in NEU, referred to as “Protocol-GMP”. The Protocol-UQ/UA thus can be seen as an appendix to Protocol-GMP. We will not further discuss the many reasons for carrying out UQ/UA here but refer to AEAT/ENV/R/1039 (see paragraph 1.2 below) for supporting arguments.

Originally, in the NEU “Description of Work”, two uncertainty protocols were envisaged, one for plot scale modelling and one for regional scale modelling. However, at the March and May 2006 meetings of NEU the overlap between the two protocols was considered to be too great to justify keeping them separate.

As further agreed at the May 2006 NEU-meeting, the Protocol-UQ/UA (like the Protocol-GMP) is not designed to be too prescriptive. Therefore the protocol largely consists of a list of recommended activities and methods (most of which will not be explained in detail, but summarized with a reference to literature), rather than a rigidly-to-be-followed procedure. Only a small fraction of this protocol consists of activities required from all modellers in NEU, e.g. the general obligation to carry out some form of UQ or other. These required activities are not new additions to everyone’s work plan but were already included in the project plan (NitroEurope IP 2006, Annex 1 – “Description of Work”) and the commitments were confirmed at the kick-off meeting of NEU (see, for example, Campbell & Van Oijen, 2006).

The emphasis of the protocol is on uncertainty itself, not on the role of uncertainty in, for example, model verification/validation (which will be treated in NEU C6.1).

1.2 Uncertainty in established protocols

Various international organisations related to the environmental and natural sciences have drafted protocols for UQ/UA. We mention the following:

“AEAT-2688”	NAEI (1998).Treatment of Uncertainties for National Estimates of Greenhouse Gas Emissions. http://www.aeat.co.uk/netcen/airqual/naei/ipcc/uncertainty/contents.html
“AEAT/ENV/R/1039”	NAEI (2003). Estimation of Uncertainties in the National Atmospheric Emissions Inventory. http://www.airquality.co.uk/archive/reports/cat07/AEAT1039_finaldraft_v2.pdf
“GPG2000”	IPCC (2000). Good Practice Guidance and Uncertainty Management in National Greenhouse Gas Inventories. http://www.ipcc-nggip.iges.or.jp/public/gp/english/
“GPG-LULUCF”	IPCC (2003). Good Practice Guidance for Land Use, Land-Use Change and Forestry. http://www.ipcc-nggip.iges.or.jp/public/gpglulucf/gpglulucf_contents.htm

"GUM"	ISO et al. (1993, corrected and reprinted 1995). Guide to the Expression of Uncertainty in Measurement.
"NIST-TN1297"	NIST (1994). Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results. NIST Technical Note 1297. http://physics.nist.gov/Document/tn1297.pdf
"UK-GHG-1990-1999-A8"	NETCEN (2001). UK Greenhouse Gas Inventory, 1990 to 1999, Appendix 8, Uncertainties. http://www.aeat.co.uk/netcen/airqual/reports/ghg/ukghgi_90-99_append_7-9.pdf

The Protocol-UQ/UA is consistent with these documents but does not repeat what is stated in them. In sections 4-9, we refer to the documents where they provide additional information.

2. Materials

This protocol is partly based on the following NEU-documents, in chronological order:

- NitroEurope IP (2006). Annex I – "Description of Work"
- De Bruijn & Butterbach-Bahl (2006). Current status and uncertainties in plot scale models for Nr and GHG. Minutes of NEU kick-off meeting parallel session 14 March, 9.00h.
- Verweij & de Vries (2006). Defining good practice in numerical modelling: includes establishment of "Task force on Good Modelling Protocols". Minutes of NEU kick-off meeting parallel session 14 March, 16.00h.
- Campbell & Van Oijen (2006). Uncertainty assessment in data and modelling. Minutes of NEU kick-off meeting Working Group 7, 16 March 8.30h.
- Butterbach-Bahl (2006). Scientific report of the joined COST-NitroEurope Workshop in Garmisch-Partenkirchen (03.05.-05.05.2006) on "Quantifying and assessing parametric and non-parametric model uncertainty in models".
- Van Oijen (2006). Modelling in NEU: QA/QC, task force and protocols. Discussion paper 4 Aug. 2006: 3 pp.

Further, the protocol builds on the pre-existing protocols described in section 1.2 and on literature listed in the References.

3. Definitions

3.1 *Terminology used in this protocol*

For the purposes of this protocol we use the following definitions:

Input	All the information needed to run a *model* that is not incorporated in the model itself. Input is of three types: (1) Initial values (= values of state variables at start of simulation), (2) Model parameters, (3) Environmental constants and variables.
Model	A computer program that transforms *input* into *output*.
Output	*Model* results for given *input*.
Uncertainty	Incomplete knowledge. Uncertainty is of three types: (1) *input* uncertainty, (2) uncertainty about *model* structure, (3) *output* uncertainty.
UA	*Uncertainty* analysis, i.e. attribution of overall *output* uncertainty (whose magnitude is determined in the process we call *UQ*) to the different *input* uncertainties. UA is specific to a given model and does not address possible errors in coding or model structure.
UQ	*Uncertainty* quantification, i.e. quantification of the *model output* uncertainty caused by uncertainty in the *inputs*. The degree to which each input, through the uncertainty associated with it, is responsible for output uncertainty, is determined in the process we call *UA*. UQ does not quantify output uncertainty associated with uncertainty about coding or model structure.

3.2 *Definitions recommended for use within NitroEurope*

For the purposes of communication within NitroEurope, we suggest that the following definitions are used in addition to the protocol-terminology given in the preceding paragraph.

[Note: the definitions were taken from GPG2000 (for full reference and weblink, see section 1.2). Most of the definitions were not written for GPG2000 but taken from some of the other sources mentioned in section 1.2, and in particular from GUM (ISO 1993, 1995).]

Bias — A systematic error of the observation method, whose value in most cases is unknown. It can be introduced by using measuring equipment that is improperly calibrated, by selecting items from a wrong population or by favouring certain elements of a population, etc. In statistical terms, bias is the difference between the expected value of a statistic and the parameter which it estimates.

Error — The difference between an observed (measured) value of a quantity and its 'true' (but usually unknown) value.

Sensitivity — A measure of how responsive one quantity is to a change in another related quantity. The sensitivity of a quantity Y that is affected by changes in another quantity X, is defined as the change in Y divided by the change in X that caused the changes in Y.

Systematic and Random Errors

Systematic error is the difference between the true, but usually unknown, value of a quantity being measured, and the mean observed value as would be estimated by the sample mean of an infinite set of observations. The random error of an individual measurement is the difference between an individual measurement and the above limiting value of the sample mean.

Variability — This refers to observed differences attributable to true heterogeneity or diversity in a population. Variability derives from processes which are either inherently random or whose nature and effects are influential but unknown. Variability is not usually reducible by further measurement or study, but can be characterised by quantities such as the sample variance.

3.3 *Other glossaries*

Other useful glossaries of terms in modelling and statistics are:

Holland, D.M., W.M. Cox, R. Scheffe, A.J. Cimorelli, D. Nychka, and P.K. Hopke, (2003). Spatial Prediction of Air Quality Data, EM, August 2003, 31-35.

Minka, T. — A Statistical Learning/Pattern Recognition Glossary.
http://research.microsoft.com/~minka/statlearn/glossary/glossary.html

NIST — Engineering Statistics Handbook
http://www.itl.nist.gov/div898/handbook/glossary.htm

Collections of glossaries can be found in:

NASA — Global Change Master Directory List of Earth Science Acronyms, Glossaries and Gazetteers
http://gcmd.nasa.gov/Resources/FAQs/acronyms.html

NOAA — Climate Glossaries and Weather Tools
http://www.cdc.noaa.gov/ClimateInfo/tools.html

NVTB — Glossaries, Dictionaries and Encyclopedias
http://www.bio.vu.nl/nvtb/Dict.html

4. The Protocol-UQ/UA in brief

In this section we provide a brief summary of the protocol. The subsequent sections will provide more detail.

4.1 ***In NEU, all modellers are committed to:***

- **Carry out UQ of their model and report: (1) The chosen UQ-method, (2) Output uncertainty. The UQ-method is chosen by the modeller but can be one of the recommended methods described in the following sections. In reporting on UQ, the model version and data used for UQ need to be documented – preferably by storing copies of the relevant files in read-only format. UQ is carried out at least twice: near beginning (year 1 for plot-scale modellers, year 2 for regional scale modellers) and end of the NEU-project, but it is recommended to do this whenever significant new data have become available and whenever the model has changed.**

- **Carry out UA for their model and report: (1) The chosen UA-method, (2) The inputs that account for most of output uncertainty. It is recommended to carry out UA after each UQ and use the results for improvement of the modelling as well as to guide the collection of data.**

4.2 *In NEU, all modellers are recommended to:*

- Carry out a model comparison, possibly in collaboration with other groups, and report: (1) The method used for model comparison, (2) The relative probabilities of different models, model versions and/or model algorithms (i.e. submodels for specific processes or groups of processes) of being correct. Systematic comparison of completely different models is recommended for near the end of the project, systematic comparison of different model versions is recommended to be repeated throughout the period of model development in the first years of NEU.

For similar recommendations concerning UQ, see AEAT/ENV/R/1039.

5. UQ step 1: Quantification of uncertainty in model inputs

5.1 *Probability distribution functions (pdf's)*

UQ quantifies the propagation of model input uncertainty to outputs, so the first step is quantifying the uncertainty in the model inputs themselves. Ideally, we would like to express our collective uncertainty about all model inputs in the form of a single joint probability distribution function (pdf). In practice, it is more convenient to specify the uncertainty about each input separately, in its own "marginal" pdf, and assume independence of inputs, unless we know that certain combinations of inputs are more likely than others, in which case we need to specify correlations as well. For example, for spatially distributed variables or dynamic variables one must include spatial and/or temporal autocorrelation: see standard geostatistical and time series texts on how to represent and estimate space-time correlations.

More information about specifying pdf's is given in GPG2000 (§§ 6.1, 6.5). The GPG2000 also gives examples of quantifying uncertainty for emission factors of greenhouse gases like N_2O (§§ 4.7.1.6, 4.8.1.6). Similar examples are given in GPG-LULUCF (§§ 3, Appendix 3a.2). NIST-TN1297 shows how to quantify uncertainty for physical measurements.

5.2 *Expert judgment*

For many model inputs the available information may be very limited. A model may, for example, contain various parameters for which no measurements are available or which can not even be measured at all. In such cases expressing the pdf becomes a matter of expert judgment (by modeller or colleague) and every effort should be made to quantify the expert's uncertainty fairly to avoid subsequent under- or overestimation of model output uncertainty.

More details on eliciting expert judgment can be found in GPG2000 (§§ 4.8.1.6, 6.2.5)

5.3 *From data to pdf*

When data on inputs are available, formulating the appropriate pdf can still be very complicated as the uncertainty arises from many sources, e.g. imprecise or inaccurate measurement, sampling, upscaling, interpolation etc.

We refer to Heuvelink (1998), GPG2000 (§§ A1.3 A1.4) and GPG-LULUCF (§4.2.4.2.1) for more details on sources of data uncertainty and to Brown and Heuvelink (in press) for the freely available DUE-software that can assist in quantifying input uncertainty.

5.4 *Bayesian approach*

Quantifying input uncertainty accurately is especially critical if UQ is seen as a uni-directional procedure, from inputs to outputs. However, the modeller in NEU wishing to quantify uncertainty may opt to follow a Bayesian approach which is two-directional and includes a flow of information from measurements on outputs to the pdf for inputs. In Bayesian calibration, large uncertainty about inputs may gradually diminish when measurements on outputs show that certain input values are less likely than others.

For more details on the Bayesian approach see Van Oijen et al, 2005, section 7 below, and Van Oijen (2008), the latter document being available from the NitroEurope website.

6. UQ step 2: Uncertainty propagation from inputs to outputs

6.1 *Monte Carlo (MC)*

A model transforms input into output, so it also transforms input uncertainty into output uncertainty. To quantify output uncertainty, we can begin by taking a random sample from the input pdf consisting for example of hundreds or more choices for the inputs. Taking such a sample and running the model for each member of the sample is the so-called Monte Carlo method of UQ. The result of the MC exercise is a sample from the pdf for model output. The output sample can be summarized graphically in the form of histograms for individual outputs or statistically by output means and their variance-covariance matrix.

For more details on MC methods see GPG2000 (§§ 6.3-6.5, A1.4.3.2) and GPG-LULUCF (§5.2.2). A detailed example of an UQ using MC – for the UK Greenhouse Gas Inventory – is given in UK-GHG-1990-1999-A8.

6.2 *Latin Hypercube Sampling*

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.

For more details on LHS see GPG2000 (§6.5) and Pebesma & Heuvelink (1999), the latter for LHS in a spatial context.

6.3 *Screening*

To reduce sample size it may also be practical to first identify the major inputs that contribute to output uncertainty and thereafter ignore the others in the subsequent UQ. 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).

7. UQ step 3: Updating uncertainties when new data become available

7.1 *Non-Bayesian approaches*

When new data about model inputs become available at a later stage, after a first UQ has already been carried out, it is possible to simple start again from scratch and do a new UQ as explained in the previous two sections.

7.2 *Bayesian Calibration (BC): general approach and MCMC*

Bayesian Calibration is a means of updating the input pdf directly when new data come in, not by starting again from scratch with UQ but by applying Bayes Theorem to the input pdf that was created before. In Bayesian terminology, the input pdf that was created before is called the “prior” pdf. If we later receive new data, on inputs or outputs or both, with associated data-uncertainty, Bayes Theorem tells us how to modify the prior accordingly, thus producing a so-called “posterior” pdf, which will include correlations between inputs. When later again new data come in, the posterior becomes the prior in a new cycle of BC leading to a further refined posterior pdf. BC thus is a continuous learning process well suited for environmental models that are repeatedly confronted with uncertain new data. BC typically uses a special form of Monte Carlo sampling, called Markov Chain Monte Carlo (MCMC) to generate samples from the posterior pdf.

Details of BC and MCMC are provided by Van Oijen et al (2005) and a guidance document with example programs can be downloaded from the NEU website (Van Oijen, 2008).

7.3 *Bayesian Calibration (BC): universal vs site-specific inputs*

In BC, we should clearly distinguish universal inputs from site-specific ones. For example, some process parameters may be considered to have the same value everywhere whereas initial values of state variables are typically site-specific. When using the posterior pdf generated at one site as the prior for further calibration on data from another site, we can only use that part of the first posterior pdf that relates to the universal parameters.

8. Uncertainty concerning model structure

8.1 *Bayesian Model Comparison (BMC)*

Bayesian Calibration has a straightforward extension to model comparison. Whereas BC (like any other method for UQ) requires defining a prior pdf for model inputs, BMC requires defining a prior pdf over the set of models that are going to be compared. The standard choice for this is initially assigning each model in the comparison the same probability of being the correct one. After that the procedure for generating the posterior pdf for models (rather than model inputs) follows the same method as BC, including the use of MCMC. The result of BMC is an objective statement about the probability of each model in a set of being the correct one given the information available to the BMC. Uncertainty about model structure is highest when the different models all have equal probability.

For more details on Bayesian Model Comparison see Kass & Raftery (1995).

8.2 *MODEVAL.XLS*

The EXCEL-spreadsheet program MODEVAL.XLS (Smith et al. 1997; program available on request from Jo & Pete Smith) provides another means of comparing different models, by calculating various measures of goodness-of-fit of the models to data. Goodness-of-fit (GOF) statistics do not provide a formal quantification of the relative probabilities of the correctness of models but the GOF-statistics are easily calculated and commonly used indicators of model quality.

9. UA

9.1 *Uncertainty Analysis (UA)*

UA is attribution of output uncertainty to the uncertainties of the different inputs. Many highly technical methods for UA exist, including ones based on analysis of variance (ANOVA) (e.g. Jansen et al, 1994). Here we recommend a simple and generally applicable method that is ANOVA-like in essence but less formal. The method consists of simply repeating the UQ as many times as there are different inputs, with each time one of the inputs assumed to be completely known (i.e. not having any associated uncertainty). Thereafter, we can tabulate, for each of the main output variables of the model, which inputs affect output uncertainty the most as quantified by a reduced output standard deviation.

For documentation on identifying key sources of uncertainty see GPG2000 (§7.2.1.2) and GPG-LULUCF (§5.4.2.2).

10. References (see also section 1.2 for other protocols and section 2 for NEU-documents)

Brown, J.D. & Heuvelink, G.B.M. (in press). The Data Uncertainty Engine (DUE): a software tool for assessing and simulating uncertain environmental variables. Computers and Geosciences.

Heuvelink, G.B.M. (1998). Uncertainty analysis in environmental modelling under a change of spatial scale. Nutrient Cycling in Agroecosystems 50: 255-264.

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

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

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

Saltelli, A., Chan, K. & Scott, E.M. (Eds) (2000). Sensitivity Analysis. Wiley, Chichester: 475 pp.

Smith, P., Smith, J.U. & Powlson et al. (1997) Evaluation and comparison of nine soil organic matter models using datasets from seven long-term experiments. In: Evaluation of soil organic matter models using datasets from seven long-term experiments. (eds: P. Smith, D.S. Powlson, J.U. Smith & E.T. Elliott) Geoderma 81: 153-225.

Tedeschi, L.O. (2006). Assessment of the adequacy of models. Agric. Syst. 89: 225-247.

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.

Van Oijen, M. (2008). Bayesian Calibration (BC) and Bayesian Model Comparison (BMC) of process-based models in NitroEurope: Theory, implementation and guidelines. Internal Report NitroEurope: 10 pp. Available from the NitroEurope website.