Realistic quantification of input, parameter and structural errors of soil process models

Gerard B.M. Heuvelink[A]

[A]Environmental Sciences Group, Wageningen University and Research Centre, Wageningen, The Netherlands, Email gerard.heuvelink@wur.nl

Abstract
Error propagation analysis with soil process models requires realistic quantification of errors in model inputs, model parameters and model structure. Once this is achieved, the error propagation analysis itself is relatively straightforward, and can for instance be done by employing a Monte Carlo simulation approach. Input error assessment is often complicated because it must include spatial, temporal and cross-correlations of input errors and must assess these at the right spatio-temporal support. Data-driven methods are preferred, but when data availability is poor, a people-driven method using expert elicitation can be used. Errors in model parameters can best be derived using Bayesian calibration, which requires that sufficient model output observations are available at the right support, and that the calibration procedure accounts for model input and structural errors. Bayesian model averaging is advocated for model structural error quantification, but this will only work when multiple models are available that cover the entire space of plausible models. If this cannot be guaranteed, a more sensible approach is to use a stochastic model that incorporates model structural error as system noise.

Key Words
Pedometrics, expert elicitation, bayesian calibration, bayesian model averaging, stochastic systems theory.

Introduction
Soil scientists know better than anyone else that the outputs of soil process models are not perfect. The reasons are well known: there are errors in the model input, model parameters and model structure. These errors propagate through the model in ways that often cannot easily be predicted without the help of specific tools. Therefore, in the past decades many approaches have been developed, implemented and applied to analyse error propagation in environmental and soil process models (e.g. Hyvonen *et al.* 1998; Bishop *et al.* 2006; Brown and Heuvelink 2007; Castrignano *et al.* 2008; Dean *et al.* 2009; Heuvelink *et al.* 2009). The most flexible and most often used approach is the Monte Carlo method, which is remarkably simple and easily implemented. First, the errors about the various ‘inputs’ to the model are characterized by probability distributions. Next, a pseudo-random number generator is used to sample from these distributions, and the model is run with the sampled inputs. This process is repeated many times, each time running the model with a new sample of inputs and storing the result. The spread in the so-obtained set of model outputs characterizes the model output error.

Although error propagation analysis with the Monte Carlo method may look simple and straightforward, it turns out to be difficult when concepts are to be put into practice. Important challenges are:
- realistic quantification of error in model inputs, parameters and model structure;
- keeping the required computation time within acceptable bounds;
- ensuring that all important error sources are included in the analysis;
- controlling the Monte Carlo sampling error;
- assessing the contribution of individual error sources to the output error;
- assessing error in spatio-temporal aggregates of model outputs;
- validation of the outcome of an error propagation analysis.

In this paper we only address the first of these challenges, because this is arguably the most crucial problem and space limitations prohibit a comprehensive analysis of all challenges. However, it should be noted that all are important and deserve attention. Also, the list may not be exhaustive.

Realistic quantification of error in model inputs, parameters and model structure
Although the distinction between model inputs and model parameters is not always obvious and models may have ‘inputs’ that are in the ‘grey zone’ between input and parameter (e.g. hydraulic conductivity,

weathering rate), it is useful to separate error assessment for inputs from that for parameters. Inputs are defined as real-world properties that exist regardless of the model and can in principle be observed. Parameters are only defined within the context of a model and loose their meaning when there is no model (e.g. regression coefficients).

Input error assessment

Probability distributions associated with errors in model inputs can be derived in various ways, such as by analysing replicates in a laboratory to quantify laboratory measurement error, comparison of ground-truth data with mapped data to assess generalisation and classification errors, and use of geostatistics to quantify spatial interpolation error. These ‘data-driven’ methods are well-developed and are continuously improved, such as in geostatistics where the ordinary kriging paradigm has gradually been replaced by more elaborate approaches such as regression kriging (Hengl *et al.* 2004) and generalized linear models for geostatistical data (Diggle and Ribeiro 2007). Basically, improvement focuses on making more realistic assumptions. For instance, ordinary kriging assumes that the soil property of interest is a realization of a second-order stationary random function that has a constant mean (Webster 2000), whereas regression kriging allows the mean of the soil property to depend on external explanatory variables. Note, however, that assumptions must always be made, because the amount of data is insufficient to uniquely derive the entire probability distribution of the input error, which should include spatial, temporal and cross-correlations when relevant. One important issue that is rarely addressed in data-driven approaches but that needs attention is that the data used to quantify the error in the model input may have non-negligible observation error. Input error will be systematically overestimated if this is ignored. Pedometricians know that the ‘support’ of the observations is also crucially important when deriving error distributions. For instance, the error associated with the nitrate concentration of the soil solution at a ‘point’ in space and time is much larger than that associated with the annual average of an entire field because ‘hot spots’ in time and space will average out over the larger support. Thus, it is imperative that input error quantification is done at the support required by the model (Heuvelink 1998). Data-driven approaches are less developed for categorical soil properties. Only few approaches exist that derive the entire probability distribution of spatially distributed categorical variables (e.g. Finke *et al.* 1999; Hartman 2006; Brus *et al.* 2008), and most of these are cumbersome, make unrealistic assumptions or have severe limitations.

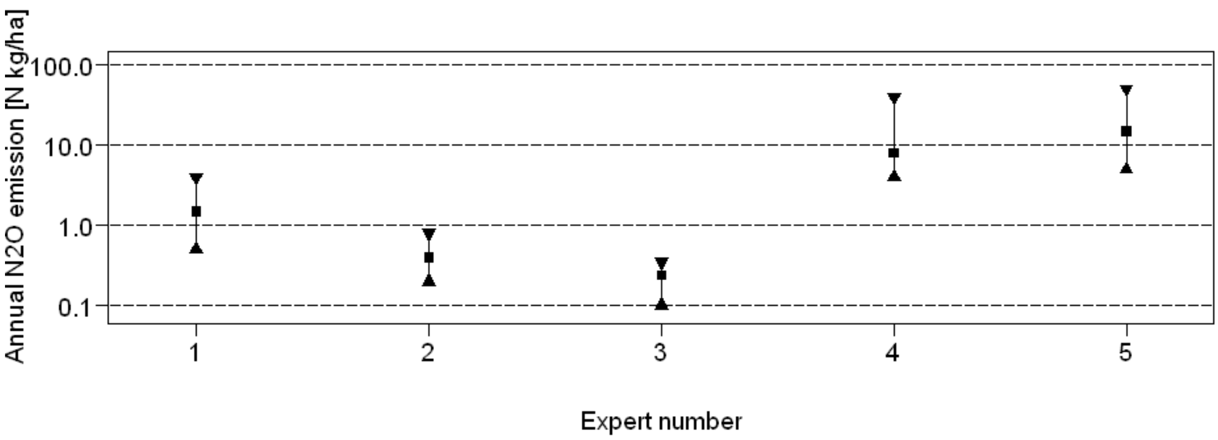

Figure 1. Error in annual N_2O emission (kg N/ha) from 1 ha plots on arable land on clay soils across Europe, independently estimated by five experts. The probability distribution of the error is characterized by the 25 (triangle point up), 50 (square) and 75 (triangle point down) percentiles.

Although the data-driven approach is the preferred option, in many practical cases it may fail for lack of sufficient data, leaving the ‘people-driven’ approach as the only alternative (Brown and Heuvelink 2005). Here, expert elicitation is used to derive probability distributions of model inputs. As an example, Figure 1 reports the quantified error about the annual nitrous-oxide emission (kg N/ha) for 1 ha plots on arable land on clay soils across Europe, estimated independently by five experts (Shang 2009). There is much disagreement between experts, which makes it difficult to merge their assessments. Figure 1 also shows that it is risky to rely on just one expert, which seems to be the common approach in people-driven assessment of input error (e.g. De Vries *et al.* 2003; Lesschen *et al.* 2007). It is imperative that we learn more about expert elicitation, which is well-developed in the risk analysis literature (e.g. Kaplan 1992; Ayyub 2001; Cooke and Goossens 2004). It must be adapted to the type of applications which pertain to soil science and extended to the quantification of support-dependent spatial- and cross-correlations.

Parameter error assessment
Errors in model parameters can only be realistically assessed by inverse methods, in which model predictions are compared with model output observations and parameter error is assigned such that it explains the observed differences. Common approaches are PEST (http://www.sspa.com/pest/) and GLUE (Dean *et al.* 2009). Recently, Bayesian calibration was introduced to the environmental sciences and also to soil science (Reinds *et al.* 2007). Starting with user-defined a priori probability distributions, Bayesian calibration uses Markov Chain Monte Carlo methods to update these distributions with information derived from the observations. Bayesian calibration is attractive because it is flexible, mathematically sound, easily implemented, and yields the full joint probability distribution of the model parameters. It is computationally demanding and standard application ignores the contribution of errors in inputs, model structure and parameters not included in the analysis. Also, the decision whether to assume that parameters are constant or variable in space and/or time turns out to be crucial (e.g. Reinds *et al.* 2007). Pedometricians and soil process modellers must take a closer look at these issues and ensure that the methodology is properly applied.

Model structure error assessment
Bayesian calibration has been extended to include model structural error. This is known as Bayesian model comparison or Bayesian model averaging (Raftery *et al.* 1997). Multiple models are considered and each gets assigned a prior probability of being the 'true' model. Next these prior probabilities are updated to posterior probabilities based on the amount of agreement between observed and predicted model outputs. The methodology works well with statistical (regression) models, where a large number of candidate models can easily be formulated simply by including or excluding explanatory variables, but extension to physically-based models is cumbersome. Refsgaard *et al.* (2006) present a framework for dealing with model structural error in hydrological modelling that uses multiple model structures, but acknowledge that the range of models must span the entire space of plausible models. The latter will be difficult in practice, because most models borrow concepts from each other, are built by people that have the same education, meet at conferences and read each others work. In addition, the development of a complex soil process model is a time consuming affair that may involve many man years of work. These are all disadvantages of the Bayesian model averaging approach to soil process modelling. The advantage of Bayesian model averaging is that it can help choosing the optimal degree of model complexity, which is a persistent problem in soil process modelling that as yet has not been satisfactorily resolved.

As an alternative to Bayesian model averaging, we may fall back to models that represent structural errors as (additive) noise terms. This leads to stochastic models or so-called state-space models, for which a rich theory has been developed (e.g. Pugachev and Sinitsyn 2002). Perhaps these models are somewhat restrictive in the way that structural error is represented, but the practical advantages are evident. Also, stochasticity can be defined at the level of the underlying differential equations, which seems physically plausible. The use of stochastic models and associated data assimilation methods, such as ensemble Kalman filtering and particle filtering, is abundant in disciplines such as hydrology, meteorology and oceanography. However, in soil science their use has been very restricted. There is no reason to believe that model structural error is less important in soil science, and if we want to address it thoroughly we need to get more involved in these approaches.

Conclusion
Quantification of error in the inputs, parameters and structure of soil process models needs more attention because model outputs should be accompanied by accuracy measures and realistic assessment of these errors is indispensible for sound error propagation analysis. In this respect, soil science still lags behind compared to other disciplines within the earth and environmental sciences. Most published studies only focus on the propagation of errors in model input, but this is only one component of the total error. Also, input error assessment must benefit more from developments in the expert elicitation literature. Bayesian inverse modelling approaches for quantification of errors in parameters and model structure are useful too, but it is important that these make comparisons at the right support and include all error sources (errors in inputs, parameters, structure and in observations of model output), because otherwise error estimates for individual error sources will be flawed.

References
Ayyub BM (2001). Elicitation of Expert Opinions for Uncertainty and Risks. CRS Press, Florida.

Bishop TFA, Minasny B, McBratney AB (2006). Uncertainty analysis for soil-terrain models. *International*

Journal of Geographical Information Science **20**, 117-134.

Brown JD, Heuvelink GBM (2005). Assessing uncertainty propagation through physically based models of soil water flow and solute transport. In 'Encyclopedia of Hydrological Sciences'. (Ed MG Anderson), pp. 1181-1195. (Wiley: Chicester).

Brown JD, Heuvelink GBM (2007). The Data Uncertainty Engine (DUE): a software tool for assessing and simulating uncertain environmental variables. *Computers & Geosciences* **33**, 172-190.

Brus DJ, Bogaert P, Heuvelink GBM (2008), Bayesian Maximum Entropy prediction of soil categories using a traditional soil map as soft information. *European Journal of Soil Science* **59**, 166-177.

Castrignano A, Buttafuoco G, Canu A, Zucca C, Madrau S (2008). Modelling spatial uncertainty of soil erodibility factor using joint stochastic simulation. *Land Degradation & Development* **19**, 198-213.

Cooke RM, Goossens LHJ (2004). Expert judgement elicitation for risk assessments of critical infrastructures. *Journal of Risk Research* **7**, 643-656.

Dean S, Freer J, Beven K, Wade AJ, Butterfield D (2009). Uncertainty assessment of a process-based integrated catchment model of phosphorus. *Stochastic Environmental Research and Risk Assessment* **23**, 991-1010.

De Vries W, Kros J, Oenema O, De Klein J (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 PJ, Ribeiro Jr. PJ (2007). 'Model-based Geostatistics'. (Springer: New York).

Finke PA, Wladis D, Kros J, Pebesma EJ, Reinds GJ (1999). Quantification and simulation of errors in categorical data for uncertainty analysis of soil acidification modelling. *Geoderma* **93**, 177-194.

Hartman LW (2006). Bayesian modelling of spatial data using Markov random fields, with application to elemental composition of forest soil. *Mathematical Geology* **38**, 113-133.

Hengl T, Heuvelink GBM, Stein A (2004). A generic framework for spatial prediction of soil properties based on regression-kriging. *Geoderma* **120**, 75-93.

Heuvelink GBM (1998). Uncertainty analysis in environmental modelling under a change of spatial scale. *Nutrient Cycling in Agro-ecosystems* **50**, 255-264.

Heuvelink GBM, Van Den Berg F, Burgers SLGE, Tiktak A (2009). Uncertainty and stochastic sensitivity analysis of the GeoPEARL pesticide leaching model. *Geoderma* **155**, 186-192.

Raftery AE, Madigan D, Hoeting JA (1997). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association* **92**, 179-191.

Hyvonen R, Agren GI, Bosatta E (1998). Predicting long-term soil carbon storage from short-term information. *Soil Science Society of America Journal* **62**, 1000-1005.

Kaplan S (1992). Expert information versus expert opinions - another approach to the problem of eliciting combining using expert knowledge in probabilistic risk analysis. *Reliability Engineering & System Safety* **35**, 61-72.

Lesschen JP, Stoorvogel JJ, Smaling EMA, Heuvelink GBM, Veldkamp A (2007). A spatially explicit methodology to quantify soil nutrient balances and their uncertainties at the national level. *Nutrient Cycling in Agro-ecosystems* **78**, 111-131.

Pugachev VS, Sinitsyn IN (2002). 'Stochastic Systems: Theory and Applications'. (World Scientific Publishing: Singapore).

Refsgaard JC, Van Der Sluijs JP, Brown J, Van Der Keur P (2006). A framework for dealing with uncertainty due to model structure error. *Advances in Water Resources* **29**, 1586-1597.

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

Shang H (2009). Estimation of terrestrial greenhouse gas emissions in Europe. MS thesis, Wageningen University.

Webster R (2000). Is soil variation random? *Geoderma* **97**, 149-163.