Reinds, Gert Jan; Van Oijen, Marcel; Heuvelink, Gerard B. M.; Kros, Hans. 2008 Bayesian calibration of the VSD soil acidification model using European forest monitoring data. *Geoderma*, 146 (3-4). 475-488. doi:10.1016/j.geoderma.2008.06.022

Bayesian calibration of the VSD soil acidification model using European forest monitoring data

Gert Jan Reinds[1,2]

Marcel van Oijen[3]

Gerard B.M. Heuvelink[1]

Hans Kros[1]

[1] Alterra, Wageningen University and Research Centre (WUR), P.O. Box 47, NL-6700 AA Wageningen, Netherlands.

[2] Corresponding author. Gertjan.Reinds@wur.nl. Tel. +31 317 486508, Fax +31 317 419000

[3] CEH-Edinburgh, Bush Estate, Penicuik EH26 0QB, United Kingdom

---

[1] Alterra, Wageningen UR

Abstract

Over the past years, Bayesian calibration methods have been successfully applied to calibrate ecosystem models. Bayesian methods combine prior probability distributions of model parameters, based on assumptions about their magnitude and uncertainty, with estimates of the likelihood of the simulation results by comparison with observed values. Bayesian methods also quantify the uncertainty in the updated posterior parameters, which can be used to perform an analysis of model output uncertainty. In this paper, we apply Bayesian techniques to calibrate the VSD soil acidification model for 182 intensively monitored forest sites in Europe. Prior distributions for the model parameters were based on available literature. Since the available literature shows a strong dependence of some VSD parameters on, for example, soil texture, prior distributions were allowed to depend on soil group (i.e. group of soils with similar texture or C/N ratio). The likelihood was computed by comparing modelled soil solution concentrations with observed concentrations for the period 1996-2001. Markov Chain Monte Carlo (MCMC) was used to sample the posterior parameter space. Two calibration approaches were applied. In the singe-site calibration, the plots are calibrated separately to obtain plot-specific posterior distributions. In the multi-site approach priors are assumed constant in space for each soil group, and all plots are calibrated simultaneously yielding one posterior probability distribution for each soil group. Results from the single-site calibrations show that the model performs much better after calibration compared to a run with standard input parameters. Posterior distributions for H-Al equilibrium constants narrow down, thus decreasing parameter uncertainty. For base cation weathering of coarse texture soils the posterior distribution is shifted to larger values, indicating an initial underestimation of the weathering rate for these soils. Results for the parameters related to nitrogen modelling show that the nitrogen processes model formulations in VSD may have to be reconsidered as no evidence is found for a relationship between nitrogen immobilization and the C/N ratio of the soil, as assumed in VSD. The multi-site calibration also strongly decreases model error for most model output parameters, but model error is larger than the median model error from the singe-site calibration. Because the large number of plots calibrated at the same time provides very many observations, the Markov chain converges to a very narrow parameter space, leaving little room for posterior parameter uncertainty. For an uncertainty analysis with VSD on the European scale, this study provides promising results, but more work is needed to investigate how the results can be used on a European scale by looking at regional patterns in calibrated parameters from the site calibration or by calibrating for regions instead of all of Europe.

Keywords: Parameter Estimation, Markov Chain Monte Carlo, Soil Chemical Processes , Forest Soils

1. Introduction

Over the last decade, reductions of sulphur and, to a lesser extent, nitrogen emissions from industry, traffic and agriculture have led to a decrease in the deposition of acidifying compounds in Europe. Under the Convention of Long-range Transboundary Air Pollution (LRTAP) within the United Nations Economic Commission for Europe (UNECE), so-called critical loads for sulphur and nitrogen have been used as an indicator of the sensitivity of natural ecosystems to acidification and eutrophication. Critical loads provide the maximum deposition on an ecosystem that, on an infinite time-scale and according to current knowledge, will not lead to significant harmful effects (Nilsson and Grennfelt, 1988). Critical loads and are mostly computed using simple steady state mass balance models (Sverdrup and De Vries, 1994; UBA, 2004). As shown by for example Hettelingh and co-workers (Hettelingh et al., 2001; 2007), exceedances of critical loads for sulphur and nitrogen have been reduced and will even be further reduced in the future if current emission reduction agreements come into force. In areas where the critical load was exceeded in the past and where the present or future deposition is smaller than the critical load, ecosystems are expected to recover from acidification. Critical loads, however, do not provide information about the (speed of) recovery of an ecosystem. Therefore, dynamic acidification models such as SMART (De Vries et al., 1989), MAGIC (Cosby et al., 2001; Cosby et al., 1985) and SAFE (Warfvinge et al., 1993) have been used to evaluate the effects of future deposition scenarios on recovery. Ideally, the dynamic model should extend the critical load model so that dynamic processes are taken into account while all other process descriptions are compatible with those in the critical load model. To this end the Very Simple Dynamic model (VSD) has been developed (Posch et al., 2003) that extends the widely used Simple

Mass Balance (SMB) critical load model (Sverdrup and De Vries, 1994) with dynamic soil processes, including cation exchange.

Because a critical load model provides the maximum allowable deposition for an ecosystem on an infinitely long time scale, the model cannot be calibrated or validated directly. Dynamic models, on the other hand, simulate soil solution concentrations as a function of atmospheric inputs and soil processes, and can thus be calibrated and/or validated on plots where measurements are available. Until recently, calibration of most dynamic acidification models was performed by fitting the simulation results to (a series of) observations by 'trial and error' procedures: the model is re-run with different settings until the observation(s) are reproduced well. Sometimes a set of parameters is calibrated using various soil (solution) concentrations simultaneously (De Vries et al., 2003a), whereas in other studies only one parameter, such as the initial base saturation, is calibrated (Aherne et al., 1998; Belyazid et al., 2006). A more advanced model calibration was performed by Kros et al. (2002) who calibrated the SMART2 acidification model to soil observation data scaled up to 5×5 km blocks using the Gauss-Levenberg-Marquardt algorithm. The MAGIC model is normally calibrated sequentially, by first calibrating deposition of S compounds using stream water chemistry, then calibrating N concentrations by adjusting uptake functions and finally calibrating base saturation using an optimization procedure (Jenkins et al., 1997). Most of these studies do not take into account the uncertainty in the observations and model input parameters: only the run that provides the best estimate of the observations is accepted as the calibrated parameter set. It is clear, however, that observations of soil solution concentrations are uncertain, mainly due to spatial variability within a plot. Furthermore, several combinations of input parameters may give the same model result. For example, base cation concentration in the soil is influenced by base cation weathering, deposition and uptake. Adjusting growth rates, base cation contents in stem wood or weathering rates for a simulation could all lead to good fit of the model to the observations of base cation concentrations. Calibration methods that include these uncertainties and interactions are thus to be preferred over methods that use a simple fit through a (set of) observation(s) yielding one set of calibrated parameters without their uncertainty. In recent years, Bayesian calibration methods have been used for calibration of ecosystem models e.g. Larssen et al. (2006), Van Oijen et al. (2005), Vrugt et al.(2006). Larssen et al. (2006) performed a Bayesian calibration of the hydrogeochemical model MAGIC. Using simulated and observed stream water concentrations from a catchment in southern Norway, they quantify uncertainties and examine the propagation of these uncertainties in forecast simulations for 3 different deposition scenarios. Van Oijen et al. (2005) present a Bayesian calibration for process based forest models, illustrated by the calibration of the BASFOR forest growth model using data from a site in Sweden. De Vrugt et al. (2006) present the Shuffled Complex Evolution Metropolis (SCEM-UA) global optimization algorithm for environmental models. They provide examples of its effectiveness for 3 different case studies, among which an application for a watershed model.

Bayesian methods combine probability distributions of model parameters, based on prior assumptions about their magnitude and uncertainty, with estimates of the likelihood of the simulation results in view of the observed, uncertain values for model output variables. They use the combined information to quantify uncertainty in parameters and use the updated parameter uncertainty to perform an analysis of model output uncertainty. Bayesian calibration (BC) can be seen as a twofold extension of Maximum Likelihood estimation (ML). First, BC uses more information than just output data by including prior information about model parameters. Secondly, BC goes beyond identifying a single parameter vector with maximum probability to also estimate its uncertainty. In other words, BC provides a complete multi-variate probability distribution.

In this study, we applied Bayesian techniques to calibrate the VSD model for 182 intensively monitored forest sites in Europe for which data on atmospheric inputs, soil solution chemistry and soil properties are available (De Vries et al. (2003b). Two calibrations of the VSD model were carried out: (i) each plot was calibrated individually, called 'single-site calibration' hereafter and (ii) the entire set of plots was calibrated simultaneously, called 'multi-site calibration', to investigate effects of the grouping of data on calibrated parameters. In this way we investigate the applicability of the calibration method on a large set of plots and analyse whether the available measurements allow a site-specific calibration of input parameters such that the model performance after calibration is significantly improved compared to that using parameter settings based on regional information alone. Eventually, results of this study could be helpful for parameterization of the VSD model when applied on a European scale.

2 Materials and methods

2.1 Location of the plots

The data for the calibration of VSD were derived from the set of EU/UNECE Intensive Monitoring plots (De Vries et al., 2003b), using those 182 plots where sufficient data on soil solution, soil solid phase, forest growth and deposition were available. Sixty validation locations were selected by k-means clustering of the 182 locations into 6o clusters, using the Cartesian x- and y-coordinates of the 188 locations as classification variables (de Gruijter et al, 2006). The locations closest to the centroids of the clusters were selected as validation location. This selection procedure ensures that the validation locations cover the study area as good as possible. The remaining 122 plots were used to calibrate VSD.

Figure 1 shows the location of the calibration and validation plots, that mainly consist of pine- (43 plots), spruce- (80 plots), beech- (29 plots) and oak forest (21 plots). Most of the plots cover about 0.25 ha. Measurements from surveys on soil condition and forest growth were used to obtain variables such as soil cation exchange capacity (CEC), soil bulk density and the average stem growth at the sites. Measured soil solution concentrations below the root zone from the period 1996 to 2001 were used to assess the VSD model performance by comparing these measurements with concentrations simulated by the model.

Figure 1

2.2 The VSD model

As its name implies, the VSD model (Posch et al., 2003, Posch and Reinds, 2008) is a very simple dynamic model that simulates soil solution chemistry and soil nitrogen pools for natural or semi-natural ecosystems. The VSD model can be seen as the simplest extension of the simple mass balance (SMB) critical load model. The SMB model (De Vries and Posch, 2003a; Posch and De Vries, 1999) computes the maximum input of S and N to an ecosystem (i.e. critical load) that will not lead to harmful effects, using simple mass balance equations. VSD also consists of a set of mass balance equations, describing the soil input-output relationships of ions, and a set of equations describing the rate-limited and equilibrium soil processes. The soil solution chemistry in VSD depends solely on the net element input from the atmosphere (deposition minus net uptake minus net immobilisation) and the geochemical interactions in the soil (CO_2 equilibrium, weathering of carbonates and silicates, and cation exchange). Soil interactions are described by simple rate-limited reactions (e.g. nutrient uptake and weathering), first order processes (denitrification) and by equilibrium reactions (e.g. cation exchange). VSD models the exchange of Al, H and Ca+Mg+K with the Gaines-Thomas or Gapon equations. Solute transport is described by assuming complete mixing of the element input within one homogeneous soil compartment with a constant density and a fixed depth. VSD is a single layer soil model that neglects vertical heterogeneity. It predicts the concentration of the soil water leaving this layer (mostly the root zone). Validation of the model should thus be based on measurements from soil solution just below the root zone. The annual water flux percolating from this layer is taken equal to the annual precipitation excess. The model strongly resembles the SMART model (De Vries et al., 1989) but leaves out some of the processes modelled by SMART such as aluminium mass balance and the soil solution chemistry in carbonate rich soils. The time-step of simulations is one year.

2.3 Model input data

Input to VSD consists of a set of 24 parameters, listed in Table 1. We have chosen to calibrate only model process parameters such as equilibrium constants and denitrification and immobilization fractions. These parameters are (highly) uncertain because estimates are often based on small data sets and the most import VSD output parameters (pH, [Al], [Bc] and base saturation) are sensitive to these process parameters. Furthermore, these parameters cannot be measured directly and were therefore selected for calibration. Parameters for which measurements were available at the plot were not calibrated: the measurement was assumed to be the best estimate at the plot, even though it can be uncertain due to within-plot spatial variability and measurement error. Calibrating such inputs may improve model performance but results will be hard to use in any upscaling as the calibration results will probably be very site specific.

Table 1

2.3.1 Parameters not subject to calibration

The thickness of the root zone was set to 50 cm, except for a few plots where evidence exists that the soil is very shallow. Bulk density was obtained from plot data unless it was not measured, in which case it was computed using a transfer function between bulk density and clay- and organic matter content (Reinds et al., 2001). The hydrological characteristics (soil water content and precipitation surplus) were derived from simulations with the hydrological model WATBAL (Starr, 1999). The validity of the hydrological model was tested using the chloride budgets at the plots and turned out to be quite satisfactory for the vast majority of plots (De Vries et al., 2003a).

Cation exchange capacity, organic anion concentrations and all deposition fluxes were obtained from measurements at the plots (De Vries et al., 2003a). Total deposition was computed from measured bulk and throughfall data, using an adapted version of the canopy budget model (De Vries et al., 2001; Draaijers and Erisman, 1995; Ulrich and Pankrath, 1983). Deposition of SO_x and total N outside the measurement period were computed by applying the trend in deposition modelled by EMEP (2001) for the 50 × 50 km EMEP grid cell that the plot is located in to the measured data; historical trends before 1960 were obtained from Schöpp et al. (2003). Base cation deposition was assumed constant in time and was set to the average measured value from the measurement period as neither historical nor future projections of base cation depositions were available.

C/N ratio and carbon pool at the beginning of the simulations (in the year 1880) were back-calculated from the measured C/N ratio and carbon pool and historical N inputs.

2.3.2 Calibration parameters

As the 182 Intensive Monitoring plots cover a wide range of soil and forest types, it is unlikely that one prior probability distribution function (pdf) for each parameter would be sufficient to arrive at a successful calibration. It is known that e.g. weathering rates and exchange constants vary strongly with soil texture (UBA,2004, De Vries and Posch, 2003b). Since we want to use the best available information, prior distributions were defined for each *soil group*. Such a soil group consists of soils with the same soil texture and C/N ratio in organic matter. For each plot, a prior distribution for the uncertain parameter was defined appropriate for the local soil type. For example, plots in the soil group consisting of sandy soils receive only values selected from the distributions of *lgKAlBc* and *BCwe* that are valid for sandy soils. Even if the priors are identically defined for all soil groups, the calibration can still yield different posterior distributions for soil groups. An overview of the prior distributions is provided in Table 2. It lists for each calibration parameter its VSD code, a description, the distribution assumed (uniform or (truncated) normal), the mean, the standard deviation (*Stddev*), the minimum (*Min*, if any) and maximum value (*Max*, if any).

Table 2

Weathering rates were calibrated for six soil groups (i.e. weathering rate classes) that consist of a combination of soil texture and parent material. Initial distributions are obtained from De Vries et al., (1994) and Reinds and Posch (2001). Exchange constants, the H-Al equilibrium constant and exponent were defined as a function of the soil texture class. Initial values for exchange constants are obtained from De Vries and Posch (2003b) and are based on an analysis of hundreds of measurements from Dutch forest soils. Exchange constants are not independent, the covariance of the logarithms of *KAlBc* and *KHBc* for each texture class was obtained from data provided by De Vries and Posch (2003b) and inserted in the initial covariance matrix so that the candidate values for these parameters are properly correlated. The initial distributions of the log of the equilibrium constant *KAlox* was defined using the standard value of 9 as the average for all texture classes with a standard deviation of 2. Uptake of nutrients (Ca, Mg, K and N) was computed as forest growth rate multiplied with nutrient content (Jacobsen et al., 2002). Principally these contents are uncertain as well but the measurements do to distinguish between the weathering and uptake of base cations. We therefore decided to calibrate only base cation weathering. For nitrogen, the various sinks (uptake, denitrification, immobilization) cannot be distinguished either. Nevertheless, we decided to calibrate both the denitrification fraction assuming a relatively small uncertainty and two parameters in an immobilization function. Denitrification was calibrated for two different drainage classes: well-drained and moderately well-drained. Prior estimates of denitrification fraction were obtained from Reinds et al. (2001). Poorly or very poorly drained soils were not included in the set of 182 calibration plots as no hydrological data were available. N immobilization can be substantial in the current environmental situation, (De Vries et al., 2001; Gundersen et al., 1998) and was estimated as a function of the average N deposition for 1960-1990 according to:

$$N_{imm} = a + b \cdot N_{dep} \qquad (1)$$

where *a* was set to 1 kg N as an estimate of the long-term immobilization even at very low N input (UBA, 2004), whereas the parameter *b* was calibrated. Parameter *b* thus expresses the proportion of the N input that is immobilized additional to the long-term constant immobilization. Since some evidence exists that there is a (weak) relationship between C/N ratio in organic matter and immobilization rates of N (Gundersen et al., 2006; Gundersen et al., 1998), three prior distributions for *b* were defined: one for soils with low C/N ratio (< 25), one for soils with intermediate C/N ratios and one for soils with high C/N ratio (> 35). For each class a prior uniform distribution between 0 and 1 was assumed as there is also evidence that within C/N ratios classes, immobilization fractions can strongly vary (De Vries et al., 2001). This is the reason that for the calibration the original VSD formulation of N immobilization as a function of C/N ratio was replaced by Equation 1.

If no literature data were available, expert judgment was used to define the prior distributions of some parameters. Prior distributions for these parameters were taken such that the range was wide enough to cover all possible values. This limitation is to a large extent compensated by the calibration procedure that provides posterior distributions that are strongly determined by the likelihood and only partly by the prior distributions.

2.4 Observations

To assess the likelihood of the simulation results, comparisons were made between simulated and measured soil solution concentrations below the root zone. Measured concentrations of H, Ca + Mg + K, and NO_3 were used. Free Al was recomputed externally from the measurements of total Al, by modelling complexation of Al with organic anions, using measured DOC values and dissociation and Al complexation constants provided by Santore et al. (1995). Measurements were aggregated to average annual concentrations, yielding about 5-6 values per plot (1996-2001). Furthermore, a single measurement of the base saturation of the soil solid phase was used. The uncertainty in the measured soil solution concentrations cannot be obtained from the data directly as the samples collected from various samplers within the plot were pooled into a mixed sample before analysis. It is likely that the uncertainty in the measured concentrations is mostly determined by spatial variability within the plot. A study at some monitoring plots in the Netherlands showed that this variability varies between 20 and 60% depending on depth and ion (De Vries et al., 1999). An uncertainty of 30% was used as an uncertainty estimate for all plots in Europe as this was about the uncertainty for the major ions at 40-60 cm depth in the Dutch plots. For very low concentrations that often are observed for Al and NO_3, 30% uncertainty is probably an underestimate. For measurements equal to or lower than the detection limit of the most widely used analysis equipment (Inductive Coupled Plasma spectrometry , ICP) the uncertainty was therefore set at two times the detection limit of the ICP.

2.5 Calibration method

2.5.1 Bayes' theorem

The VSD model was calibrated using Bayesian calibration. Our implementation of Bayesian calibration strongly resembles the method applied by Van Oijen et al. (2005). In the Bayesian approach the prediction of uncertain parameters is taken to be conditional on data, and their conditional probability is given by:

$$p(\theta \mid D) = c \cdot p(D \mid \theta) \cdot p(\theta) \quad (2)$$

Where $p(\theta \mid D)$ is the posterior (conditional) probability of the parameter (vector) θ given the data (observations) D, $p(D \mid \theta)$ is the likelihood-function for θ and $p(\theta)$ is the prior distribution of θ. The value of c ($=1/p(D)$) is fixed and usually needs not be computed explicitly (Van Oijen et al., 2005).

The prior distributions, $p(\theta)$, for a number of VSD input parameters (listed in Table 1) must be defined based on best available knowledge. Next, the posterior distribution of input parameters given data on model output, $p(\theta \mid D)$, is computed based on comparison of the outcome of VSD with the set of measurements of soil (solution) chemistry at the Intensive Monitoring plots, including their uncertainty Equation (2) shows that the posterior probability for θ increases with an increased prior probability and an increased likelihood, i.e. when the selected set of parameter values have larger a priori probability and when the model is able to reproduce the measurements. The prior probability is directly computed from the probability of a candidate point of θ; for candidate points from normal distributions close to the mean the probability density will be large, for points in the 'tail' of the distribution the probability density will be small. The likelihood $p(D \mid \theta)$, is computed assuming measurement errors are Gaussian and uncorrelated (Van Oijen et al., 2005):

$$\log p(D \mid \theta) = \sum_{i=1}^{n} \left(-0.5\left(\frac{O_i - S_i}{M_i} \right)^2 - 0.5\log(2\pi) - \log M_i \right).(3)$$

Where *S* is the simulated and *O* the observed soil solution concentrations, *n* is the number of observations and M_i. is the standard deviation of the measurements.

When calibrating VSD we assume the model is correct and thus ignore model error. The final estimates of uncertainty are thus only estimates of parameter uncertainty. After site specific calibration, the remaining error, here expressed in the normalized root mean square error (NRMSE), could be an indication of model error but also includes measurement error {Larssen, 2006 #2111}.

2.5.2 Markov Chain Monte Carlo

In practice, Equation (2) has to be evaluated numerically, by using e.g. a simulation approach. In such cases, Bayesian techniques rely on carrying out of a large number of simulations, often in the form of a Markov Chain Monte Carlo (MCMC) approach. To calculate the posterior $p(\theta \mid D)$, we used a simple MCMC algorithm, known as the Metropolis-Hastings random walk (Robert and Casella, 1999). In this method, the multi-dimensional parameter space (where the number of dimensions, N_p, equals the number of uncertain model parameters) is explored by randomly stepping through this space and running the model for each visited point. The combination of number of steps and step size should be chosen such that the parameter space is adequately sampled during the MCMC. In the first step, a candidate value for each parameter is chosen. The model is run for N plots (with N =1 for the single site-calibration or N = a set of plots for the multi site-calibration) and the likelihood is determined. We started the Markov chain by selecting for each parameter the midpoint of the prior distribution. To test the sensitivity of the method to the starting point of the Markov chain, we have also calibrated the model using a starting point that consists of parameter values randomly chosen within the 95% confidence interval of the prior distribution. For subsequent runs a new candidate point (consisting of N_p parameter values) is computed by moving randomly away from the current point in the multi-dimensional parameter space. If, for the new point, the product of prior probability and likelihood exceeds that of the current point, the new point is accepted. On the other hand, if the ratio of new and current products of the prior and likelihood is between 0 and 1, the new point can still be accepted, but with a probability that is equal to that ratio. The algorithm is inherently stochastic since candidate points that perform (somewhat) worse than previous points can still be accepted. If the candidate point is not accepted, the previous candidate point is duplicated in the chain of accepted points. Eventually, this procedure yields a chain of points in the N_p-dimensional parameter space. To account for a 'burn-in' of the chain, i.e. to remove the effect of the choice of the starting point, the first 10% of the runs are removed from the chain. The remaining chain contains all accepted (or duplicated) parameter values. From this chain we derived the posterior distribution of each parameter. Furthermore, correlation and covariance matrices were computed from the chain using the standard statistical routine CORVC from the IMSL statistical library (Visual Numerics, 1997). By running multiple chains with different starting points, we have verified convergence of the calibration.

2.5.3 Practical implementation

In the single-site calibration every plot was calibrated separately so that the posterior distributions are computed on a plot by plot basis. In the multi-site calibration all plots within a soil group are calibrated simultaneously, thus obtaining posterior distributions for the entire soil group. In the single site calibration we assume that soil parameters vary in space, i.e. even within a soil group parameters values can be different for different sites. For each plot the calibration yields a plot-specific estimate of the posterior distribution of parameters. In the multi-site calibration we assume that parameters are constant in space. For each group of plots with uniform soil characteristics, one posterior distribution will be computed. Calibrating for each plot separately can provide insight in the variation and patterns in posterior distributions over Europe. Such patterns can provide means to extrapolate the results to an European scale. Calibrating all plots simultaneously yields posterior distributions that can be directly used on an European scale but only with the assumption that for a given soil group parameters are constant in space.

3 Results and discussion

3.1 Results for single-site calibrations

In the following, the success of the BC application to the VSD model is evaluated by checking whether (a) the goodness-of-fit of VSD has increased after calibration and (b) the posterior distribution is narrower than the prior, indicating reduced parameter uncertainty (c) running the calibrated model on the 60 validation plots. The calibration used a chain length of 50000 and the mean step length, i.e. the standard deviation of the proposal distribution, is set at 3% of the width of the prior.

3.1.1 Likelihood

To quantify the gain in model performance, a comparison was made between the goodness-of-fit using a run at every calibration plot with parameters set to the mean of their prior distributions and the average goodness of fit over all accepted runs in the Markov Chain (that form the posterior distribution).. Goodness-of-fit is expressed here as the Normalized

Root Mean Square Error (NRMSE), defined as the root of the mean squared difference between measurements and simulations, divided by the mean of the measurements. Figure 2 shows the cumulative frequency distributions of NRMSE for pH, Al, NO_3 and base saturation (EBc) based on all 122 calibrated plots before and after calibration.

Figure 2

Figure 2 shows, as expected, that for all VSD output parameters the average NRMSE computed from all excepted runs from the MCMC is much lower that the NRMSE of the run using the mean of the priors. Especially for pH and NO_3, the model prediction error decreases strongly: The median error after calibration is less than half the error prior to calibration. For Al and NO_3 errors for about 10-20% of the plots are zero as both the measured as well as the simulated concentrations are zero. Figure 2 also shows that for a limited number of plots (about 10%), model prediction errors in especially NO_3 and base saturation remain large despite calibration. For EBc these are probably the plots where the combination of observed EBc and observed pH cannot be reproduced by the VSD model. Since there is only one observation on EBc and mostly 5-6 observations for pH, it is likely that the calibration procedure will move towards a good fit on pH rather than on EBc. Large residual NRMSE can be due to model error, but could also be caused by measurement error, indeicated by e.g. unlikely combinations of measured pH and base saturation.

Figure 3

Figure 3 shows the spatial pattern in the NRMSE before calibration minus the NRMSE after calibration, for pH and NO_3. The larger the value, the larger the gain in model performance. Figure 3a shows that for most plots a substantial gain in model performance is achieved for pH, but that in parts of Germany, the UK and the plots in southern Europe the NRMSE does not strongly decrease. Part of the explanation is given by Figure 3b that shows the gain in NRMSE for NO_3. This figure shows a strong increase in model performance for most of the plots for which pH did not strongly improve. Obviously, the calibration at these plots leads to much better nitrate concentrations but the parameterisation improves the simulation of total acidity to a lesser extent. Figure 3b also shows that for many plots in Southern Scandinavia, Eastern parts of Germany and southern Europe, the calibrated model performs much better for NO_3 with calibrated parameters than with the default mean parameter values.

3.1.2 Posterior distributions of model parameters

The main result of the Bayesian calibration procedure is the joint posterior distribution for the model parameters. The posterior distribution also contains correlations between parameters. However, it is difficult to visualize a multi-dimensional distribution for many sites, so in the following we shall focus on the marginal distributions for individual parameters. If the measurements are conclusive enough, broad prior distributions will narrow down and thus demonstrate a reduction of parameter uncertainty. Posterior distributions may also lie in the high or low parts of the prior distributions, indicating that the data forced a small probability to parameter vectors that were considered plausible before. If the a priori assumed relationship between soil characteristics and parameter values is confirmed by the data, one may expect that the mean of the posterior probability distribution functions will not differ much from the mean of the prior pdf, the but uncertainty could be significantly reduced.

Figure 4 shows the prior and combined posterior distribution of the base cation weathering rates for the low (1), intermediate (3) and high (6) weathering rate classes. For the other classes only a few plots (< 15) were available for so the posterior distribution may be more strongly determined by the prior distribution. It shows, as expected, that weathering rates increase with increasing weathering rate class. It also shows that the prior for weathering rate class 1 was underestimated: after calibration the distribution is at much larger values than the prior, with the median weathering rate being about 2.2 times as large as initially assumed. Posterior weathering rates for class 6 (rich clay soils) tend to be skewed towards small values, whereas the posterior for weathering rate class 3 (clay soils) indicates that the observations do not significantly change the prior estimate, although the posterior distribution function is more uniform than the prior normal distribution.

Figure 4

In Figure 5, the distribution functions are displayed for the *lgKAlox* parameter, i.e. the logarithm of the constant defining the equilibrium between H and Al concentrations. It clearly shows that the data strongly narrow the prior distributions; uncertainties in the posterior parameter distributions are much smaller than the a priori assumed Furthermore, the posterior distributions show that there is little difference between the three texture classes. As expected *lgKAlox* values decrease somewhat with increasing clay content.

Figure 5

Figure 6

Figure 6 shows the prior and posterior distributions for the exchange constants *KAlBc* and *KHBc* for sand and clay. It shows that the posterior distributions of *KAlBc* and *KHBc* for sand hardly differ from their priors, indicating that the data cannot improve the accuracy of the initial estimates of the exchange constants. For clay soils, the posterior distributions differ somewhat from the prior. For *KAlBc* there is a shift towards somewhat smaller values, for *KHBc* there is a shift to somewhat higher values.

In Figure 7 the prior and posterior distribution of the fraction of nitrogen after N uptake and denitrification that is immobilized (parameter *b* in equation 1) is shown for the two C/N ratio classes, low and medium. For the C/N class 'high' only a few plots were available. The graphs clearly show that the calibration procedure confirms findings based on other datasets (e.g. Gundersen et al., 2006), namely that the immobilisation fraction varies widely even within different C/N ratio classes but that low retention occurs mainly at sites with at low C/N ratios. A majority of the sites retain most of the incoming nitrogen (parameter values close to one). Figure 7 indicates some differences in nitrogen retention between plots with low and medium C/N ratios: as expected low N retention fractions occur more frequent for low C/N ratios (indicating nitrogen saturation) than for high C/N ratios. In the majority of plots high C/N ratios more than 80% of the incoming N is retained.

Figure 7

3.1.3 Influence of the starting point of the Markov chain

In the standard run, the Markov chain is started at the mean of each prior parameter distribution. Ideally the starting point should have no influence on the posterior distributions. Different posterior distributions may be obtained at different starting points only if e.g. the parameter space is not fully explored . Figure 8 shows the posterior distributions for *lgKAlox* for sand and clay and for the *BCwe* for soils with texture class 1 for different starting points. The figure shows that the posteriors are very similar, indicating that the Markov chain converged to the same part of the parameter space, independent of its starting point.

Figure 8

3.1.4 Validation

For validation, the NRMSE for the VSD output parameters was computed for each of the 60 validation plots. One run was made using parameter values equal to the mean of the priors (uncalibrated run) whereas in the second run input parameters were set to the median of the posterior parameter distributions obtained from the calibration. For each validation plot, the proper posterior distribution was selected based on the soil texture, parent material and topsoil C/N ratio of the plot. The cumulative frequency distributions of the NRMSE for pH, Al, NO_3 and base saturation (EBc) before and after calibration show that NRMSE decreases for pH, NO_3 and to a lesser extent for EBc if the model is run at the validation plots using the results from the calibration (Figure 9). The error in the simulated aluminium concentration is about equal in both runs. For pH, NO_3 and EBc_c the NRMSE decreases for about 65-75% of the plots when using calibrated parameters, for Al the use of calibrated parameters leads to more accurate simulations for only half of the plots.

Comparing figures 9 and 2 shows that the gain in NRMSE for the validation plots is lower than for the calibration plots. This is according to expectations. For the VSD application to the validation plots the median value from the posterior distributions obtained from the entire set of calibration plots was used, whereas the calibration plots were calibrated individually thus obtaining the best fit for each plot. Nevertheless, the calibration has shown to be successful as an application of the calibrated model to the validation plots yields a (much) better fit to the observed soil solution concentrations (except for aluminium) than an application with parameter values at the mean of the prior distribution.

3.2 Multi-site calibration

3.2.1 Calibration

In the multi-site calibration, for every soil group (i.e. a set of sites with the same soil characteristics), the vector of model parameters is selected from a single prior multivariate distribution and the overall prior probability is computed. We thus assume that the parameter values per soil group are constant. For example, all plots with poor sandy soils are assigned the same *lgKAlox* value selected from the prior for poor sandy soils. Then, simulations are made for all plots and the overall likelihood, being the product of all the likelihoods over all plots is computed. It is thus assumed that observation and model errors are independent. Next a new candidate point is selected by taking a step in the multi-dimensional parameter space following the same procedure as for the single-site calibration. This procedure thus calibrates all sites simultaneously and uses MCMC to obtain values of the N_p model parameters that give a good fit for all plots within a soil group simultaneously. In this case the Markov chain converges to a very narrow posterior distribution due to the large number of observations the uncertainty about the model parameters becomes very small (Figure 10). The figure shows the prior and very narrow posterior distributions of *lgKAlox* for sand and clay and *BCwe* for weathering rate class 3. The vast amount of observations (120 plots and 3-5 years of soil solution measurements) causes this very low parameter uncertainty. Because we assume VSD to be correct, and thus ignore model error, the uncertainty shown here is parameter uncertainty only, given VSD as is.

Figure 10

Comparing the NRMSE from the multi-site calibration with the single-site calibration shows that the NRMSE at the calibration plots from the multi-site calibration exceeds the NRMSE from the single-site calibration for all parameters. Largest differences occur for pH where the median error over all 122 plots from the multi-site calibration is is about twice that of the single-site calibration. For NO_3, Al and EBc differences are much smaller (20-25% in the median NRMSE). This shows that assuming constant parameter values for a soil group leads to a larger model prediction error than calibrating the parameters for each plot separately. Although parameter uncertainty after calibration is low, the model uncertainty is large, i.e. VSD parameterized with spatially constant parameters has a larger model error than VSD with local parameter settings.

3.2.2 Validation

The multi-site calibration was verified by running VSD for the 60 validation plots with parameter values sampled from the posterior distributions obtained from the multi-site calibration and comparing the NRMSE with the NRMSE obtained with parameters set to the mean from their prior distributions (see also section 3.1.4).. Figure 11 shows that for pH and NO_3 the calibration leads to (much) smaller simulation errors, but for aluminium no improvement is achieved. The same was observed in the validation of the singe site calibration. For EBc the NRMSE decreases for part of the plots but increases for other plots, but the the largest errors disappear due to the calibration. Comparing Figure 11 with Figure 11 shows that the NRMSE at the validation plots using the multi-site calibration exceeds the NRMSE from the singe-site calibration for EBc. For pH, Al the errors are about equal. For a number of plots the error in NO_3 when using the results from the multi site calibration is substantially lower than the error from the single-site calibration. These are most likely the plots in the validation set were despite a high N input, all N is retained. In the multi site calibration the median f_{nim} is close to one and thus almost all incoming N is retained, but in the singe site calibration, also low f_{nim} values are present leading to a lower median f_{nim}. Using this lower value leads to large errors for some of the validation plots. Furthermore, because the NRMSE is defined as the root of the mean squared difference between measurements and simulations divided by the mean of the measurements, NRMSE can become very high if simulated nitrate concentrations exceed observed concentrations and the observed nitrate concentrations are very low.

To check that the Markov chain in the multi-site calibration converges to the area with the combination of highest prior probability and highest likelihood, the Markov chain was run (a) with another, random, starting point and (b) by crudely assuming that all parameters have a uniform distribution (with lower and upper bounds set at the 5 and 95 percentile of the normal distribution, respectively). In case (b) we eliminated the effect of the form of the prior probability and gave the Markov chain more liberty in exploring parameter space. Results from the run with the uniform distribution and the run with a random starting point of the Markov chain show also a very low acceptance (about 0.3-0.5%) and the posterior distributions only slightly differ from the ones of the standard run. This is illustrated in Figure 12, which shows the cumulative frequency diagrams for the three runs for *lgKAlox* for sand and clay and *BCwe* for weathering rate class 3. In all cases the posterior distributions are very narrow and very similar. This shows that the Markov chain of the multi site calibration converges to the same posterior pdf, irrespective of the starting point or assumptions about the distribution type of the prior, proving the robustness of the procedure.

Figure 12

4 Discussion and conclusions

The Bayesian calibration technique using MCMC applied in this study has proven a successful method to calibrate the VSD model at a large set of locations. By applying a single-site calibration and sampling from a set of prior distributions defined as a function of site characteristics such as soil texture, the fit of the model on the soil solution measurements strongly improved compared to model runs with standard values for the input parameters. At the same time the parameter uncertainty was quantified, and the posterior distributions can be used for uncertainty assessments with VSD at the plots. This gives a clear advantage over simple calibration techniques used for VSD previously (De Vries et al., 2003a). An application of the calibrated model (using the median parameter values from the posterior distributions obtained from the calibration) to the 60 validation plots showed that a (much) better fit to the observed soil solution concentrations is obtained than an application with parameter values at the mean of the prior distribution. Only for the aluminium concentration no improvement in fit was achieved.

Posterior distributions of base cation weathering rates show that for plots in the sandy texture class, often relatively high weathering rates compared to literature data, such as De Vries et al.(1994), are required to simulate the observed base cation concentrations. This can be explained by the fact that the prior base cation weathering for class 1 is mainly based on data from pure sandy soils. Texture class 1 though, is a texture class encompassing pure sandy soils as well as soils with clay contents up to 8 per cent. It seems that a substantial number of plots with weathering class 1 have a relatively rich mineralogy and/or high clay content, leading to higher weathering rates than initially assumed.

Comparison of posterior distributions with the prior distributions showed that the data are conclusive for the *lgKAlox* parameter, because the posterior distribution is much narrower than the prior. For the exchange constants, the calibration does not lead to narrower distributions, but some shifts do occur.

The simulation of the nitrogen cycle in VSD is very simple and may sometimes be too simple to adequately model the nitrogen fluxes observed in the field (Evans et al., 2006). In VSD, nitrogen immobilization is modelled as a function of C/N ratio in the soil. In this study we have modelled immobilization using a linear function, where immobilization is modelled irrespective of C/N ratios as a fraction of the remaining nitrogen after uptake and denitrification. By defining three identical prior distributions for different C/N ratios, we could judge by the posterior whether the data yielded a clear difference between the three classes. The posterior distribution of the immobilized fraction of nitrogen for soils with a low C/N ratio is similar to that of soils with a higher C/N ratio although low N retention fractions are clearly more frequent at low C/N ratios than at medium C/N ratios. Also other studies have shown that the fraction of N immobilized can strongly vary within C/N classes (Gundersen et al., 1998). Consequently, the way in which N removal is currently modelled in the standard VSD model needs to be reconsidered. However, the fact that many processes influence the N balance (uptake, denitrification, immobilization, leaching), makes it difficult to adequately model and calibrate the associated parameters.

The multi-site calibration in which all sites were modelled simultaneously aiming at the highest combination of prior probability and likelihood over all plots within a soil group also strongly improved the goodness of fit of the model results compared to a run with standard values for model parameters for the 60 validation plots. The median *NRMSE* for pH after calibration was about 2 times as high as the median *NRMSE* from the validation of the single-site calibration, but for Al, NO_3 and EBc a difference in NRMSE of only 20-25% was observed. Due to the large amount of data and the assumption that VSD is 'correct', the multi-site calibration leads to a very small parameter uncertainty with very narrow posterior distributions. NRMSE increases for the validation plots compared to the single-site calibration because we assume that parameters are constant in space for a given soil group. This assumption limits the flexibility of the model to adapt to the local situation. A compromise could be to perform multi-site calibrations within environmental zones (Metzger, 2005) instead of entire Europe. This could reduce model error and still allow the use of the calibration results in European-wide applications of VSD.

Further study is required to make a full in-depth analysis of the multi-site calibration to confirm that the narrow posterior distributions are caused by the large amount of data only that makes parameter uncertainty small, and that the result is not strongly influenced by (sets of) plots with certain (deviating) characteristics or specific measurements. We have noticed in the site by site calibration that a few measurements with small uncertainty can strongly limit the acceptance in the MCMC procedure. Such small uncertainty in data must be realistic to justify small uncertainty in the posterior distributions. Therefore we also feel that more data on the (spatial) variation of the soil solution measurements are needed to improve the uncertainty estimates as the current assumption of a fixed 30% standard deviation was a crude assumption based on a single data set.

In the multi-site calibration the parameter uncertainty becomes very low because we do not include a model error term in the computation of the likelihood. To assess the uncertainty in model structure, a Bayesian model comparison of VSD could be used in which VSD is compared with e.g. more detailed models . Another way to assess the error due to model structure would be to run VSD at plots with completely independent validation data using the posterior distributions of the calibrated parameters and then compare the model error with the model error at the calibration plots (Heuvelink and Pebesma, 1999; Kros et al., 2002), but this requires a.o. a correct quantification of the measurement error of the observations.

Table 1. VSD parameters

Parameter	Description	Calibration	Parameter	Description	Calibration
thick	Thickness of the rootzone	no	f_de	Denitrification fraction	Yes
bulkdens	Bulk density	no	percol	Precipitation surplus	No
Theta	Soil water content	no	Bcwe	Base cation weathering	Yes
pCO2fac	Partial pressure of CO_2 in soil	no	ctNst	N content in stems for N uptake	No
CEC	Cation exchange content	no	ctCast	Ca content in stems for Ca uptake	No
lgKAlox	Equilibrium constant H-Al	yes	Ca_dep	Ca deposition	No
lgKAlBc	Exchange constant Al-BC	yes	SO2_dep	SO_2 deposition	No
lgKHBc	Exchange constant H-BC	yes	NOx_dep	NOx deposition	No
Nim_acc	N immobilization	yes	Mg_dep	Mg deposition	No
Cpool_0	Initial C pool	no	K_dep	K deposition	No
CNrat_0	Intial CN ratio	no	Na_dep	Na deposition	No
cRCOO	Organic anion concentration	no	Cl_dep	Cl deposition	No

Table 2. Prior distributions of VSD calibration parameters

Parameter	Description	Distribution	Mean	Stddev	Min	Max
lgKAlox_sand	Log10 of Equilibrium constant H-Al sandy soils $(mol.l^{-1})^{-2}$	normal	8	1		
lgKAlox_clay	Log10 of Equilibrium constant H-Al clay soils	normal	8	1		
lgKAlox_hclay	Log10 of Equilibrium constant H-Al heavy clay soils	normal	8	1		
lgKAlBc_sand	Log10 of Exchange constant Al-BC sandy soils	normal	0.503	0.6		
lgKAlBc_clay	Log10 of Exchange constant Al-BC clay soils	normal	-0.6	0.7		
lgKAlBc_hclay	Log10 of Exchange constant Al-BC heavy clay soils	normal	-0.6	0.7		
lgKHBc_sand	Log10 of Exchange constant H-BC sandy soils	normal	3.3	0.35		
lgKHBc_clay	Log10 of Exchange constant H-BC clay soils	normal	3.6	0.5		
lgKHBc_hclay	Log10 of Exchange constant H-BC heavy clay soils	normal	3.6	0.5		
fde_mod	Denitrification fraction moderately-well drained soils (-)	truncated normal	0.4	0.1	0	1
fde_well	Denitrification fraction well drained soils	truncated normal	0.2	0.075	0	1
BCwe_iwr1	BC weathering acid sandy soils $(eq.m^{-2}.yr^{-1})$	truncated normal	0.025	0.025	0	
BCwe_iwr2	BC weathering intermediate/basic sandy soils	truncated normal	0.075	0.0375	0	
BCwe_iwr3	BC weathering acid loamy/clayey soils	truncated normal	0.125	0.0625	0	
BCwe_iwr4	BC weathering intermediate loamy/clayey soils	truncated normal	0.175	0.0875	0	
BCwe_iwr5	BC weathering basic loamy/clayey soils	truncated normal	0.225	0.1125	0	
BCwe_iwr6	BC weathering heavy clay soils	truncated normal	0.275	0.1375	0	
bNim_cnh	Immobilisation fraction soils with high C/N (-)	uniform			0	1
bNim_cnm	Immobilisation fraction soils with intermediate C/N	uniform			0	1
bNim_cnl	Immobilisation fraction soils with low C/N	uniform			0	1

Table 3. Average simulation errors before and after calibration

	Uncalibrated NRMSE	Calibrated NRMSE
pH	0.16471	0.10552
[Al] (eq/m^3)	5.17217	0.98769
[NO$_3$] (eq/m^3)	10.90941	0.23611
EBc	0.71144	0.82169

Figure captions

Figure 1. Location of the plots used for the calibration

Figure 2. Cumulative frequency distributions of NRMSE for pH (a),Al (b), NO3(c) and EBc(d) before (dashed) and after calibration (solid).

Figure 3. Gain in NRMSE for pH (a) and NO3(d) after calibration.

Figure 4 Prior (light grey) and posterior (dark grey) distribution functions for weathering rates classes 1 (A),3 (B) and 6 (C).

Figure 5. Prior (light grey) and posterior (dark grey) distributions of lgKAlox for sand (A), clay (B) and heavy clay (C)

Figure 6 Prior (light grey) and posterior (dark grey) distributions for the exchange constants KAlBc for sand (A) and for clay (B) and KHBc for sand (C) and for clay (D).

Figure 7. Prior (light grey) and posterior (dark grey) distributions of the N immobilization fraction b for soils with a low C/N ratio (A) and intermediate C/N ratio (B)

Figure 8. Posterior distributions of lgKAlox for sand (A) and clay (B) and of weathering rate for texture class 1 (C) as a function of the starting point of the Markov chain (dark grey = mean, light grey = random).

Figure 9. Prior (light grey) and posterior (dark grey) distributions for the parameter lgKAlox for sand (A) and clay(B) and for base cation weathering for weathering rate class 3 (C); multi site calibration.

Figure 10. Cumulative frequency diagrams for the three runs for lgKAlox for sand (A) and clay (B) and BCwe for weathering rate class 3(C) for the standard run (solid), run with uniform prior (dot) and run with random start (dash)

References

Aherne, J., Sverdrup, H., Farrell, E.P. and Cummins, T., 1998. Application of the SAFE model to a Norway spruce stand at Ballyhooly, Ireland. Forest Ecology and Management, 101(1-3): 331-338.

Belyazid, S., Westling, O. and Sverdrup, H., 2006. Modelling changes in forest soil chemistry at 16 Swedish coniferous forest sites following deposition reduction. Environmental Pollution, 144(2): 596-609.

Cosby, B.J., Ferrier, R.C., Jenkins, A. and Wright, R.F., 2001. Modelling the effects of acid deposition: refinements, adjustments and inclusion of nitrogen dynamics in the MAGIC model. Hydrology and Earth System Sciences, 5(3): 499 - 517.

Cosby, B.J., Wright, R.F., Hornberger, G.M. and Galloway, J.N., 1985. Modelling the effects of acid deposition: assessment of a lumped parameter model of soil water and streamwater chemistry. Water Resources Research, 21: 51-63.

De Gruijter, J.J., D.J. Brus, M.F.P. Bierkens & M. Knotters, 2006. Sampling for natural resource monitoring. Springer, Berlin, Heidelberg, New York.

De Vries, W. and Posch, M., 2003a. Critical Levels and Critical Loads as A Tool For Air Quality Management. In: C.N. Hewitt and A.V. Jackson (Editors), Handbook of Atmospheric Science. Principles and Applications. Blackwell Science, Oxford, UK, pp. 562-602.

De Vries, W. and Posch, M., 2003b. Derivation of cation exchange constants for sand, loess, clay and peat soils on the basis of field measurements in the Netherlands. Alterra Report 701, Wageningen.

De Vries, W., Posch, M. and Kämäri, J., 1989. Simulation of the long-term soil response to acid deposition in various buffer ranges. Water, Air, & Soil Pollution, 48(3): 349-390.

De Vries, W., Reinds, G.J., Deelstra, H.D., Klap, J.M. and Vel, E.M., 1999. Intensive Monitoring of Forest Ecosystems in Europe. Technical Report 1999. UN/ECE, EC, Forest Intensive Monitoring Coordinating Institute, 173 pp.

De Vries, W., Reinds, G.J. and Posch, M., 1994. Assessment of critical loads and their exceedance on European forests using a one-layer steady-state model. Water, Air, & Soil Pollution, 72(1/4): 357-394.

De Vries, W. et al., 2003a. Intensive Monitoring of Forest Ecosystems in Europe. Technical Report 2003. UN/ECE and EC, Forest Intensive Monitoring Coordinating Institute, Geneva and Brussels, 170 pp.

De Vries, W. et al., 2001. Intensive Monitoring of Forest Ecosystems in Europe. Technical Report 2001. UN/ECE, EC, Forest Intensive Monitoring Coordinating Institute, Geneva and Brussels, 177 pp.

De Vries, W., Reinds, G.J. and Vel, E., 2003b. Intensive monitoring of forest ecosystems in Europe: 2. Atmospheric deposition and its impacts on soil solution chemistry. Forest Ecology and Management, 174(1-3): 97-115.

Draaijers, G.P.J. and Erisman, J.W., 1995. A canopy budget model to assess atmospheric deposition from throughfall measurements. Water, Air, & Soil Pollution, 85(4): 2253-2258.

EMEP, 2001. Transboundary acidification and eutrophication and ground level ozone in Europe., EMEP Summary Report 2001 EMEP Report 1/01, July 2001. Joint CIAM & CCC & MSC-W report Norwegian Meteorological Institute.

Evans, C.D. et al., 2006. Modelling nitrogen saturation and carbon accumulation in heathland soils under elevated nitrogen deposition. Environmental Pollution, 143(3): 468-478.

Gundersen, P. et al., 2006. Carbon - nitrogen interactions in forest ecosystems, Forest & Landscape Working Papers, Copenhagen, pp. 62.

Gundersen, P., Callesen, I. and De Vries, W., 1998. Nitrate leaching in forest ecosystems is related to forest floor C/N ratios. Environmental Pollution, 102(1): 403-407.

Hettelingh, J.-P., Posch, M. and De Smet, P.A.M., 2001. Multi-effect critical loads used in multi-pollutant reduction agreements in Europe. Water, Air, & Soil Pollution, 130: 1133-1138.

Hettelingh, J.-P. et al., 2007. Critical loads and dynamic modelling to assess European areas at risk of acidification and eutrophication. Water, Air, & Soil Pollution: Focus, 7: 379-384.

Heuvelink, G.B.M. and Pebesma, E.J., 1999. Spatial aggregation and soil process modelling. Geoderma, 89: 47-65.

Jacobsen, C., Rademacher, P., Meesenburg, H. and Meiwes, K.J., 2002. Gehalte chemischer Elemente in Baumkompartimenten., Niedersächsische Forstliche Versuchsanstalt Göttingen; im Auftrag des Bundesministeriums für Verbraucherschutz, Ernährung and Landwirtschaft (BMVEL). Bonn.

Jenkins, A. et al., 1997. Modelling surface water acidification in the UK. , Institute of Hydrology, Wallingford.

Kros, J., Mol Dijkstra, J.P. and Pebesma, E.J., 2002. Assessment of the prediction error in a large-scale application of a dynamic soil acidification model. Stochastic Environmental Research and Risk Assessment, 16(4): 279-306.

Larssen, T. et al., 2006. Forecasting acidification effects using a Bayesian calibration and uncertainty propagation approach. Environ Sci Technol, 40(24): 7841-7847.

Metzger, M.J., 2005. European vulnerability to global change, Wageningen University, Wageningen, 192 pp.

Nilsson, J. and Grennfelt, P., 1988. Critical loads for sulphur and nitrogen. Report from a Workshop held at Skokloster Sweden March 19-24 1988. Miljø rapport 1988: 15. Copenhagen Denmark Nordic Council of Ministers, 418 pp.

Posch, M. and De Vries, W., 1999. Derivation of critical loads by steady-state and dynamic soil models. In: S.J. Langan (Editor), The impact of nitrogen deposition on natural and semi-natural ecosystems. Kluwer, Dordrecht, The Netherlands, pp. 213-234.

Posch, M., Hettelingh, J.-P., Slootweg, J. and (eds.), 2003. Manual for dynamic modelling of soil response to atmospheric deposition. 259101.012, Bilthoven, The Netherlands.

Posch, M. and Reinds, G.J., 2008. VSD – An interactive Very Simple Dynamic soil acidification model for scenario analyses and target load calculations. Submitted.

Reinds, G.J., Posch, M. and De Vries, W., 2001. A semi-empirical dynamic soil acidification model for use in spatially explicit integrated assessment models for Europe. 084, Alterra, Wageningen (Netherlands).

Robert, C.P. and Casella, G., 1999. Monte Carlo statistical methods, New York, 507 pp.

Santore, R.C., Driscoll, C.T. and Aloi, M., 1995. A model of soil organic matter and its function in temperate forest soil development. In: W.W. McFee and J.M. Kelly (Editors), Carbon forms and functions in forest soils. Soil Sci. Soc. of America, Madison, Wisconsin, pp. 275-298.

Schöpp, W., Posch, M., Mylona, S. and Johansson, M., 2003. Long-term development of acid deposition (1880–2030) in sensitive freshwater regions in Europe. Hydrology and Earth System Sciences, 7(4): 436-446.

Starr, M., 1999. WATBAL: a model for estimating monthly water balance components, including soil water fluxes. In: S. Kleemola and M. Forsius (Editors), 8th Annual report 1999. UN ECE ICP Integrated Monitoring. The Finnish Environment 325. Finnish Environment Institute, Helsinki, Finland, pp. 31-35.

Sverdrup, H. and De Vries, W., 1994. Calculating critical loads for acidity with the simple mass balance method. Water, Air, & Soil Pollution, 72(1/4): 143-162.

UBA, 2004. Manual on methodologies and criteria for modelling and mapping critical loads & levels and air pollution effects, risks and trends, UNECE convention on Long Range Transboundary Air Pollution, Federal Environmental Agency (Unweltbundesamt), Berlin.

Ulrich, B. and Pankrath, J., 1983. Effects of Accumulation of Air Pollutants on Forest Ecosystems. D. Reidel Publ. Co., Dordrecht, The Netherlands, 389 pp.

Van Oijen, M., Rougier, J. and Smith, R., 2005. Bayesian calibration of process-based forest models: bridging the gap between models and data. Tree Physiology, 25(7): 915-927.

Visual Numerics, 1997. IMSL Stat/Library, Volume 1, Houston.

Vrugt, J.A. et al., 2006. Application of parallel computing to stochastic parameter estimation in environmental models. Computers & Geosciences, 32(8): 1139-1155.

Warfvinge, P., Falkengren-Grerup, U., Sverdrup, H. and Andersen, B., 1993. Modelling long-term cation supply in acidified forest stands. Environmental Pollution, 80: 209-221.