Article (refereed)

De Vries, W.; Wamelink, G.W.W.; van Dobben, H.; Kros, J.; Reinds, G.J.; Mol-Dijkstra, J.P.; **Smart, S.M.**; **Evans, C.D.**; **Rowe, E.C.**; Belyazid, S.; Sverdrup, H.U.; van Hinsberg, A.; Posch, M.; Hettelingh, J-P.; Spranger, T.; Bobbink, R.. 2010 Use of dynamic soil-vegetation models to assess impacts of nitrogen deposition on plant species composition: an overview. *Ecological Applications*, 20 (1). 69-79. 10.1890/08-1019.1

Running Head. – modelling nitrogen deposition impacts

Use of dynamic soil-vegetation models to assess impacts of nitrogen deposition on plant species composition and to estimate critical loads: an overview

Wim de Vries[1], Wieger Wamelink[1], Han van Dobben[1], Hans Kros[1], Gert Jan Reinds[1], Janet Mol-Dijkstra[1], Simon Smart[2], Chris Evans[2], Ed Rowe[2], Salim Belyazid[3], Harald Sverdrup[3], Arjen van Hinsberg[4], Maximilian Posch[4], Jean-Paul Hettelingh[4], Till Spranger[5] and Roland Bobbink[6]

[1] Alterra, Wageningen University and Research Centre, P.O. Box 47, 6700 AA Wageningen, The Netherlands

[2] Centre for Ecology and Hydrology, Orton Building, Deiniol Road, Bangor LL57 2UP, UK

[3] Lund University, Department of Chemical Engineering, Box 124, S-22100, Lund, Sweden

[4] Netherlands Environmental Assessment Agency (PBL), PO Box 303, 3720 AH Bilthoven, The Netherlands

[5] German Federal Environment Agency (UBA), Wörlitzer Platz 1, 06844 Dessau, Germany

[6] Research Centre B-WARE, Radboud University Nijmegen, PO Box 9010, 6500 GL Nijmegen

Abstract

Field observations and experimental data of effects of nitrogen (N) deposition on plant species diversity have been used to derive empirical critical N loads for various ecosystems. The great advantage of such as approach is the inclusion of field evidence, but there are also restrictions, such as the absence of explicit criteria regarding significant effects on the vegetation, and the impossibility to predict future impacts when N deposition changes. Model approaches can account for this. In this paper, we review the possibilities of static and dynamic multi-species models in combination with dynamic soil - vegetation models to (i) predict plant species composition as a function of atmospheric N deposition and (ii) calculate critical N loads in relation to a prescribed protection level of the species composition. The similarities between the models are presented, but also several important differences, including the use of different indicators for N and acidity and the prediction of individual plant species versus plant communities. A summary of the strengths and weaknesses of the various models, including their validation status, is given. Furthermore, examples are given of critical load calculations with the model chains and their comparison with empirical critical N loads. We show that linked biogeochemistry-biodiversity models for N have potential for applications to support European policy to reduce N input, but the definition of damage thresholds for terrestrial biodiversity represents a major challenge. There is a also a clear need for further testing and validation of the models against long-term monitoring or long-term experimental datasets and against large-scale survey data. This requires a focused data collection in Europe, combing vegetation descriptions with variables affecting the species diversity, such as soil acidity, nutrient status and water availability. Finally, there is a need for adaptation and upscaling of the models beyond the regions for which dose-response relationships have been parameterised, to make them generally applicable.

Key words: soil-vegetation models, model validation, nitrogen deposition, critical loads, plant species composition, biodiversity, terrestrial ecosystems, plant communities

1 Introduction

Impacts of nitrogen deposition on plant species composition

During the past two decades the reduction of sulfur (S) emissions and the persistence of a high N pressure on terrestrial and aquatic ecosystems shifted attention from effects of S-deposition and acidification towards effects of N-deposition and eutrophication. In Europe, N is the most important air pollutant affecting plant species diversity. Evidence suggests that increasing N availability often causes an overall decline in plant species diversity (Tilman, 1987; Bobbink et al., 1998) even at long-term low-N inputs (Clark and Tilman, 2008). In some cases, especially under very nutrient poor conditions, however, an increase in plant species diversity has been observed due to the expansion of nitrophilic species (Emmett, 2007). Effects of N deposition, either in the form of ammonia (NH_3), ammonium (NH_4), nitrogen oxide (NO_x) or nitrate (NO_3), are now recognised in nearly all oligotrophic and mesotrophic (semi-)natural ecosystems. An overview of effects on plant species diversity, including impacts on mosses, lichens and mycorrhizae, in forests, grasslands, heathlands, oligotrophic wetlands (mire, bog and fen), and coastal habitats, mainly in Europe, with related empirical critical N loads, is presented in Achermann and Bobbink (2003). More recently , an overview of effects of N deposition on a global scale, distinguishing Arctic and Alpine ecosystems, boreal forests, temperate forests and tropical forests, heathlands and grasslands, Mediterranean vegetation, tropical savannas and arid vegetation (desert and semi-desert) is presented in Bobbink et al. (2009).

Critical loads and their use in policy making

In order to set standards and targets for emission reduction policy, the concept of critical load has been developed. The general definition of a critical load is ‘a quantitative estimate of an exposure to one or more pollutants below which significant harmful effects on specified sensitive elements of the environment do not occur according to present knowledge' (Nilsson and Grennfelt, 1988). Critical loads are defined for specific combinations of pollutants, effects and receptors. They reflect spatially variable sensitivities, thus leading to regionally defined emission-reduction needs. The concept is most commonly used in connection with the atmospheric deposition of S and N (acidification and eutrophication) and in these cases the critical load is the maximum flux (in kg N $ha^{-1}.yr^{-1}$ or keq H $ha^{-1}.yr^{-1}$) that an ecosystem is able to sustain. Since 1994, critical loads for N and acidity have played an important role in European air pollution abatement (Hettelingh et al., 2001; Spranger et al., 2008). European critical load exceedances, calculated and mapped using the latest methods and datasets for critical loads, deposition, and emission scenarios, are presented in Hettelingh et al. (2007) and Slootweg et al. (2007). Results show that the area where critical loads of acidity are exceeded and will continue to decrease, even if no new legislation is implemented, while high (> 10 kg N $ha^{-1}.y^{-1}$) exceedances for critical N loads remain widespread especially in north-western European areas dominated by ammonia emissions.

Exceedances of the critical load of acidity and N have been used in European pollution abatement policy for defining emission-reduction targets, i.e. in the UNECE Convention on Long-range Transboundary Air Pollution (LRTAP Convention) and the European Union (National Emission Ceilings Directive, 2001; EC, 2005). Integrated assessment models (e.g. the RAINS model, Amann et al., 1999) use these data and methods in scenario analyses. The exceedance of critical loads of N is also used as an indicator for risk to biodiversity by the European Environment Agency (EEA, 2007). The general organization of effects-based European air pollution policies is described in www.unece.org/env/lrtap.

Critical load approaches for nitrogen used in policy making

Critical loads for N, as used in European environmental policy, are estimated empirically or by simulation. The empirical approach uses experimental fields or 'mesocosms' where various levels of N fertiliser have been added. In that case, the critical load is determined as the level of deposition where a decrease in biodiversity just starts to occur (Achermann and Bobbink, 2003). The great advantage of the use of empirical critical N loads, based on N-addition experiments, is that there is field and/or experimental evidence for a relationship between N deposition and effects. There are, however, several requirements for empirical critical N loads to be reliable, including: (i) long-term experiments (preferably > 4 year) to show long-term effects and (ii) studies in low background-N deposition areas to ensure that major effects have not already occurred. In high background N deposition, high-N additions may be needed before (additional) effects show up and N removal experiments should be used instead (Emmett, 2007). The empirical critical N loads are thus mainly based on long-term field-addition experiments and mesocosm studies in low-N-deposition areas with realistic N loads (<100 kg $N.ha^{-1}.yr^{-1}$). However, since such experiments are time- and labour-intensive, results are only available for a rather limited group of broadly defined ecosystems. The reliability (range) in empirical critical loads, being the level of the lowest N addition where effects occur, is mainly influenced by the chosen interval in N additions and by an uncertain background N deposition that has to be added to this level (Sutton et al., 2003). An aspect that also limits a strict comparability is the lack of fixed criteria regarding significant effects for which a critical load is derived (as e.g. in critical limits for toxic substances), and it is also impossible to predict future impacts when N deposition and other environmental conditions change simultaneously.

The model-based critical load approach, used in European environmental policy making, is based on an ecosystem mass balance, which balances the deposition load to an ecosystem with its long-term capacity to buffer this input or to remove it from the system without harmful effects inside or beyond the system (Hettelingh et al., 2001; Spranger et al., 2008). The harmful effects are defined in terms of critical limits above which a negative effect is assumed to occur. An overview of those limits is given in De Vries et al. (2007). The model calculates a critical load as the deposition level leading to soil conditions that are just tolerated by a given ecosystem. The model-based critical load approach as used up to now, however, is by definition based on the sustainable state of a given ecosystem that is invariable in time (steady-state) and excludes non-permanent buffering processes such as temporary N release and retention and cation exchange. This long-term critical load may therefore differ from the atmospheric deposition actually affecting the ecosystem, since ecosystems differ in sensitivity to perturbation depending on their current state and recent history. Nitrogen deposition thresholds may vary during forest stand development, as for example shown also with dynamic-model approaches (Tietema et al., 2002). These differences form the core of resilience and sustainability theories. This aspect has not been included in model based critical loads until now, and their steady state concept implies that the exceedance of such critical loads does not allow a prognosis of ecosystem status at any point in time.

Need for dynamic model approaches and aim of this paper

Both empirical critical loads and steady-state models do not allow prediction of the temporal response of ecosystems to deposition scenarios, for example, in terms of impacts on plant species diversity. This requires the use of the dynamic integrated soil-vegetation models. Such models can also be used to assess critical loads, while accounting for differences in sensitivity to perturbation depending on their current state and recent history. In the context of ecological theory, N deposition is a form of disturbance; i.e. an external influence that moves the system away from its stable state (Gunderson, 2000). If such a disturbance is not too large, the ecosystem has the ability to return to its former state ('resilience', see e.g. Gunderson, 2000); in the case of N deposition, this might happen by incorporation of N in refractory soil organic material or N leaching. However, if the disturbance is larger or extends over a prolonged period, the system may move towards an alternative stable state, from which it will not be able to return to its former state without a new external influence (Ludwig et al., 1997). N deposition will ultimately stimulate the growth of more productive species, that usually produce more easily degradable litter and reach a greater height, thus increasing both the deposition itself, and the amount of N cycling in the system. In these terms, the critical load is the highest deposition that will not cause an ecosystem to shift to an alternative (more productive, usually species-poorer) state. In this alternative state the quantity of N cycling through the system will be much larger than in its original state and, because of the tight cycling of N- a return to the former, N-poor state will only be possible by physically removing the excess N even if deposition decreases (Wamelink et al., 2008). Critical-load assessments including such aspects can be only be included in dynamic model approaches, simulating delays in damage due to buffering processes and delays in recovery due to time take to restore soils to its original state.

In this overview, we describe the possibilities of multi-species models in combination with dynamic soil - vegetation models to (i) predict plant species composition or diversity as a function of atmospheric N deposition and (ii) calculate critical N loads in relation to an acceptable plant species diversity change. First, we present the two main model approaches that are presently employed in Europe: (i) a simple soil acidification and nutrient cycling model (SMART2 or MAGIC) combined with field-based empirical relationships with plant species responses (MOVE, GBMOVE or NTM) and (ii) a detailed, mechanistic, soil-acidification and nutrient-cycling model (ForSAFE) with a process-based description of plant species responses (VEG). An explanation of the model acronyms and main features appears in Table 4 and 5. The overview includes a description of each modelling approach, followed by application examples illustrating the model validation status and the use of the models in critical-load assessments. In a final section we discuss the potential of linked biogeochemistry-biodiversity models to support European pollution abatement policy, including (i) strengths and weaknesses of the two major model approaches, (ii) the use of different indicators for N availability, (iii) the validation status of each model, (iv) the potential of the models to assess critical loads, (v) the need for additional field surveys and (vi) relevant extensions to the modelling approaches.

2 *Modelling approach*

Integrated soil-vegetation models are used at present in Europe to predict plant species composition as a function of atmospheric deposition of N and acidity, as illustrated in Figure 1. The principle of such model-based approaches is that a dynamic soil model (SMART2, MAGIC, (For)SAFE) predicts the changes in water and nutrient status (e.g. as N availability or C/N ratio) and soil acidity (e.g. as soil pH or base saturation) in response to atmospheric deposition, whereas a statistical model (NTM, MOVE) or a process-based model (SUMO, VEG) predicts vegetation succession or changes in plant species composition in response to the changes in water, nutrient and acidity status, using plant species-specific information on habitat preferences. Such coupled models can be used in an inverse way to determine critical loads. In that case, critical values for abiotic factors (e.g. N availability or soil pH) have to be empirically determined per vegetation type, either directly or from information per species (step 1) and subsequently used in the coupled soil model (step 2) to back-calculate the critical N and acid loads.

Simulation of critical loads according to the above principle was carried out in the Netherlands (Van Dobben et al., 2006), where (i) the critical pH and N availability per vegetation type (association) were determined on the basis of a large set of vegetation relevés (vegetation description of a small plot), often , (cf. MOVE model, Latour and Reiling, 1993), and (ii) the dynamic soil model SMART2 (Kros et al., 1995) was used to calculate the critical loads at which the above critical limits were not exceeded in the long term. Other models use critical limits for other abiotic variables, such as the C/N ratio in the GBMOVE model (Smart et al., 2003) and the BERN model (Schlutow & Hübener, (2004). or the soil N, P, BC availability, soil moisture, pH, light and grazing pressure in the ForSafe-VEG model (Belyazid et al., 2006; Sverdrup et al., 2007).

Below, we discuss the two major model approaches used at present, i.e. (i) dynamic soil models linked with empirical static-vegetation models (the SMART2(-SUMO)-MOVE/NTM and MAGIC(-SUMO)-GBMOVE model chains) and (ii) a deterministic, dynamic ecosystem model integrating hydrology, growth, biogeochemical cycles and vegetation dynamics. (ForSafe-VEG)

2.1 Linked dynamic soil models with empirical static vegetation models

Two major comparable model chains of dynamic soil models linked with static vegetation models are SMART2(-SUMO)-MOVE/NTM and MAGIC(-SUMO)-GBMOVE, which are developed and used in the Netherlands (NL) and the United Kingdom (UK), respectively. The model chains consist of: (i) the soil models SMART2 (NL) or MAGIC (UK) that simulate the cycling of nutrients in the soil and predict soil acidity and N availability, (ii) the succession model SUMO that simulates the cycling of nutrients (N, P, K, Ca, Mg) in the plant-soil system, including biomass growth through photosynthesis and biomass removal through management and (iii) multiple regression equations between species presence and abiotic factors that define the realized niches of a substantial proportion of the vascular flora of each country (and in the UK also bryophytes) (MOVE in NL and GBMOVE in UK), or of plant communities (NTM in NL). In the Dutch MOVE and NTM models, abiotic factors are groundwater table, soil pH and N availability, derived via Ellenberg indicator values, whereas a version of the UK GBMOVE model also includes three climatic variables, i.e. the minimum January temperature, maximum July temperature and precipitation. Ellenberg's indicator values are classes of plants species with similar ecological niches, which are derived for about 2720 central-European vascular plants (Ellenberg et al., 1992). Ellenberg derived values for the following ecological factors: light (E_L), temperature (E_T), continentality (E_K), moisture (E_F), soil pH (E_R), nutrients/nitrogen (E_N), and others (salinity, heavy metal resistance) (E_{sonst}). SMART2 can be used both in its original, dynamic form, allowing the calculation of target loads, and as a steady state version, allowing the calculation of steady state critical loads, as used in policy making. The use of SUMO is optional in both model chains. Changes in species composition are modelled by first simulating the effects of N and S deposition on soil conditions followed by simulating the impacts of changed soil conditions on species composition.

Modelling the relation between atmospheric deposition and soil conditions

SMART2 and *MAGIC* are dynamic, process-oriented models that predict changes in soil chemistry at a given level of N and S deposition. Changes of N and S deposition on soil variables such as pH, C/N ratio or N availability by SMART2 (Kros et al., 1995; Kros, 2002) or MAGIC (Cosby et al., 2001). Both SMART2 and MAGIC include the major hydrological and biogeochemical processes in the soil compartment, to calculate the long-term effects of atmospheric deposition of NO_x, NH_y, SO_x and base cations (BC^{2+}) on soil-solution chemistry, and in case of MAGIC also the surface-water chemistry. The models have a high degree of process aggregation to minimize their data requirements, which allows application on a regional scale. They consists of a set of mass-balance equations, describing the soil input-output relationships, and a set of equations describing the rate-limited and equilibrium soil processes. Apart from pH, the models predict changes in aluminium, base cation, ammonium, nitrate and sulfate concentrations in the soil solution and solid phase. Key parameters include the input and output fluxes of base cations and strong acid anions, the soil cation exchange capacity, and the fraction of this capacity that is occupied by Ca, Mg, Na and K ions. Nitrogen dynamics in MAGIC are based on empirical relationships between net N retention and the current C/N ratio in the soil, whereas in SMART2 litterfall, mineralization, root uptake and immobilization are modelled explicitly. Both SMART2 and MAGIC have an internal simplified growth module which enable the models to calculate nutrient cycling detached from SUMO. The detached version of SMART2 has been used for the calculation of critical loads and target loads. For the computation of target loads, a procedure was developed to iteratively run SMART2 until the N and S deposition used, lead to the critical pH or N availability for a given vegetation type. Furthermore, a steady-state version of

SMART2 has been developed that computes the critical N and acid load that in steady state leads to a given combination of N availability and pH. A complete overview of these models, and the differences between SMART2 model and MAGIC can be found in the various references mentioned and in De Vries et al. (2007).

SUMO is a process-based model that simulates biomass growth under given soil, climate and management conditions (Wamelink, 2007). The basis of the model is a maximum growth that is being reduced by a series of linear and non-linear reduction factors to constrain growth. These reduction factors convey the effect of changes in the availability of light, N, phosphorous, water and temperature. SUMO distinguishes five functional plant types (climax trees, pioneer trees, shrubs, dwarf shrubs and herbs) that compete for light and nutrients. Their competitive balance is governed by vegetation structure, i.e. canopy height and biomass of roots and leaves per functional type. Management is simulated as biomass removal by mowing, grazing, cutting or turf stripping. The accumulation of biomass in the five functional types determines the succession stage (e.g. pioneer, grassland, heathland, forest). SUMO can be coupled to niche models. e.g. NTM or MOVE, through vegetation structure and soil chemical conditions (pH and nutrient availability) simulated by SMART2 or MAGIC. For these soil models litterfall is a crucial input term that is generated by SUMO. In each time step there is feedback between SMART2 or MAGIC and SUMO; the models exchange information about N and P, litterfall, and vegetation structure.

<u>Relationships between plant species occurrence and soil conditions</u>

MOVE and NTM: The models MOVE (Latour and Reiling, 1993; Latour et al., 1994) and NTM (Schouwenberg et al., 2000; Wamelink et al., 2003a) are based on response curves in which the probability of plant species (MOVE) or plant community (NTM) occurrence is determined by vegetation structure and the abiotic site conditions of groundwater table, soil pH and N availability. The probability of occurrence is a simple bell-shaped curve derived for 914 species by

2nd-order logistic regression based on presence/absence, representing species occurrence along an environmental gradient. These relationships are based on the realised niche, i.e. they account for competitive exclusion, rather than responses of the species in isolation. Since MOVE and NTM focus on more than one abiotic factor, the curves are multi-dimensional. The probabilities of occurrence are determined per vegetation type relative to soil pH and N availability, estimated on the basis of Ellenberg's (1992) indicator values for N (E_N) and acidity (E_R). In the critical-load approach of Van Dobben et al. (2006), the 20- and 80-percentiles of these frequency distributions were used as the critical limits, i.e. the range between these percentiles was considered as the optimal range for each vegetation type.

The above frequency distributions were determined in a database of 160 000 vegetation relevés that were labelled in terms of vegetation type (Schaminée et al., 1989), originally developed for a revision of the Dutch classification of plant communities. In a separate procedure, the Ellenberg values (which are on an arbitrary scale) were translated into physical units that can be used as input to dynamic models. This translation requires a training set where vegetation and soil conditions (at least pH and N availability) have been recorded simultaneously. In the past few years, much effort has been put into the collection of such data (Wamelink et al., 2007, see www.abiotic.wur.nl). Various translation functions between Ellenberg values and physical units have been derived (e.g. (Alkemade et al., 1996; Ertsen et al., 1998; Wamelink et al., 2002; Van Dobben et al., 2006). Those of Van Dobben et al. (2006) for example run:

$$\mathrm{pH} = 3.1 + 0.53 \cdot \mathrm{E_R} \quad R^2 = 0.43;\ n = 3630 \tag{1}$$

$$\mathrm{N_{av}} = 6.19 + 0.64 \cdot \mathrm{E_N} + c \cdot \mathrm{vegtype} \quad R^2 = 0.24;\ n = 6911 \tag{2}$$

Where N_{av}= N availability ($kMol.ha^{-1}.yr^{-1}$) and the constants c per vegetation type are -1.182 for grass, -1.898 for heath, -0.274 for coniferous forests and 0 for deciduous forest

GBMOVE: As with MOVE, multiple logistic regression was used to construct empirical equations that predict habitat suitability for higher and lower plants representative of British plant communities, based on their abundance along key environmental gradients as recorded by extensive relevé data (e.g. Roy et al., 2000). Each equation consists of regression coefficients that apply to either four or seven explanatory variables, depending on whether climate variables (minimum January temperature, maximum July temperature and precipitation) are included or not. Important interaction terms are also included. These quantify the extent to which a species' response on one gradient is conditioned by another gradient (e.g. Pakeman et al., 2008). The data used to derive each equation were assembled from a variety of sources as described in De Vries et al. (2007) and covered more than 40 000 vegetation relevés. The regression was based on presence/absence data for each plant species in each plot paired with values of climatic variables (derived from the plot's geographical position) and plot-averaged Ellenberg indicator values. The final number of species having GBMOVE regression models is 327 for bryophytes and 803 for vascular plants in non-coastal habitats (74 in coastal habitats).

As with MOVE, soil pH and soil C/N ratio are translated instantaneously into mean Ellenberg E_R and E_N values respectively, using paired soil measurements and mean Ellenberg values from the Countryside Survey 1998 database (Smart et al., 2003). The limitations of the assumption that species presence immediately changes in response to soil conditions are discussed in detail in Section 5. Relationships thus obtained are:

$$\ln(C/N) = 3.61 - 0.63 \cdot \ln E_N \quad R^2 = 0.62;\ n = 256 \tag{3}$$

$$pH = 2.5 + 0.61 \cdot E_R \quad R^2 = 0.61;\ n = 256 \tag{4}$$

The mean Ellenberg E_R and E_N values per relevé are terms in the GBMOVE regression equations. At each time step the simulated values of soil C/N, soil pH, % soil moisture and cover-weighted canopy height are translated into Ellenberg units and put in the regression equation, resulting in predicted probability of species occurrence over time. Changes in soil pH and C/N ratio are predicted with the dynamic soil model MAGIC. Canopy height can be changed arbitrarily using pre-existing knowledge of the pace of succession in a particular location, or on a more process-linked basis, by the SUMO succession model. Climate variables can be changed to mimic expectations under different climate change scenarios. Likewise, soil moisture can also be changed to mimic drainage or drought.

2.2 The integrated dynamic FORSAFE-VEG model

The only fully integrated dynamic soil and plant species diversity (vegetation) model that is presently available in Europe is the ForSAFE-VEG model chain. This model chain, developed in Sweden, consists of: (i) the ForSAFE model, aimed at the dynamic simulation of changes in soil chemistry, soil organic matter, hydrology and tree biomass growth in relation to changes in environmental factors (Wallman et al., 2005); and (ii) the VEG submodel, which simulates changes in the composition of the ground vegetation in response to changes in biotic and abiotic factors such as light intensity at the forest floor, temperature, grazing pressure, soil moisture, soil pH and alkalinity in addition to competition between species based on height and root depth (Belyazid et al., 2006; Sverdrup et al., 2007). For each time step, defined by the resolution of the input data, ForSAFE simulates the changes in state variables in response to environmental changes (temperature and precipitation, atmospheric deposition, forest management). These state variables are read by the VEG module, where the occupancy strength is calculated for each plant group. The plant groups are defined by the user. The single occupancy strengths are then used to calculate the relative occupancy of each plant group.

If a stress factor would eliminate a certain species, the disappearance of this species will not be instantaneous, but will happen with a delay, which depends on the lifespan of the species. Unlike the model chains with MOVE and GBMOVR, this aspect is included in ForSAFE-VEG. The change in occupancy of a specific plant group, *dX/dt,* depends on the actual occupancy of the plant group (*X*), the target occupancy (referred to as equilibrium occupancy X_{eq}), and the specific regeneration time of the plant group (τ) according to the following equation:

$$\frac{dX}{dt} = \frac{1}{\tau} \cdot (X_{eq} - X) \tag{5}$$

The regeneration time τ is related to the lifespan of a specific plant group. The life span depends on site factors, such as drought. The equilibrium occupancy of a plant group *i*, $X_{eq,i}$, is the ratio between the strength of the species under the specific environmental conditions and the sum of the strengths of all present species according to the following equation:

$$X_{eq,i} = \frac{S_i}{\sum_{j=1}^{j=plantgroup} S_j} \tag{6}$$

where S_i is the individual strength of the plant group *i*. The sum of plant group strengths is also used as an indicator of the density of the ground cover, referred to as the mass index (MI). The strength of each plant group is the product of the following drivers: (i) soil solution N concentration (mol.l^{-1}), (ii) soil solution phosphorus concentration (mol.l^{-1}), (iii) soil acidity ([H^+], [BC^{2+}], [Al^{3+}] (eq.l^{-1}), (iv) soil water content (m^3 water.m^{-3} soil), (v) soil temperature (°C), (vi) light reaching the ground (µmol photon.m^{-2}.s^{-1}), (vii) grazing (moose units km^{-2}), (viii) wind tatter and wind chill damage, (ix) plant competition based on above-ground competition for light and below-ground competition for water and nutrients and (x) air CO_2 concentration. ForSAFE-VEG thus simulates the ground vegetation occupancy based on the individual response of plant groups to these controlling factors. The effects are multiplicative and have the same weights in affecting the plant strength. . Each plant group represents various individual plant species, varying from less than ten up to several hundreds.

For each plant group indicator that has been selected, response functions were parameterised for Sweden from published laboratory and field data, by approximations from empirical data, or by scaling the response with respect to other plant groups for which the response is known. Scaling of plant groups towards known responses was based on generic knowledge and expert opinions from Swedish plant ecologists. However, the basic shape of each response function does not vary between the plant groups. For example, all plant groups will respond positively to an increase in water availability in the soil up until a certain level where anaerobic conditions in the saturating soil may hinder the plant's growth. The distinction between the plant groups is the minimal water content required for survival, optimal water content for growth, and the point at which water becomes damaging. The individual response functions are described in detail in Belyazid (2006) and De Vries et al. (2007).

3 Model validation on changes in soil and vegetation data

Results of both types of model chains were compared with measurements on changes in soil and vegetation data. The performance of the models was calculated by two measures that are often applied for this purpose; the Normalized Root Mean Square Error (NRMSE; eq. 7), and the Normalized Mean Error (NME; eq. 8) (Janssen and Heuberger, 1995).

$$\mathrm{NRMSE} = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(P_i - O_i\right)^2}}{\overline{O}} \qquad (7)$$

$$\mathrm{NME} = \frac{\overline{P} - \overline{O}}{N} \qquad (8)$$

Where:

Pi = predicted value: model output for N2O emissions

Oi = observed value: field value for N2O emissions

$\overline{P}$ = average for the predicted values

$\overline{O}$ = average for the observed values

N = number of observations

NRMSE describes the deviations between the measurements and the predictions in a quadratic way and is thus rather sensitive to extreme values. Optimally, it should be 0. The NME compares predictions and observations over the entire time-span, on an average basis. It expresses the bias in average values of model predictions and observations, and gives a rough indication of overestimation (NME>0) or underestimation (NME<0).

3.1 Linked dynamic soil models with static vegetation models

Below, examples are given of the validation of the SMART2/MAGIC-GBMOVE/NTM model chains with respect to soil chemical data (focus on MAGIC), aboveground biomass data (SUMO) and time series of observed species composition and species richness (focus on GBMOVE in combination with MAGIC).

Validation of MAGIC on time series for soil chemical data: Data from plot-scale N-manipulation studies in the UK have been used to test the ability of MAGIC to predict changes in soil C/N under different addition levels (Evans et al., 2006). For two sites with high-quality soil C and N data (Figure 2), the model successfully reproduced observed decreases in C/N under three treatments. These simulations incorporated an (observed) increase in C storage as a consequence of N deposition, which slowed down the rate of C/N change. The NME's derived were 0.0388 for the Ruabon site and 0.0813 for the Budworth site, indicating a slight overestimation of the predictions. The calculated NRMSE's were very low, i.e. 0.0065 and 0.014, respectively.

MAGIC was also validated on data on C/N ratios for the Parkgrass experimental site at Rothamsted, which are available for a 100-year period. N removal was calculated by multiplying hay removal, for which accurate measurements are available, by the proportion of N in hay biomass. The uncertainty in N concentration in hay has a large effect on the net addition (deposition minus removal) and thus on the historic C/N trajectory (Table 1). Note that raising N deposition would have the same effect as reducing N removal.

Validation of SUMO on time series of above-ground biomass: Biomass growth predicted by SUMO was validated using data collected at two unfertilized grassland sites, one in the Netherlands and one in the UK, using site specific historical deposition data. The first grassland site is situated near Wageningen in the Netherlands and is part of a long-term field experiment started in 1958 on former agricultural land (Elberse et al., 1983). The second grassland site is the Parkgrass experimental site at Rothamstead in the UK mentioned before. The Parkgrass site was mown twice a year and the harvested biomass was weighed and averaged over ten year periods. The experiment started in 1856 and still continues today. The trends in herbage yields are extensively described by Jenkinson et al. (1994). At the Dutch site, the measured biomass varies greatly between years due to yearly differences in rainfall and temperature. The simulated biomass does not vary much and remains within the range of the measured biomass (Figure 3). The high measured biomass in the first year is probably caused by the former agricultural use of the field. Both the measured and the simulated biomass show a decrease over the years, due to the yearly biomass removal. The results for Rothamstead show that the harvested biomass is fairly well simulated by SUMO. The reduction in biomass harvest between 1850 and 1900, due to exhaustion of the soil, the stabilisation of the harvest when the effect of N deposition compensated for the exhaustion between 1900 and 1950, and the increase of the harvest later on due to the further increase in deposition since 1950 are simulated quite well. The effect of N deposition since approximately 1960 is underestimated, however (Figure 3 right). Overall, the NME of the Ossekampen was -0.012 and -0.0033 for Rothamsted, implying that on average, the predictions are almost equal to the observations. The values for the NMRSE are 0.290 and 0.195, respectively indicating a considerable deviation for defined years. SUMO was also validated on a heathland and a forest site in the Netherlands, as described in De Vries et al. (2007).

Validation of MAGIC-GBMOVE on time series of observed species composition and species richness: A number of tests have been carried out to determine how successfully MAGIC-GBMOVE could reproduce the observed species composition in sampled plots. Observations were compared with predictions generated initially by populating a simulated set of plots with species conditioned on probability of occurrence values generated by GBMOVE and a Poisson distribution of mean species-richness values with proportional variance predicted by a separate General Linear Mixed Model using the same explanatory variables as GBMOVE. This statistical model is fully described in Smart et al. (2005). To assess the influence of uncertainty in the calibration equations relating soil properties to mean Ellenberg scores, predictions of species composition based on soil C/N and pH generated by MAGIC were compared with predictions based on observed mean Ellenberg scores. These comparisons were carried out for control plots at the long-term continuous Park Grass hay experiment at Rothamsted (unimproved neutral grassland) and the Hard Hills grazing and burning experiment at Moorhouse National Nature Reserve (ombrogenous bog), described in Smart et al. (2005). The results for Rothamsted indicated that when mean Ellenberg scores based on observed species composition were used as input to GBMOVE, on average 67% of species observed were actually predicted. However, when predictions were based on MAGIC simulations of soil C/N and pH as input to GBMOVE, the percentage of all species correctly predicted decreased substantially (Figure 4 left). The main reason for the poor performance at Rothamsted appears to be that observed soil changes were inconsistent with observed vegetation changes. This is most likely due to sampling practices in the experimental plots, avoiding a thin mat of persistent litter that had developed in the O horizon over the course of the experiment. This resulted in C/N measurements indicating a higher fertility in the rooting zone than that encountered by at least some of the more shallow rooting species present. Also at Moorhouse, both the predictions from MAGIC linked to GBMOVE and the predictions based solely on observed mean Ellenberg values did not compare well with the species actually observed in the control plots, although the key dominants in the vegetation were predicted to be present by both models. (Figure 4 right). When predicted species lists for both GBMOVE and MAGIC+GBMOVE were examined, key absences included a range of bryophytes. Possibly, bryophytes are more responsive to direct deposition effects, and less so to changes in soil chemistry, making the approach less useful for lower plants.

From this model comparison it was concluded that it is unlikely that both generally applicable yet highly accurate models can be developed, because of the dependence of current species composition on site-specific aspects of patch and wider landscape history. Because of this, probabilities of occurrence from GBMOVE are no longer used as expectations of species presence, but rather interpreted as indices of habitat suitability where target species ought to be able to persist and increase in population size in the absence of constraints to dispersal and establishment. The further validation work therefore focuses on comparing predicted trends through time with observed changes, based on species present on-site and in the local species pool, rather than attempting to predict the entire species assemblage on a specific site.

Validation of MAGIC-GBMOVE on temporal changes among plant species: At Moorhouse, observed versus predicted species changes over time were summarised as slope coefficients for each species in a linear regression line relating (i) observed abundance over the years and (ii) predicted habitat suitability from MAGIC and GBMOVE across the same time period. Despite considerable scatter there was a positive correlation between observed change in species frequency and predicted change in habitat suitability (Figure 5). A chi-square test of observed versus predicted directions of change was significant (p=0.016). While the correlation between observed and predicted slopes was also significant, predicted rates of change covered a narrower range than observed species changes, which may be due to weather fluctuations or sampling errors.

3.2 The integrated dynamic FORSAFE-VEG model

The ForSAFE-VEG model was validated on changes in soil chemistry (Belyazid et al., 2006), standing wood biomass (Belyazid, 2006) and in ground vegetation cover (Sverdrup et al., 2007) at 16 Swedish forest sites that are part of the ICP Forest level II monitoring network (International Co-operative Program on Assessment and Monitoring of Air Pollution Effects on Forests, http://www.icp-forests.org). At the sites, 42 plant groups and 9 tree seedling types have been identified. These plant groups were assumed to be potentially present throughout Sweden, but are only expected to manifest where environmental conditions are favourable. The sites cover a wide range of climatic conditions, soils, fire regimes, atmospheric deposition gradients and management histories. ForSAFE-VEG was used to simulate the changes in soil chemistry, hydrology and tree biomass according to these conditions, and the composition of the ground vegetation was subsequently derived. Atmospheric deposition data for NO_3^{+}+NH_4^{+} and SO_4^{-} were derived on the basis of EMEP model estimates according to the 1999 LRTAP Gothenburg protocol (Schöpp et al., 2003). The sites were subject to different histories of fire regimes, alterations between open fields and forest cover as well as different harvesting regimes depending on the location of each site.

Validation on soil chemical data. Simulated soil organic matter contents showed a reasonable correlation between the measured and modelled values of soil organic carbon (C) and N at the 16 study sites. The NME's derived were -79 g/m^2 for organic C and -3.04 g/m^2 for organic N, indicating a slight underestimation of the predictions. The calculated NRMSE's were quite high, i.e. 0.58 for both organic C and N.

The model reconstructs the pH profiles at the 16 study sites quite well (Figure 6). The NME for the 16 sites varied from -0.056 to + 0.097 while the NMRSE varied from 0.016 to 0.210, indicating an appropriate prediction of the average pH and a limited deviation with measurements at various depth. The model, however underestimates the acidity at the deeper soil layers (Figure 6). This inconsistency is probably due to the fact that the model considers only a limited amount of roots at the deep layers, thus underestimating uptake and the presence of organic matter and its decomposition. Also important for the ground-vegetation community is the soil base cation to aluminium ratio (BC/Al ratio). The variation in both the measured and modelled BC/Al ratios was large for most of the sites, but the correspondence between the model and the measurements was reasonably good. More information on the validation of soil organic C and N and of soil pH is given in Belyazid (2006) and Belyazid et al. (2006), respectively.

Validation of standing tree biomass and ground vegetation composition: Predicted values for the ground occupancy of the 42 identified plant groups calculated with ForSAFE-VEG for the year 1995 were plotted against measurements from the same year to establish the validity of the model outputs at the 16 sites (Sverdrup et al., 2007). Results are presented for two representative sites, i.e. Brattfors

(Figure 7 left) and Svartberget (Figure 7, right). The model predicts fairly well the occupancy of the present vegetation groups. The NMRSE varied from 0.167 for Brattfors to 0.197 for Svartberget, while the NME is close to 0. Recently, the model has also been validated on two Swiss forest plots (Aeschau and Bachtel). A comparison of the model output for these test sites to ground vegetation assessment showed that only 40-55% of the species present at these two sites were also modelled.. The presence of major species, i.e. *Vaccinium myrtillus, Blechnum spicant, Dryopteris dilatata, Polytrichum formosum, Rubus fruticosus* and *Oxalis acetosella* were, however, forecasted correctly in both sites, for the latter two species even predicted with a correct estimate of the cover degree. The observed sensitive reaction of *Rubus fructicosus* cover to N deposition, was also predicted well ForSAFE-VEG. Finally, tree biomass was predicted well for the two Swiss test sites.

4 Model application: assessment of critical nitrogen loads

4.1 Application of the SMART2-MOVE model for Dutch vegetation types

To date, the MAGIC-GBMOVE model has not been applied in 'inverse mode' to estimate critical loads based on biodiversity targets. The SMART2-MOVE model, however, has been used in an inverse way to assess critical loads and target loads for major vegetation types in the Netherlands and to compare results with empirical critical N loads (Van Dobben et al., 2006). The empirical critical loads and the calculated critical loads correspond reasonably well (Table 2), i.e. their ranges usually overlap, although there is no significant correlation between the range midpoints of the two methods. The comparison of methods is slightly hampered by the fact that simulated values are determined for a very detailed typology (associations or sub-associations in the sense of Braun-Blanquet, 1964; Schaminée et al., 1995) while the empirical values were determined for so-called 'EUNIS' classes, where EUNIS stands for European Nature Information System (Davies and Moss, 2002). On average the midpoints of the empirical ranges are 3.4 kg N $ha^{-1}.y^{-1}$ below the simulated range midpoints, and this difference is nearly significant ($P \approx 0.07$). There are various reasons for the lower empirical values compared to the simulated ones, the most important probably being that the empirical critical loads tend to be based on the most sensitive components of an ecosystem, often under abiotic conditions that enhance sensitivity still further (cf. Achermann and Bobbink, 2003). In contrast, in the simulation approach all environmental conditions are usually set to 'mean' or 'most probable' values.

In the Netherlands recent attempts to integrate the empirical and the simulation method have made use of the virtues of both: the broad scientific acceptance (at least in Europe) of the empirical values, and the ecological detail of the simulated ones (Van Dobben and van Hinsberg, 2008). To this end, both the EUNIS typology and Schaminée et al.'s (1995) typology were translated into the European Habitat typology (Commission of the European Communities, 2003), and critical load ranges were determined according to both methods. For each Habitat type a unique critical load value was determined as the midpoint of the simulated range when this midpoint was with the empirical range; otherwise either the upper or the lower extreme of the empirical range was used. This method was developed in response to the policymaker's need for unique critical load values per Habitat type, to be used for the assessment of human activities in European 'Natura 2000' areas.

4.2 Application of the ForSAFE-VEG model at 16 Swedish forest sites

ForSAFE-VEG does not run in an inverse mode to derive critical loads. Actually, this is impossible, as the model is too complex to be used in an inverse way. Instead the “critical load” is determined to be passed at the time one can observe unwanted significant shifts in vegetation composition or abundance. This time is used for estimating the critical load, which is defined in this case as the deposition of N at the point in time of significant unwanted vegetation change. Actually, this value is dependent on the site history. To estimate the critical loads of N, a preliminary definition was adopted by which a 95% of the natural ground vegetation composition is preserved. This definition excludes the effect of other factors than N on the ground vegetation composition.

Critical-load estimates for 16 forested sites in Sweden thus derived are given in Table 3. The table presents the year when the acceptable change in ground vegetation composition occurred, and the value of the deposition at that year. A reduction from today's deposition values can then be deduced to lower the deposition to the historic value that preceded the undesired change in the ground-vegetation composition (Table 3). The estimates set the critical load as the deposition at the time the change occurs, probably leading to a slight overestimate of the critical load. Results show that all sites have significant exceedance, and in order to protect 95% of the area, a 90% reduction of present deposition is required, implying an average atmospheric deposition in southern Sweden of 1.1 kg $N.ha^{-1}.yr^{-1}$. Using a protective level to 50%, still 55% reduction in present deposition will be required, implying an average deposition in southern Sweden of 2.8 kg $N.ha^{-1}.yr^{-1}$.

5 Comparison and evaluation of the model chains

Comparison of model approaches and evaluation of the model chains

The modelling approaches described in this article consist of a combination of a biogeochemical model of nutrient (including N) behaviour in the soil, connected with a vegetation model predicting water, N and acidity impacts on biodiversity. The biogeochemical models discussed are SMART2 (either or not in connection with SUMO), MAGIC and ForSAFE. These models differ with respect to the included processes and management options (Table 4). Models of vegetation succession are included in ForSAFE, and in the model chain SMART2-SUMO-MOVE/NTM, with SUMO being a specific model for vegetation succession. Vegetation-succession models are intermediates between biogeochemical models and species-composition models since they simulate changes in elemental budgets and biomass distribution. Both SUMO and FORSAFE thus simulate the development of vegetation biomass and stocks of nutrient elements in relation to events such as fire, grazing, mowing or turf stripping. For example, grazing increases light availability and thus favours the growth of short-growing plants.

A comparison of the characteristics of the vegetation models and succession models predicting N impacts on biodiversity (MOVE/NTM, VEG and SUMO) is given in Table 5. The major strength of the SMART2/ MAGIC-GBMOVE/NTM approach is the empirical determination of the relation between plant species composition and soil moisture, nutrient availability and soil acidity. Furthermore, the relationships are based on species-response curves of a large number of higher and lower plant species (e.g. in MOVE about 900 plant species are covered (Wiertz et al., 1992) . By using vegetation relevés identified on the level of vegetation types, it was possible to estimate critical limits for these vegetation types based on percentile values of the Ellenberg indicators for nutrients and pH. Thus the strength of the resulting empirical niche models is that the weight of data reduces noise relative to species-environment relationships, at least in the Ellenberg domain. GBMOVE also includes climate and management besides N and acidity, again based on survey data, and thus incorporates the impacts of climate and management on plant species diversity and its modifying effect on critical loads.

The major weakness of the SMART2/MAGIC-GBMOVE/NTM approach is that a relationship is needed between Ellenberg indicators for N and moisture availability and acidity and measured values for these abiotic variables. Such calibration equations increase uncertainty because soil pH, soil C/N and soil moisture do not explain the total variation in mean Ellenberg scores. The greater the scatter about each regression line the more likely it is that predictions of mean Ellenberg values from soil measurements will differ from actual observations. The relationship with N indicators, such as N availability used in MOVE and soil C/N ratio used in GBMOVE, is rather weak, especially in high-fertility ecosystems. Especially the uncertainty in the Ellenberg indicator for nutrient availability is large and can be the main source of uncertainty in the end result (Schouwenberg et al., 2000; Wamelink et al., 2002). Ideally the use of Ellenberg indicator values should thus be avoided and response curves should be estimated from actual measurements of soil pH and N availability (Wamelink et al., 2005). Furthermore, it is not likely that the relations between Ellenberg indicator values and actual conditions derived for the Netherlands or the United Kingdom are valid for other countries. Therefore, to use these models in other countries it is necessary to analyse local vegetation relevés in order to assign critical site factors to ecosystems. Finally, the output of the model chains is the potential vegetation on a site, whereas the observed vegetation may differ due to time lag effects. The MOVE and GBMOVE models are based on empirical observations recorded at different times in the past 30-70 years across Dutch and British ecosystems while the resulting regression models assume equilibrium between species and environment and the niche of each species is thus static. Another weakness is the lack of feedback between vegetation change and the soil model, at least when SUMO is not included.

The major strength of ForSAFE-VEG is the mechanistic approach relating (many) abiotic parameters to plant species diversity including: (i) ground vegetation community competition, feedbacks from climate and from grazing animals and forest management and (ii) the mechanistic integration of the N cycle with process kinetics and feedbacks to the chemistry, organic matter decomposition and growth cycles. Furthermore, the model is field tested in Sweden. The major weakness of the ForSAFE-VEG approach is the high data demand. This holds specifically for the driving variables that consist mainly of descriptions of events, in particular the timing and intensity of grazing and other management events, which is also a limitation of SUMO. Furthermore, the complexity of the model makes interpretation of the results difficult, especially how different factors like acidity, nitrogen, management and climate change are all linked to biodiversity.

Use of nitrogen indicators in view of impacts on plant species occurrence

To connect theory on N dynamics in soils with models of plant species occurrence, a measure of N exposure, i.e. of plant-available N, is required. There are different measures to integrate N exposure into a single indicator. Some of these indicators give direct information on an N flux to the ecosystem, whereas other indicators only give indirect information based on correlations with fluxes. Most effects of N are due to an excess of N, either in the form of NH_4 or NO_x. In the soil models, a differentiation is made between NH_4 and NO_3 but not in the vegetation response models, even though there are indications that plants are more sensitive for NH_4 than for NO_3 (see e.g. Bobbink et al., 2003). The knowledge is, however, considered too limited to include in the vegetation effect models.

Major direct indicators of N availability are: (i) gross mineralization and nitrification rates, reflecting the internal N cycling and thus potentially the maximum inorganic-N pool available to plants in competition with microbial uptake and (ii) N deposition. The N-deposition flux accurately reflects the exposure of species with limited root systems, particularly bryophytes and lichens. In other systems, however, the transformation of N by soil microorganisms modifies plant exposure and N-deposition flux is thus not a good measure of exposure for plants rooting in soil. A better indicator is the sum of N deposition and N mineralization, as used in the SMART2-SUMO-MOVE approach, although the link to biodiversity is only expert-based, namely through the Ellenberg N indicator.

Indirect indicators, which are correlated with N availability, include (Rowe et al., 2005): (i) soil solution N concentration, (ii) plant tissue N concentration, (iii) soil C/N ratio and (iv) indicators based on the plant species assemblage. Apart from various other factors, use of a soil solution N concentration forms the basis of the ForSAFE-VEG model approach. The advantage of using this indicator is that the soluble N pool is immediately available to plants. However, soil solution only reflects the N in excess of uptake and leaching losses and thus may underestimate total N availability to plants. Concentrations are also very dynamic, both spatially and temporally, and single measurements of soil-solution N concentrations are thus of limited use. Measures integrated over time are thus more reliable indicators of N status. Furthermore, species differ in their ability to use different forms of soluble N, i.e. NO_3, NH_4 and DON and the ratio of ammonium to nitrate in solution may provide information relevant to species occurrence and also the potential for microbial uptake of nitrate. Measures of plant chemistry are not yet used in any of the models. A disadvantage of the use of tissue concentrations is that they vary considerably in time (seasonally), among species, plant parts, tissue age/phenological stage and with nutrient supply, grazing or other management (Rowe et al., 2005). Nevertheless, if these factors can be controlled (e.g. by sampling a standard part, from a single species or group, at a standard time of year), tissue concentrations of N and amino acids may be good indicators of N exposure and in principle could be outputs from biogeochemical models. The soil C/N ratio, as used in the GBMOVE approach, is not directly influencing plant response but represents a readily measurable proxy for important processes (e.g. mineralization or nitrification). In general, the relationship is weak, since total soil N is largely inactive, and it is not a good indicator of N availability (Tamm, 1991). Finally, Ellenberg indicators (Ellenberg et al., 1992) are used in GBMOVE, MOVE and NTM to describe assemblages of European vascular plants and bryophytes. Mean Ellenberg fertility scores have been shown to be reasonable indicators of soil N availability (Van Dobben, 1993), although the relationship usually shows large variation (Wamelink et al., 2002) and appears to correlate best with annual above-ground biomass production rather than soil nutrient status (Hill and Carey, 1997). However, in systems which are limited by other nutrients, e.g. phosphorous or potassium, this may not be the case.

In summary, plants do not respond to a single abiotic variable, and there are problems with all variables that could potentially be used as input to the vegetation models. Those considered most useful are direct indicators of N availability, such as N deposition plus N mineralization, followed by indirect indicators, such as soil solution N concentrations in the rooting zone, foliar N concentration or soil C/N ratio.

Model validation status

The validation status of the various models differs, specifically with respect to the comparisons between measured and modelled changes in plant species composition. In general, the biogeochemical models used (SMART2, MAGIC and ForSAFE) have a good validation status. For example, SMART2 has been validated on the German 'Solling site' and hundreds of intensively monitored forest plots (De Vries et al., 2003). Here we show that MAGIC is able to predict changes in observed N leaching and soil C/N in plot-scale N manipulation studies under different addition levels. The ForSAFE model also shows a good correlation between simulated and measured values of tree biomass, pools of soil organic C and N, soil pH and BC/Al ratios at 16 Swedish forest sites.

The validation status of the vegetation models is, however, much less advanced. Biomass growth of SUMO has until now only been validated on data collected at a grassland site, a heathland site and a forest site in the Netherlands and a grassland site in the UK. Regarding (GB)MOVE, a preliminary test was made whether GBMOVE in combination with MAGIC could reproduce the observed species composition in test plots, including an unimproved neutral grassland and a blanket bog. The comparisons of predicted species occurrence using measured soil C/N and pH versus mean Ellenberg scores indicate that the greatest uncertainty in model predictions, is due to the weak calibration relationships, especially those between soil C/N and mean Ellenberg N at high fertility. Current model development is therefore focusing on the establishment of a direct relation between N indicators and species composition. Finally, the ForSAFE-VEG model has until now been validated at 16 Swedish forest sites, by comparing simulated and measured values on the ground occupancy of 42 plant groups, and in Switzerland. Considering the sometimes poor correlations in all model applications, as described above, there is still a need for substantial model validation and model improvement. In this context, good model-testing data are urgently needed.

Reliability of biogeochemistry-biodiversity modelling approaches to assess critical loads

Up to now, only the SMART-MOVE model chain has been used to assess critical loads by using the model in an inverse way. Due to its complexity, t he ForSAFE-VEG model is more suited to predict response of plant species composition to environmental change than to predict critical loads. A major disadvantage is the fact that the critical N load is influenced by the N deposition history. The uncertainties in the assessed critical loads, using SMART-MOVE model chain, are specifically due to uncertainties in the calibration equations between abiotic conditions and Ellenberg scores, used to transfer information between soil models and plant-species models. A detailed uncertainty analysis, focusing on this aspect, was carried out by van Dobben et al. (2006). Results show that the uncertainty in critical loads per vegetation type is quite high (generally in a range of 15-40% of the average value), but the ranges of simulated and empirical values usually overlap, implying that the results are applicable for practical purposes. However, at the site level, uncertainty becomes very large and thus it is not yet possible to determine critical loads with any practical significance (Van Dobben et al., 2006). The uncertainties can only be reduced if more data become available on the abiotic response per species under field conditions, at least to N availability and soil pH, as described below.

Despite the uncertainties in the described model approaches, the approach provides a relevant addition to empirical critical-load estimates. First of all, even though empirical critical loads may be as good or even better than modelled critical loads at present, dynamic models allow us to explore impacts of future scenarios, where habitats may face completely novel configurations of multiple drivers, so-called “no analogue” states (Steffen et al., 2004). In these situations empirical critical loads that are based on reviews of contemporary and historical data become may be increasingly inappropriate and process based, detailed models are necessary. Critical loads can be interpreted as a means of identifying thresholds of ecosystem resilience (Gunderson, 2000). Estimating resilience thresholds using linked soil-vegetation models is therefore particularly appropriate, because of several key aspects associated with ecosystem responses to perturbation. Threshold changes and non-linear responses to disturbance, species invasion and changes in nutrient availability can result from the dynamic interplay between above- and below-ground biota and differences in the extent to which soil and vegetation store or more rapidly process and release excess nutrients or buffer pH changes (Evans et al., 2001; Craine et al., 2002). Hence resilience is a property related to the ability of soil and vegetation to buffer or to amplify the response to changing conditions. Quantifying resilience and proximity to thresholds of change then requires quantification of the dynamic relationships between ecosystem compartments (e.g. Suding et al., 2008). Policy interest also focuses on timescales for recovery; hence dynamic modelling is required to model the persistence of alternative stable states that can result when perturbations drive ecosystems into new domains of stability (Gunderson, 2000; Suding et al., 2004). While it may be possible to identify critical ecosystem state variables that are closely correlated with damage and loss of adaptive capacity, these may be above or below ground (e.g. Strengbom et al., 2001) once more emphasising the importance of jointly modelling soil and vegetation. Lastly, the critical-load approach increasingly recognises that resilience is not realistically thought of as a threshold situated along single or multiple, yet independent, abiotic axes. Multiple drivers and their interactions are important. For example, land use around a nature reserve can change the composition of the local species pool, increasing availability of species that could capitalise on changes in soil conditions driven by atmospheric pollutant deposition or conversely, reducing the availability of desirable species for recolonisation following recovery and remediation (Lindborg and Eriksson, 2004). Dynamic model development offers a way of incorporating other drivers as modifiers of the critical-load range and allowing multiple drivers to dynamically interact. The fully integrated, mechanistic ForSafe-VEG model is most appropriate to investigate such interactions.

Measurements of plant species response to environmental variables

Expert-based estimates of plant species responses to environmental variables form the basis of all models. Although the Ellenberg indicator system (Ellenberg et al., 1992) or its derivates (Diekmann, 2003) are the most frequently used systems for this, their uncertainty is quite large, it is sometimes unclear what the indicator values represent, and they may be biased (Ertsen et al., 1998; Schaffers and Sykora, 2000; Wamelink et al., 2002; Wamelink et al., 2003b; Witte and von Asmuth, 2003; Smart and Scott, 2004; Wamelink et al., 2004). Furthermore, it requires a translation from indicator values into soil chemical variables.

The most logical solution to avoid this highly uncertain step is to replace the Ellenberg indicator system by a system based on measurements (Wamelink et al., 2002). On a small scale this was carried out for France (Gégout et al., 2003) and the Netherlands (Wamelink et al., 2005). Results are promising, but Europe-wide and international data are needed, instead of only national data to ensure a wide application. Such data should consist of a vegetation description (relevés) and at least the following measured variables: geographical coordinates, soil acidity, nutrient status, and water availability. Based on this, plant species response per abiotic variable can be estimated, reviewed and tested on independent datasets. The hypothesis that plant species have different responses in different regions can be tested on the basis of such data.

Relevant extensions to the modelling approaches.

As mentioned above, in the SMART2/ MAGIC- (GB)MOVE/ NTM model chain, model output is the potential vegetation on a site, not accounting for time-lag effects. Priorities for future work on modelling N impacts on biodiversity in this model chain thus include: (i) inclusion of species-response curves based on combined field measurements of vegetation relevés and abiotic data and (ii) representation of lag times (e.g. due to species persistence, dispersal).

The present linked-model approaches further centre upon the impact of N deposition on existing species assemblages. However, changes in resource availability are also predicted to increase susceptibility to invasion by immigrants, some of which could possess traits associated with suppression of resident species and changes in nutrient cycling (Evans et al., 2001; Davis et al., 2005). If information about the composition of the regional species pool is available, then model application could be extended to estimate changes in habitat suitability for potential immigrants. This would amount to a risk assessment of the establishment of potential invaders. Incorporating the dynamic impact of invasive species on above and below-ground ecosystem compartments is possible in the presented model approaches but would require modification of the biomass growth models in SUMO and For-SAFE/VEG.

Conclusions on the potential of linked biogeochemistry-biodiversity modelling approaches

The following conclusions can be drawn regarding the modelling approaches presented here:

- Vegetation models, based on either large-scale vegetation surveys (MOVE/ GBMOVE, NTM) or mainly experimental data (VEG), have been developed and integrated with biogeochemical models, such as SMART2 (either or not in connection with SUMO), MAGIC and ForSAFE.
- There are large similarities between the models, particularly those based on survey data, but there are also several important differences, including: (i) use of different abiotic variables for N, such as N availability and Ellenberg N indicators in SMART2-MOVE/NTM, soil C/N ratio in GBMOVE and soil-solution N in ForSAFE, (ii) prediction of individual plant species (e.g. MOVE/GBMOVE) versus plant communities (NTM) and (iii) calibration based on different (national) soil and vegetation datasets.

- At their current level of development, most integrated models focus primarily on predicting the biodiversity impacts of different scenarios in terms of air-quality change and climate change.
- The model chains can also predict biodiversity-based critical loads or target loads. In doing this, the definition of reference conditions and damage thresholds for terrestrial biodiversity represents a major challenge. Although the definition of biodiversity targets is an issue for policy-makers, dynamic models can provide valuable information on realistic reference conditions and achievable recovery targets.
- The reliance on Ellenberg indicator values, used as a proxy for abiotic conditions in survey-based models such as SMART2-MOVE and MAGIC-GBMOVE, adds uncertainty to model predictions. However, Ellenberg values are likely to remain necessary in many areas due to the insufficient European coverage of combined vegetation and soil survey data.
- Models based on survey data have largest potential for country-wide mapping of critical loads in view of their limited data demands. This holds not only for Europe but also for other areas coping with elevated N deposition, such as North America and China. However, testing and adaptation of the linkage between vegetation and soil, using e.g. Ellenberg indicators is needed for other countries or ecosystems before large-scale applications can be made. Furthermore, results are quite reliable for vegetation types but at site level, the uncertainty in critical loads becomes too large to be of practical significance.
- In combination, both empirical and model-based critical loads, are powerful tools to assess a reliable value for defined ecosystems. An example of such a combination approach is presented by Van Dobben and van Hinsberg (2008).

Despite the various limitations mentioned, this overview shows that linked biogeochemistry-biodiversity models for N have great potential for applications to support policies to reduce N inputs. Apart from further model development, there is a need for further testing and validation of the models against long-term monitoring or long-term experimental datasets and against large-scale survey data. In this context, the continuation of existing programmes, where possible with improved integration of biotic and abiotic measurements, is essential to the future development of this work. Finally, there is a need for adaptation and upscaling of the models beyond the regions for which dose-response relationships have been parameterised, such as Mediterranean and Alpine regions, and Eastern Europe, based on a focused data collection combing vegetation descriptions with variables affecting the species diversity. A similar approach outside Europe would also allow the use of these models, both for predicting impacts of scenarios and assessing critical loads.

Acknowledgments

The preparation of this paper has been partially supported by the European Commission LIFE programme within the framework of the European Consortium for Modelling Air Pollution and Climate Strategies (EC4MACS) and the trust fund for the partial funding of effect-oriented activities under the Convention on Long-range Transboundary Air Pollution. We also thank the two anonymous reviewers for their detailed and helpful comments which largely improved the paper.

References

Achermann, B., Bobbink, R. (Eds.), 2003. Empirical critical loads for nitrogen: Expert workshop, Berne, 11-13 November 2002. Environmental Documentation 164, Swiss Agency for the Environment, Forests and Landscape.

Alkemade, J.R.M., Wiertz, J., Latour, J.B., 1996. Kalibratie van Ellenbergs milieu-indicatiegetallen aan werkelijk gemeten bodemfactoren. Rijksinstituut voor volksgezondheid en milieu, Bilthoven, The Netherlands. Rapport nr. 711901 016.

Amann, M., Cofala, J., Heyes, C., Klimont, Z., Schöpp, W., 1999. The RAINS model: A tool for assessing regional emission control strategies in Europe. Pollution Atmosphérique 20, 41-64.

Belyazid, S., 2006. Dynamic Modelling of biogeochemical processes in forest ecosystems. Doctoral Thesis. Reports in Ecology and Environmental Engineering 2006:1, Department of chemical Engineering, Lund University, Sweden.

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

Bobbink, R., Hornung, M., Roelofs, J.G.M., 1998. The effects of air-borne nitrogen pollutants on species diversity in natural and semi-natural European vegetation. Journal of Ecology 86, 717-738.

Bobbink, R., Ashmore, M., Braun, S., Flückiger, W., van den Wyngaert, I.J.J., 2003. Empirical nitrogen critical loads for natural and semi-natural ecosystems: 2002 update. In: Achermann, B., Bobbink, R. (Eds). Empirical critical loads for nitrogen. Swiss Agency for Environment, Forest and Landscape SAEFL, Berne, pp. 43-170.

Bobbink, R., Hicks, K., Spranger, T., Galloway, J., Alkemade, R., Ashmore, M., Bustamante, M., Davidson, E., Dentener, F., Emmett, B., Fenn, M., Gilliam, F., Nordin, A., de Vries, W., 2009. Atmospheric nitrogen deposition, biodiversity effects and critical loads - a synthesis. Ecological Applications (submitted).

Braun-Blanquet, W., 1964. Pflanzensoziologie. Grundzüge der Vegetationskunde. 3. Aufl. Springer, Wien/New York.

Clark, C.M., Tilman, D., 2008. Loss of plant species after chronic low-level nitrogen deposition to prairie grasslands. Nature 451, 712-715.

Commission of the European Communities, 2003. Interpretation manual of European Union habitats - EUR 25. DG Environment - Nature and Biodiversity.

Cosby, B.J., Ferrier, R.C., Jenkins, A., 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, 499-517.

Craine, J.M., Tilman, D., Wedin, D., Reich, P., Tjoelker, M., Knops, J., 2002. Functional traits, productivity and effects on nitrogen cycling of 33 grassland species. Functional Ecology 16, 563-574.

Davies, C.E., Moss, D., 2002. EUNIS Habitat Classification. Centre for Ecology and Hydrology. 2001 Work Programme, Final Report to the European Environment Agency European Topic Centre on Nature Protection and Biodiversity.

Davis, M.A., Thompson, K., Grime, J.P., 2005. Invasibility: the local mechanism driving community assembly and species diversity. Ecography 28, 696-704.

De Vries, W., Reinds, G.J., Posch, M., Sanz, M.J., Krause, G.H.M., Calatayud, V., Renaud, J.P., Dupouey, J.L., Sterba, H., Gundersen, P., Voogd, J.C.H., Vel, E.M., 2003. Intensive Monitoring of Forest Ecosystems in Europe. Technical Report 2003. UN/ECE and EC, Forest Intensive Monitoring Coordinating Institute, Geneva and Brussels.

De Vries, W., Kros, J., Reinds, G.J., Wamelink, W., Mol, J., van Dobben, H., Bobbink, R., Smart, S., Evans, C., Schlutow, A., Kraft, P., Belyazid, S., Sverdrup, H.U., van Hinsberg, A., Posch, M., Hettelingh, J.P., 2007. Developments in modelling critical nitrogen loads for terrestrial ecosystems in Europe. Alterra Wageningen UR, Wageningen, the Netherlands. Report 1382.

Diekmann, M., 2003. Species indicator values as an important tool in applied plant ecology - a review. Basic and Applied Ecology 4, 493-506.

EC, 2005. Thematic Strategy on Air Pollution. Communication from the Commission to the Council and the European Parliament. COM(2005) 446 final.

EEA, (European Environment Agency), 2007. Halting the loss of biodiversity by 2010: proposal for a first set of indicators to monitor progress in Europe. EEA Technical report 11/2007.

Elberse, W.T., van den Bergh, J.P., Dirven, J.P.G., 1983. Effects of use and mineral supply on the botanical composition and yield of old grassland on heavy-clay soil. Netherlands Journal of Agricultural Science 31, 63-88.

Ellenberg, H., Weber, H.E., Düll, R., Wirth, V., Werner, W., Paulißen, D., 1992. Zeigerwerte von Pflanzen in Mitteleuropa. Scripta Geobotanica 18, 1-248.

Emmett, B.A., 2007. Nitrogen saturation of terrestrial ecosystems: some recent findings and their implications for our conceptual framework. Water, Air, and Soil Pollution: Focus 7, 99-109.

Ertsen, A.C.D., Alkemade, J.R.M., Wassen, M.J., 1998. Calibrating Ellenberg indicator values for moisture, acidity, nutrient availability and salinity in the Netherlands. Plant Ecology 135, 113-124.

Evans, C.D., Caporn, S.J.M., Carroll, J.A., Pilkington, M.G., Wilson, D.B., Ray, N., Cresswell, N., 2006. Modelling nitrogen saturation and carbon accumulation in heathland soils under elevated nitrogen deposition. Environmental Pollution 143, 468-478.

Evans, R.D., Rimer, R., Sperry, L., Belnap, J., 2001. Exotic plant invasion alters nitrogen dynamics in an arid grassland. Ecological Applications 11, 1301-1310.

Gégout, J.C., Hervé, J.C., Houllier, F., Pierrat, J.C., 2003. Prediction of forest soil nutrient status using vegetation. Journal of Vegetation Science 14, 55-62.

Gunderson, L.H., 2000. Ecological resilience - in theory and application. Annual Review of Ecology and Systematics 31, 425-439.

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

Hettelingh, J.-P., Posch, M., Slootweg, J., Reinds, G.J., Spranger, T., Tarrasón, L., 2007. Critical loads and dynamic modelling to assess European areas at risk of acidification and eutrophication. Water, Air, & Soil Pollution: Focus 7, 379-384.

Hill, M.O., Carey, P.D., 1997. Prediction of yield in the Rothamsted Park Grass Experiment by Ellenberg indicator values. Journal of Vegetation Science 8, 579-586.

Janssen, P.H.M., Heuberger, P.S.C., 1995. Calibration of process-oriented models. Ecological Modelling 83, 55-66.

Jenkinson, D.S., Potts, J.M., Perry, J.N., Barnett, V., Coleman, K., Johnston, A.E., 1994. Trends in herbage yields over the last century on the Rothamsted long-term continuous hay experiment. Journal of Agricultural Science 122, 365-374.

Kros, J., 2002. Evaluation of biogeochemical models at local and regional scale. Wageningen, Alterra, 2002, Alterra scientific contributions, Wageningen, The Netherlands.

Kros, J., Reinds, G.J., de Vries, W., Latour, J.B., Bollen, M.J.S., 1995. Modelling abiotic site factors in response to atmospheric deposition and upward seepage. In: Schoute, J.F.T., Finke, P.A., Veeneklaas, F.R., Wolfert, H.P. (Eds). Scenario studies for the rural environment. Kluwer, Dordrecht, pp. 445-448.

Latour, J.B., Reiling, R., 1993. A multiple stress model for vegetation ('move'): a tool for scenario studies and standard-setting. Science of the Total Environment 134, 1513-1526.

Latour, J.B., Reiling, R., Slooff, W., 1994. Ecological standards for eutrophication and dessication: perspectives for a risk assessment. Water, Air, and Soil Pollution 78, 265-277.

Lindborg, R., Eriksson, O., 2004. Historical landscape connectivity affects present plant species diversity. Ecology 85, 1840-1845.

Ludwig, D., Walker, B., Holling, C.S., 1997. Sustainability, stability, and resilience. Conservation Ecology [online] 1, 7.

National Emission Ceilings Directive, 2001. Directive 2001/81/EC of the European Parliament and of the Council of 23 October 2001 on national emission ceilings for certain atmospheric pollutants.

Nilsson, J., 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.

Pakeman, R.J., Reid, C.L., Lennon, J.J., Kent, M., 2008. Possible interactions between environmental factors in determining species optima. Journal of Vegetation Science 19, 201-208.

Rowe, E.C., Moldan, F., Emmett, B.A., Evans, C.D., Hellsten, S., 2005. Model chains for assessing impacts of nitrogen on soils, waters and biodiversity: a review. Centre for Ecology and Hydrology. Contract report project no C02887 for DEFRA (UK) project no. Cpea 19.

Roy, D.B., Hill, M.O., Rothery, P., Bunce, R.G.H., 2000. Ecological indicator values of British species: an application of Gaussian logistic regression. Annales Botanici Fennici 37, 219-226.

Schaffers, A.P., Sykora, K.V., 2000. Reliability of Ellenberg indicator values for moisture, nitrogen and soil reaction: a comparison with field measurements. Journal of Vegetation Science 11, 225-244.

Schaminée, J.H.J., Westhoff, V., van Wirdum, G., 1989. Naar een nieuw overzicht van de plantengemeenschappen van Nederland. De Levende Natuur 90, 204-209.

Schaminée, J.H.J., Stortelder, A.H.F., Westhoff, V., 1995. De vegetatie van Nederland 1. inleiding tot de plantensociologie: grondslagen, methoden en toepassingen. Opulus Press, Uppsala.

Schlutow, A., Hübener, P., 2004. The BERN Model: Bioindication for Ecosystem Regeneration towards Natural conditions. Umwelt bundesamt. Research report 20085221.

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

Schouwenberg, E.P.A.G., Houweling, H., Jansen, M.J.W., Kros, J., Mol-Dijkstra, J.P., 2000. Uncertainty propagation in model chains: a case study in nature conservancy. Alterra, Green World Research, Wageningen. Alterra rapport 1.

Slootweg, J., Posch, M., Hettelingh, J.-P. (Eds.), 2007. Critical loads of nitrogen and dynamic modelling: CCE Progress Report 2007. Coordination Centre for Effects. MNP Report 500090001, Bilthoven, Netherlands.

Smart, S.M., Scott, W.A., 2004. Bias in Ellenberg indicator values - problems with detection of the effect of vegetation type. Journal of Vegetation Science 15, 843-846.

Smart, S.M., Scott, W.A., Whittaker, J., Evans, C., Emmett, B., 2003. Development and testing of linked soil-plant model systems. Pollution Umbrella Progress Report. May 2003.

Smart, S.M., Evans, C., Rowe, E., Wamelink, W., Roy, D., Preston, C., Hill, M., Rothery, P., Bullock, J., Moy, I., Scott, A., Emmett, B., Wright, S., 2005. Atmospheric nitrogen pollution impacts on biodiversity: Phase 1 - Model development and testing. Final report to Department of the Environment, Food and Rural Affairs, Joint Nature Conservation Committee and English Nature.

Spranger, T., Hettelingh, J.-P., Slootweg, J., Posch, M., 2008. Modelling and mapping long-term risks due to reactive nitrogen effects: An overview of LRTAP convention activities. Environmental Pollution 154, 482-487.

Steffen, W., Sanderson, A., Tyson, P.D., Jager, J., Matson, P.A., Moore III, B., Oldfield, F., Richardson, K., Schellnhuber, H.J., Turner, B.L., Wasson, R.J., 2004. Global change and the earth system: a planet under pressure. Springer-Verlag, Berlin.

Strengbom, J., Nordin, A., Näsholm, T., Ericson, L., 2001. Slow recovery of boreal forest ecosystem following decreased nitrogen input. Functional Ecology 15, 451-457.

Suding, K.N., Gross, K.L., Houseman, G.R., 2004. Alternative states and positive feedbacks in restoration ecology. Trends in Ecology & Evolution 19, 46-53.

Suding, K.N., Ashton, I.W., Bechtold, H., Bowman, W.D., Mobley, M.L., Winkleman, R., 2008. Plant and microbe contribution to community resilience in a directionally changing environment. Ecological monographs 78, 313-329.

Sutton, M.A., Cape, J.N., Rihm, B., Sheppard, L.J., Smith, R.I., Spranger, T., Fowler, D., 2003. The importance of accurate background atmospheric deposition estimates in setting critical loads for nitrogen. In: Achermann, B., Bobbink, R. (Eds). Empirical critical Loads for Nitrogen. Proceedings of an Expert Workshop in Berne, Switzerland, November 11-13. Swiss Agency for the Environment, Forests and Landscape (SAEFL) Environmental Documentation 164, pp. 231-257.

Sverdrup, H., Belyazid, S., Nihlgård, B., Ericson, L., 2007. Modelling change in ground vegetation response to acid and nitrogen pollution, climate change and forest management at in Sweden 1500–2100 A.D. Water, Air, and Soil Pollution: Focus 7, 163-179.

Tamm, C.O., 1991. Nitrogen in terrestrial ecosystems. Questions of productivity, vegetational changes and ecosystem stability. Springer Verlag, Berlin.

Tietema, A., Mol Dijkstra, J.P., Kros, J., de Vries, W., 2002. Dynamic nitrogen deposition thresholds during forest stand developments in a Douglas fir forest analysed with two nitrogen models SMART2 and MERLIN. Hydrology and Earth System Sciences 6, 375-382.

Tilman, D., 1987. Secondary succession and the pattern of plant dominance along experimental nitrogen gradients. Ecological monographs 57, 189-214.

Van Dobben, H.F., 1993. Vegetation as a Monitor for Deposition of Nitrogen and Acidity. PhD Thesis, University of Utrecht.

Van Dobben, H.F., van Hinsberg, A., 2008. Overzicht van kritische depositiewaarden voor stikstof, toegepast op habitattypen en Natura 2000 gebieden. Alterra, Wageningen. Alterra rapport 1654.

Van Dobben, H.F., van Hinsberg, A., Schouwenberg, E.P.A.G., Jansen, M.J.W., Mol-Dijkstra, J.P., Wieggers, H.J.J., Kros, J., de Vries, W., 2006. Simulation of critical loads for nitrogen for terrestrial plant communities in the Netherlands. Ecosystems 9, 32-45.

Wallman, P., Svenssson, M., Sverdrup, H., Belyazid, S., 2005. ForSAFE - an integrated process-oriented forest model for long-term sustainability assessments. Forest Ecology and Management 207, 19-36.

Wamelink, G.W.W., 2007. Simulation of vegetation dynamics as affected by nitrogen deposition. Pons & Looijen, Wageningen.

Wamelink, G.W.W., ter Braak, C.F.J., van Dobben, H.F., 2003a. Changes in large-scale patterns of plant biodiversity predicted from environmental economic scenarios. Landscape Ecology 18, 513-527.

Wamelink, G.W.W., van Dobben, H.F., Berendse, F., 2003b. Apparently we do need phytosociological classes to calibrate Ellenberg indicator values! Journal of Vegetation Science 14, 619-620.

Wamelink, G.W.W., Goedhart, P.W., van Dobben, H.F., 2004. Measurement errors and regression to the mean cannot explain bias in average Ellenberg indicator values. Journal of Vegetation Science 15, 847-851.

Wamelink, G.W.W., van Dobben, H.F., Berendse, F., 2008. Vegetation succession as affected by decreasing nitrogen deposition, soil characteristics and site management: a modelling approach. Forest Ecology and Management (accepted).

Wamelink, G.W.W., Joosten, V., van Dobben, H.F., Berendse, F., 2002. Validity of Ellenberg indicator values judged from physico-chemical field measurements. Journal of Vegetation Science 13, 269-278.

Wamelink, G.W.W., Goedhart, P.W., van Dobben, H.F., Berendse, F., 2005. Plant species as predictors of soil pH: replacing expert judgement with measurements. Journal of Vegetation Science 16, 461-470.

Wamelink, G.W.W., Goedhart, P.W., Frissel, J.Y., Wegman, R.M.A., Slim, P.A., van Dobben, H.F., 2007. Response curves for plant species and phytosociological classes. Alterra, Wageningen. Alterra report 1489.

Wiertz, J., van Dijk, J., Latour, J.B., 1992. De MOVE-vegetatie module: De kans op voorkomen van 700 plantesoorten als functie van vocht, pH, nutrienten en zout. Wageningen Bilthoven. RIVM report 711901006 / IBN-DLO report no. IBN 92/24.

Witte, J.P.M., von Asmuth, J.R., 2003. Do we really need phytosociological classes to calibrate Ellenberg indicator values? Journal of Vegetation Science 14, 615-618.

Table 1. Comparison of MAGIC-simulated and measured C/N ratios at Park Grass experimental site at Rothamsted, UK, using 100% and 50% of the approximate estimate for nitrogen offtake in hay.

Year	Measured C/N ratio (g C g^{-1} N)	Simulated C/N ratio (g C g^{-1} N)	
		100% N offtake	50% N offtake
1876	12.8	9.7	12.3
1959	12.3	11.2	12.4
2000	12.4	12.2	12.0

Table 2 Empirical (Achermann and Bobbink, 2003) and average modelled (using SMART2) critical N loads (CL) and target N loads for 2030 and 2100 (in $kg.ha^{-1}.yr^{-1}$) for European Nature Information System (EUNIS) classes.

EUNIS Class	Empirical CL	Modelled CL [1]	Modelled target load (2030)	Modelled target load (2100)
Forest (G)	10-20	16.8 (12.9 - 18.2)	8.4 (7.4 – 16.8)	14.0 (13.0 – 16.8)
Raised bogs (D1)	5-10	6.1 (6.1 – 6.1)	4.5 (3.8 – 6.1)	5.7 (5.0 – 6.1)
Salt marsh (A2.64/65) [2]	30-40	30.0 (30.0 – 34.1)	33.7 (29.9 – 33.9)	34.1 (34.0 – 34.1)
Dry and neutral grasslands (E1.7) [2]	10-20	8.0 (8.0 – 8.0)	1.4 (0.2 – 3.1)	7.9 (4.4 – 10.9)
Semi-dry calcareous grasslands (E1.26) [3]	15-25	12.4 (12.4 – 12.4)	-	-
Moist and wet oligotrophic grasslands (E3.5)	10-20	12.6 (12.6 – 12.6)	1.4 (0.5 – 6.7)	1.2 (0.4 – 12.6)
Coastal dune heaths (B1.5) [4]	10-20	15.5 (14.4 – 15.5)	3.3 (3.1 – 5.0)	12.9 (12.6 – 12.9)
Dry heaths (F4.2)	10-20	11.2 (9.4 – 17.1)	19.8 (17.0 – 21.7)	19.8 (18.5 – 21.7)

[1] Values in bracket refer to the 5 and 95 percentile

[2] Consists of a few nature types only with similar requirements regarding N status, leading to very similar values for the various percentiles.

[3] Consists of one nature type only, so all critical nutrient N load computations yield equal results

[4] Consists of a few receptors only, leading to strongly skewed distribution

Table 3 Preliminary critical loads for N based on preservation of the ground vegetation biodiversity according to the set conditions for non-effect for 16 Swedish study sites

Site	Time of vegetation response	Critical load deposition $kg.ha^{-1}.yr^{-1}$	Present deposition $kg.ha^{-1}.yr^{-1}$	Excess deposition $kg.ha^{-1}.yr^{-1}$	Required deposition reduction %
Högbränna	1910	1.1	1.5	0.4	27
Brattfors	1890	0.9	2.0	1.1	55
Storulvsjön	1925	2.0	3.5	1.5	43
Högskogen	1928	4.8	7.9	3.2	40
Örlingen	1910	3.6	8.5	3.9	52
Edeby	1918	3.9	7.8	3.9	50
Blåbärskullen	1880	1.6	8.5	6.9	81
Höka	1920	4.0	8.9	4.9	55
Hensbacka	1922	7.4	18.0	10.6	59
Söstared	1868	2.1	20.0	17.9	89
Gynge	1870	2.8	8.3	5.5	66
Fagerhult	1915	3.7	7.5	3.8	51
Bullsäng	1870	2.1	15.0	12.9	86
Timrilt	1889	3.6	23.0	19.4	84
Vång	1910	7.8	17.0	9.2	54
Västra Torup	1866	2.4	27.0	24.6	91

Table 4 Key processes represented in the biogeochemical models used in model chains for assessing impacts of nitrogen on biodiversity. SMART stands for simulation model for acidification's regional trends, SUMO for succession model, MAGIC for Model for Acidification of ground water in catchments and SAFE for Soil acidification in forest ecosystems ● = modelled dynamically; ○ = modelled indirectly or in a simplified way; k = included as constant or fitted term; - = not modelled.

Process	SMART2[1]	SMART2/SUMO[1]	MAGIC[1]	ForSAFE-VEG
Photosynthesis / tree growth	k	●	-	●
Competition / succession	-	●	-	●
Plant N uptake	●	●	○	●
Symbiotic nitrogen fixation	k	●	-	k
Litterfall	●	●	○	●
Decomposition	●	●	○	●
N mineralization	●	●	○	●
Nitrification	●	●	○	●
Denitrification	●	●	○	○
Inorganic N leaching	●	●	●	●
Organic N leaching	-	-	○	○
N immobilization	●	●	●	●
Soil carbon dynamics	●	●	○	●
SOM pools with different reactivity	●	●	-	●
Major ion chemistry/acidity	●	●	●	●
Base cation weathering	○	●	k	○
Grazing	-	●	○	●
Fire	-	●	○	●
Sod cutting	-	●	○	-
Tree felling	-	●	○	●

[1)] The combination of the vegetation model (GB)MOVE or NTM with either SMART2 or MAGIC does not include any additional process compared to the use of the individual models

Table 5 Comparison of the characteristics of MOVE/NTM, GBMOVE, VEG and SUMO where MOVE stands for Model of Vegetation and GBMOVE for the Great Britain version of MOVE, NTM for Nature technical model, VEG for vegetation model and SUMO for succession model .

Characteristic	MOVE/NTM	GBMOVE	VEG	SUMO
Methodology				
Relation between abiotic conditions and species diversity	Statistical (Logistic & Splines)	Multiple Logistic Regression	Mechanistic competition model (growth functions)	Mechanistic competition model (growth functions)
Abiotic conditions as single stressors, combined single stressors or multistressors	Multistress (water content, pH, N-availability)	Multistress (% soil moisture, pH, C/N ratio, cover-weighted canopy height)	Combined single stressors (water content, pH, N, P, light, temperature, grazing)	Combined single stressors (water content, pH, N, P, light, grazing, management)
Crucial factor for critical load calculations	pH & N-availability	pH & N-availability	N-concentration in soil solution & pH & Al-concentration & plant competition (light & nutrients)	N-availability & pH & plant competition (light & nutrients)
Link between environment and biodiversity	**(In)direct** (correlations between mean Ellenberg-indicator values of plant releveés and abiotic measurements)	**(In)direct** (correlations between mean Ellenberg-indicator values of plant releveés and abiotic measurements)	**Indirect**: model-outcome as a result of differences in species-specific growth functions	**Indirect**: model-outcome as a result of differences in plant type-specific growth functions
Applicability				
Link with biodiversity targets	**Direct** (via protected species or protected habitat types of EU-	**Direct** (via indicator species designated by statutory agencies by	Indirect by calculating a relevant indicator.	Indirect: Only possible after link with a species model

Characteristic	MOVE/NTM	GBMOVE	VEG	SUMO
	habitat directive)	habitat)	-	-
Link with SMB for CL-calculations	**Possible** via acceptable N-leaching			
Link with dynamic soil models for critical load and target-load calculations	**Operational** (SMART2)	Operational (MAGIC)	Link with dynamic models is operational (SAFE) and used for target-load calculations	Link with dynamic models is operational (SMART2) but can't be used for target-load calculations
Countries for which the method is developed and tested/used	Methodology developed for the Netherlands, tested in UK, Denmark	Methodology developed for, and tested in, the UK	Methodology developed and tested in Nordic countries	Methodology developed for Netherlands, tested in UK.

Figure legends

Figure 1 Method to predict plant species composition as a function of atmospheric deposition (top) and to calculate critical loads for nitrogen and acid deposition (bottom). The model abbreviations are explained in Table 4 and 5

Figure 2 Simulated and observed organic soil C/N ratio under ambient N deposition and three levels of long-term NH_4NO_3 *addition at two heathland experimental sites. Vertical line indicates start of experiment.*

Figure 3 Measured and simulated biomass harvest for a mown grassland site near Wageningen in the Netherlands (left) and for an experimental grassland site at Rothamstead in the UK (right).

Figure 4 Percentage of species correctly predicted in the three Park Grass control plots (left) and Moorhouse (right) based predictions by MAGIC+GBMOVE versus predictions based on observed mean Ellenberg scores only, as input to GBMOVE.

Figure 5 Predicted versus observed change in individual species in the Moorhouse Hard Hills control plots. Predicted change is the slope coefficient of a linear regression on occurrence probabilities predicted by MAGIC+GBMOVE for each year between 1973 and 2001. Observed change is the slope coefficient of a linear regression on % frequency in sample plots in each survey year. Pearson correlation coefficient = 0.568, p=0.002.

Figure 6 Modelled and measured pH values through the soil profile at 16 Swedish study sites

Figure 7 Modelled and measured ground vegetation occupancy of different plant groups at 2 Swedish study sites, i.e. Brattfors (left) and Svartberget (right). The included line is the 1:1 relation..

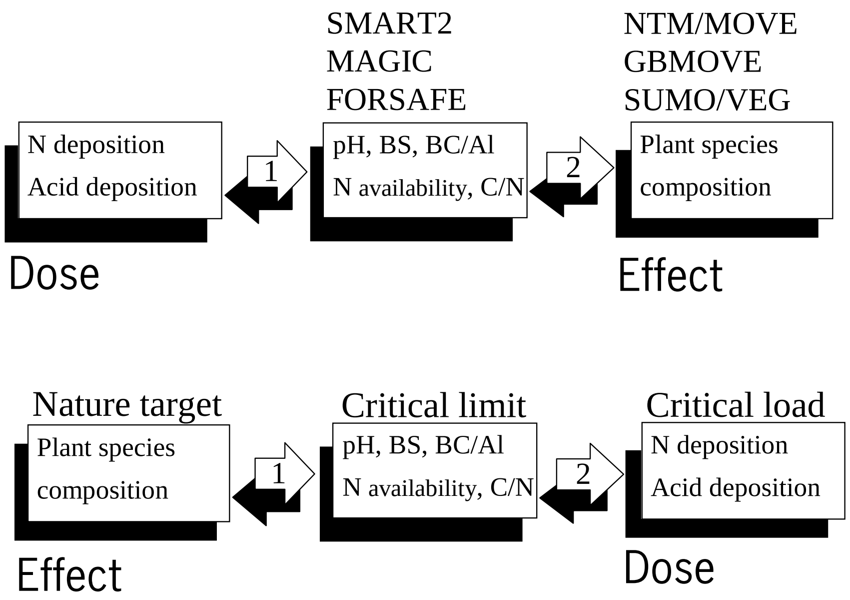

Figure 1 Method to predict plant species composition as a function of atmospheric deposition (top) and to calculate critical loads for nitrogen and acid deposition (bottom). The model abbreviations are explained in Table 4 and 5

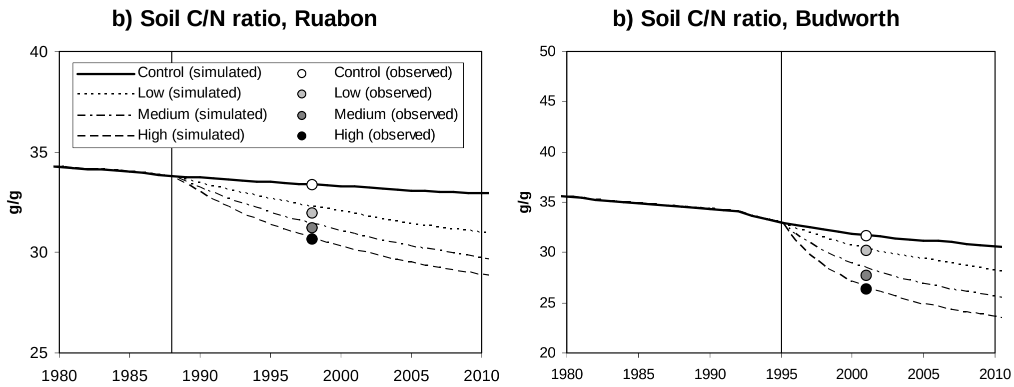

Figure 2 Simulated and observed organic soil C/N ratio under ambient N deposition and three levels of long-term NH_4NO_3 addition at two heathland experimental sites. Vertical line indicates start of experiment.

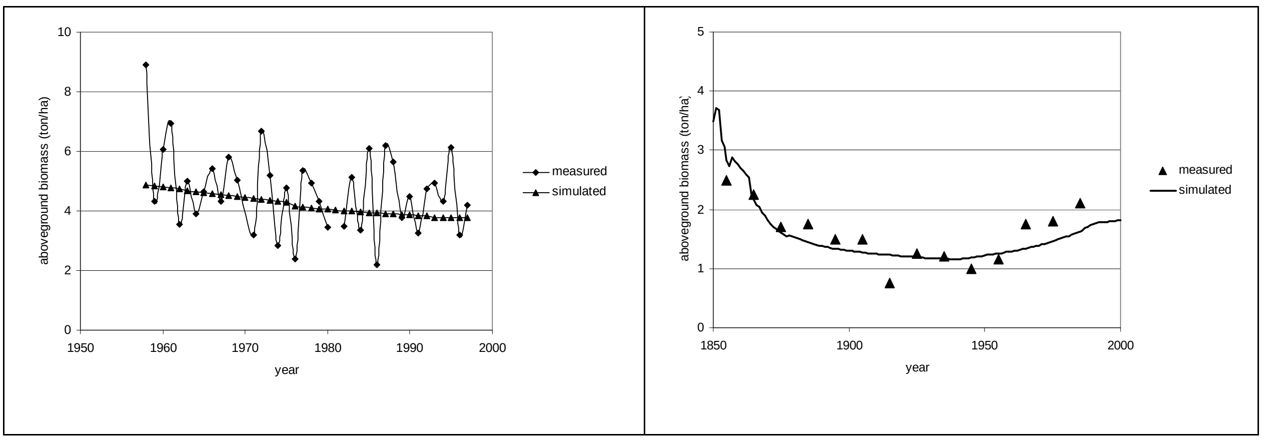

Figure 3 Measured and simulated biomass harvest for a mown grassland site near Wageningen in the Netherlands (left) and for an experimental grassland site at Rothamstead in the UK (right).

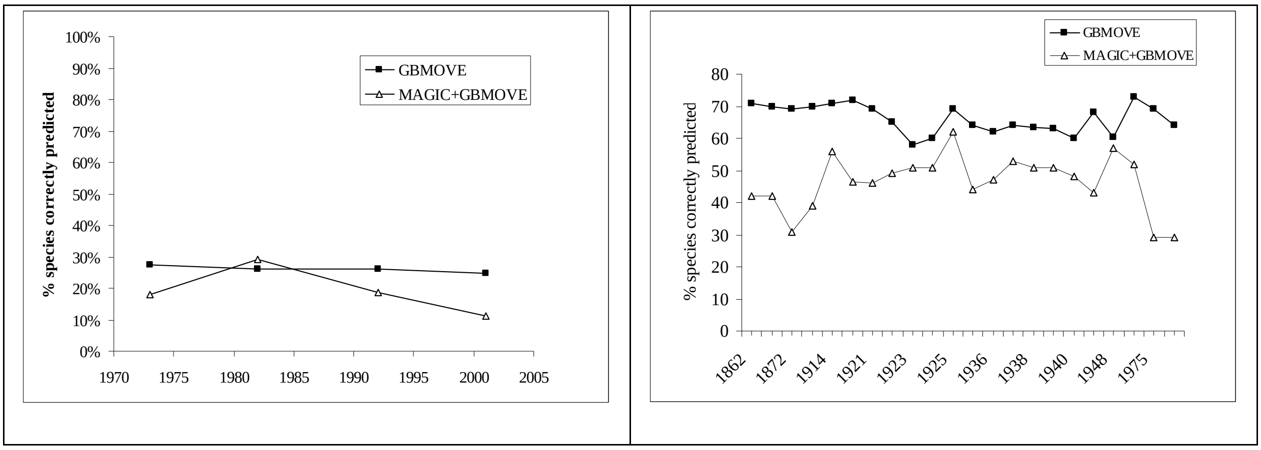

Figure 4 Percentage of species correctly predicted in the three Park Grass control plots (left) and Moorhouse (right) based predictions by MAGIC+GBMOVE versus predictions based on observed mean Ellenberg scores only, as input to GBMOVE.

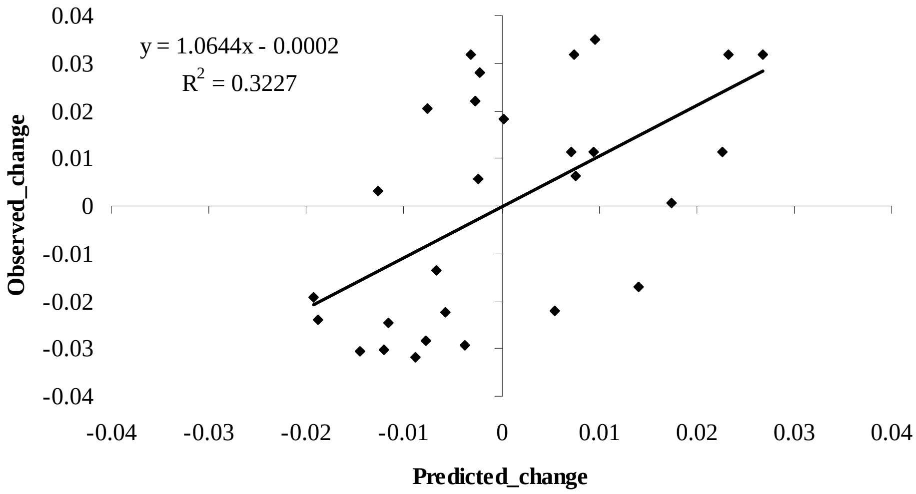

Figure 5 Predicted versus observed change in individual species in the Moorhouse Hard Hills control plots. Predicted change is the slope coefficient of a linear regression on occurrence probabilities predicted by MAGIC+GBMOVE for each year between 1973 and 2001. Observed change is the slope coefficient of a linear regression on % frequency in sample plots in each survey year. Pearson correlation coefficient = 0.568, p=0.002.

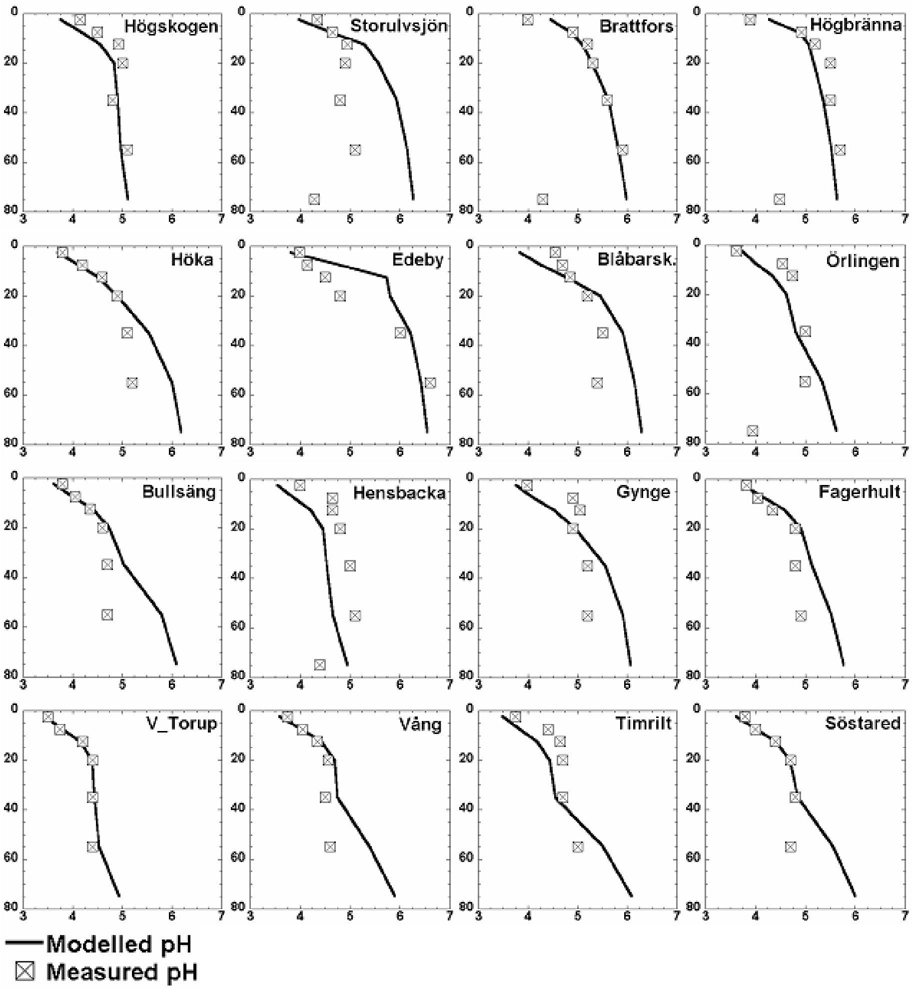

Figure 6 Modelled and measured pH values through the soil profile at 16 Swedish study sites

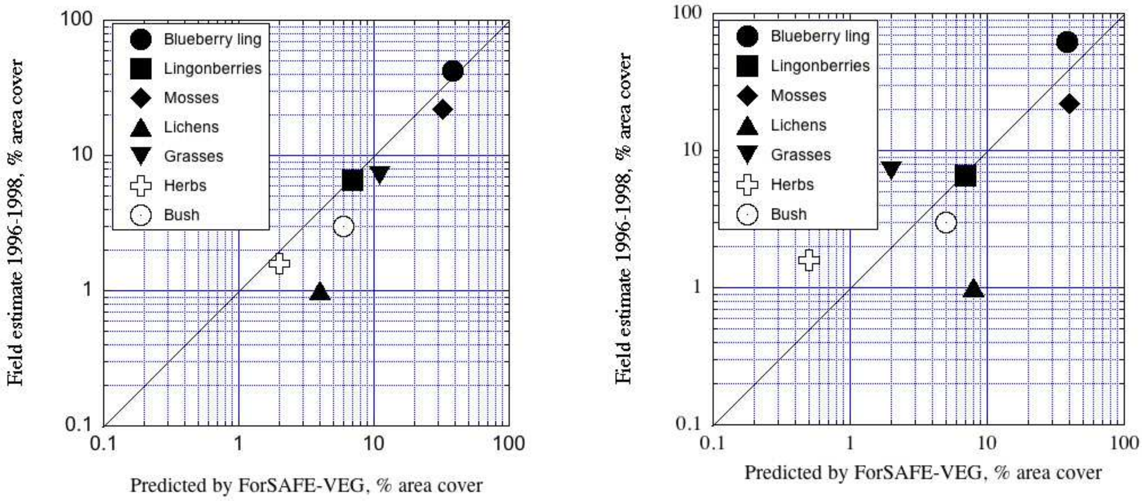

Figure 7 Modelled and measured ground vegetation occupancy of different plant groups at 2 Swedish study sites, i.e. Brattfors (left) and Svartberget (right). The included line is the 1:1 relation.