Biogeosciences Discuss., 7, 2313–2360, 2010
www.biogeosciences-discuss.net/7/2313/2010/

Biogeosciences Discussions

This discussion paper is/has been under review for the journal Biogeosciences (BG).
Please refer to the corresponding final paper in BG if available.

The use of machine learning algorithms to design a generalized simplified denitrification model

F. Oehler, J. C. Rutherford, and G. Coco

National Institute of Water & Atmospheric Research P. O. Box 11115, Hamilton, New Zealand

Received: 11 February 2010 – Accepted: 25 March 2010 – Published: 1 April 2010

Correspondence to: F. Oehler (f.oehler@niwa.co.nz)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

We designed generalized simplified models using machine learning algorithms (ML) to assess denitrification at the catchment scale. In particular, we designed an artificial neural network (ANN) to simulate total nitrogen emissions from the denitrification process. Boosted regression trees (BRT, another ML) was also used to analyse the relationships and the relative influences of different input variables towards total denitrification. To calibrate the ANN and BRT models, we used a large database obtained by collating datasets from the literature. We developed a simple methodology to give confidence intervals for the calibration and validation process. Both ML algorithms clearly outperformed a commonly used simplified model of nitrogen emissions, NEMIS. NEMIS is based on denitrification potential, temperature, soil water content and nitrate concentration. The ML models used soil organic matter % in place of a denitrification potential and pH as a fifth input variable. The BRT analysis reaffirms the importance of temperature, soil water content and nitrate concentration. Generality of the ANN model may also be improved if pH is used to differentiate between soil types. Further improvements in model performance can be achieved by lessening dataset effects.

1 Introduction

The increase of agricultural nitrogen (N) inputs favors the emission of nitrous oxide (N_2O) through nitrification and denitrification. N_2O is a well-known greenhouse gas (IPCC, 2006) involved in the ozone layer destruction (Cicerone, 1987) and soils are the main source of atmospheric N_2O (Mosier and Kroeze, 2000). Indirect emissions of N gasses (i.e., occurring after the applied nitrogen has been transformed or transferred out of the field) are still a major source of uncertainty despite their role on climate change (Crutzen et al., 2007; Mosier and Kroeze, 2000; Nevison, 2000).

Heterotrophic denitrification is the biological reduction of nitrate (NO_3^-) or nitrite

BGD

7, 2313–2360, 2010

(NO_2^-) into N_2O and di-nitrogen (N_2) in absence of oxygen (O_2). The process is influenced by many factors, is highly variable over space and time, and is thus difficult to assess at the catchment level. The difference between annual nitrogen flow measured at the catchment outlet and the nitrogen surplus do not provide a reliable estimate of the denitrification at the catchment scale, because of temporary storage processes of nitrogen in the soil, vadose zone or groundwater (Basset-Mens et al., 2006; Molenat and Gascuel-Odoux, 2002; Ruiz et al., 2002). Furthermore, losses by gaseous emission through denitrification are not evenly distributed over the catchment area since they are particularly higher in the riparian zone (Fisher and Acreman, 2004; Haag and Kaupenjohann, 2001; Martin et al., 1999; Oehler et al., 2007; Sebilo et al., 2003). As a result, it is still problematic to up-scale measured emissions to a larger, landscape scale which is the most relevant to assess the impact of agriculture practices and their management.

Models can be used to take into account these processes and the spatial and temporal variability of the driving factors. Many models integrate a more or less complex denitrification module (e.g. GLEAMS (Knisel, 1993), DNDC (Li et al., 1992), SWAT (Arnold and Fohrer, 2005), TNT2 (Beaujouan et al., 2001)) to simulate NO_3^- fluxes at the agricultural field or catchment scale. These models are often coupled to socio-economic models to provide an integrated N management tool (Leip et al., 2008; Turpin et al., 2005). Different approaches have been developed for denitrification modelling. These approaches range from (1) simplified process models (e.g. NEMIS, (Henault and Germon, 2000), (2) to soil structural models (e.g. Vinten et al., 1996), and (3) to microbial growth models (e.g. DNDC).

The accuracy of measurement techniques still needs to be improved especially to assess long term emissions, and this is particularly the case for upland terrestrial areas (Groffman et al., 2006). Our long term goal is developing a model of denitrification at the catchment scale that also addresses the significant emissions from upland areas (Oehler et al., 2007). To achieve this aim, we turned towards simplified modelling approaches also because (1) mechanistic models are developed and validated for ho-

mogeneous and simple medium, which is not necessarily appropriate at the catchment scale (Beven, 1993), (2) either the accuracy of measured emissions is poor and/or sampling is too scarce (Groffman et al., 2006), (3) simplified models need inputs that can be obtained either from relatively simple field measurements or directly from simulation models.

Simplified models have already been used in many studies and, for example, Heinen (2006b) found as many as 59 simplified models in the literature. He also analyzed the performance of the simplified model NEMIS on an extended data set (Heinen, 2006a). Following the same procedure as Heinen (2006a), NEMIS was also calibrated on another large data set (Oehler et al., 2009). Because of either measurements or modelling shortcomings, results were not fully satisfactory for a generalized use at the catchment scale. Moreover, there is a need to simulate also N_2O emissions from denitrification at the catchment scale, especially as stakeholders are looking toward the use of wetlands as nitrogen attenuation tools. Finally, it is worth reminding that in simplified approaches the global N emissions are a key parameter to estimate N_2O (using the N_2O/N_2 ratio, Henault et al., 2005; Lehuger et al., 2009) and so their estimate needs to be more robust and accurate.

Simplified models can be developed using a data-driven approach and so using a broad family of algorithms loosely defined in the literature as "machine learning" (ML). Since the "universal approximator demonstration" at the end of the 1980's (Cybenko, 1989; Hornik et al., 1989; Irie and Miyake, 1988), artificial neural networks (ANN) have probably become the most typical machine learning algorithm and have been used in many different fields like physics, chemistry, medicine, ecology and hydrology (Cote et al., 1995; Faraggi and Simon, 1995; Kralisch et al., 2003; Lek et al., 1999; Lischeid, 2001; Smits et al., 1992; Suen and Eheart, 2003; Telszewski et al., 2009). Artificial Neural Networks have been widely used to model complex non-linear relationships, particularly when the functional form of the relations between the variables involved is unknown.

Boosted Regression Trees (BRT) is a relatively new ML algorithm (partly originating from Schapire, 2003) characterized by strong predictive performance and that can give powerful insights of the variable relationships (Elith et al., 2008). However BRT are complex models in their representation (from a few hundred to few thousands of trees), and can be difficult to export from the ML environment to a separate and independent model. BRT do not entirely fit into the "simplified models", but they can efficiently describe the relationship between input variables and a system response.

In this study we designed a generalized simplified model based on ANN to simulate N emissions from the denitrification process at the field scale. To achieve this objective, we:

- assembled a large database from literature datasets;
- analysed the variable relationships and the relative influences of input variables toward total denitrification using BRT. This guided the choice of the input variables for the final ANN model;
- tested different input variables and ANN models;
- compared the predictive performance with NEMIS;
- explored the sensitivity of simulated denitrification rates to input factor variations.

2 Methods

2.1 Database and input factors

To calibrate (train) the ANN and BRT models, a large database is needed. The dasabase was built with datasets easily extractable from the literature (Cosandey et al., 2003; Henault and Germon, 2000; Luo et al., 1999 and Oehler et al., 2007). Denitrification rate (*Da*) rates were measured using the acetylene (C_2H_2) blockage technique (Ryden et al., 1987; Yoshinari et al., 1977). The soil denitrifying potential was

either a long term (days, termed Denitrification Potential (LDP) as in Henault and Germon (2000)) or a short term (hours, termed Denitrifying Enzyme Activity (DEA) as in Cosandey et al., 2003; Luo et al., 1999; Oehler et al., 2007) measure. The main differences in measurement techniques are summarized in Table 1. For the dataset of Luo et al. (1999), soil temperature (T) was estimated from national statistics. All the studies were carried out in temperate regions (France, Switzerland and New Zealand) and 40% of the measurements were in riparian or wetland areas. Our final database has 449 records: 58 from Cosandey et al. (2003); 39 from Henault and Germon (2000); 99 from Luo et al. (1999) and 253 from Oehler et al. (2007). Soil types included: cultivated and uncultivated silt loam and silty clay loam soils with OM 4–7% and pH 5–6.5; cultivated silt loam with OM 1% and pH 7.1; grazed riparian grasslands on silty clay and silty sand soils with OM 2.4–12.2% and pH 6.8–8; and pasture on silt loam with OM 6% and pH 6. All the *Da* measurements were done using a static chamber technique. The main denitrification measurements issues with the C_2H_2 blockage technique are:

- the diffusion of C_2H_2 into the soil;
- C_2H_2 can be used as a carbon source by micro-organism after a long time;
- C_2H_2 inhibits also the mineralisation, hence limiting its applicability to moderate to high [NO_3^-];
- the diffusion of O_2 into the soil samples if they are disturbed;
- low gas emission dynamic compared to the sensor sensitivity, compensated by the length of incubation time;
- soils are heterogeneous substrates.

All of this can lead to a very large measurement variability, especially for low-drainage soils (Groffman et al., 2006). Henault and Germon (2000) and Cosandey et al. (2003) measurements may be the less variable (i.e. all the measurements at 20 °C). The

DEA measurement methods are similar in their adding of substrate quantities, mixing procedures and incubation time.

Figure 1 shows the distribution of the response (*Da* rates) and independent variables (NO_3^-, *T*, WFPS, OM, *Db*, SD and pH). *Da* and NO_3^- show distributions with very long upper tails. This could impact the variance of the predictive performance computed on independent validation subset. Apart from the NEMIS model, the ANN and BRT are robust to outliers and the calibration process should not be influenced. Table 2 shows the univariate linear correlations (Pearson *r*) between variables. The r values (notice r^2 will be even smaller) show that we are not in a simple case with one or two dominant factors and linear relationships. *Da* is weakly correlated with WFPS, *T*, pH and OM, but *Da* is not correlated with NO_3^-. This does not mean that NO_3^- is not involved in the denitrification process. First, NO_3^- may have been present (mostly) in excess so that it did not limit *Da*. Also, NO_3^- soil concentration may have been a poor indicator of the rate of supply of NO_3^- (e.g., by advection, diffusion or nitrification, if its inhibition by C_2H_2 was not total) to denitrification micro sites, which is what determines *Da*. Second, the measurements from this collated dataset are far from genuinely and equally representing the different studied systems: there is a mixture of field (uncontrolled) and laboratory (some of the parameters are controlled or manipulated, like *T* or $[NO_3^-]$) measurements, with different sampling strategies and measurement technique variants (disturbed or undisturbed soil cores). There are also correlations between the input variables. Notably OM and NO_3^- are weakly correlated ($r = 0.27$) as a result of a small number of high NO_3^-/low OM points in Henault and Germon (2000) and low NO_3^-/high OM points in Cosandey et al. (2003). *Db* is weakly correlated with OM ($r = 0.42$), pH ($r = 0.63$) and SD ($r = 0.57$), and pH is weakly correlated with *T* ($r = 0.42$). This might be again the result of a sampling bias (e.g., the highest *Db* and pH soils were measured at 20 °C in Henault and Germon (2000) and Cosandey et al. (2003), and the soils with highest *Db* were also those with the highest pH). Besides conjectures, at this stage we can only suggest that the variation of *Da* is due to more than one factor and probably in a non-linear way. As we are in the presence of a multi-variable non-linear

BGD
7, 2313–2360, 2010

problem, and r values are a measure of the strength of linear dependence between two variables, they are not really meaningful. However these show that the dataset might be unbalanced (r values between known factors not close to zero), and therefore that some variable-space regions might not be equally represented.

Previous modelling has identified the most important factors influencing denitrification rate (Da) to be: T, water filled pore space (WFPS), nitrate concentration ($[NO_3^-]$), and the soil denitrifying potential. The last factor can be either a long term (days, like LDP) or a short term (hours, like DEA) measurement. The short term denitrification potential metrics (DEA) are most commonly used. Although successfully used as a denitrification indicator (Heinen, 2006a), DEA techniques are varied and have an imprecise relationship to Da (Oehler et al., 2007; Simek et al., 2000).

In addition to the controlling factors outlined above (i.e. Temperature, WFPS, NO_3^- and DEA), we tested the following factors:

- Organic matter % (OM). OM could be a useful surrogate for soil LDP which is correlated to soil physical characteristics more than DEA is (Simek et al., 2000). Some models use OM to compute a LDP (Hansen et al., 1991; Johnsson et al., 1987) which has been suggested to be more appropriate than DEA for modelling purposes (Henault and Germon, 2000).

- Bulk density (Db). Petersen et al. (2008) argued that gas diffusivity is affected by Db and influences O_2 concentration. This in turn strongly influences denitrification rates. Hence, it may be a better estimator of O_2 concentration than WFPS. Moldrup et al. (2005) modelled gas diffusivity using soil porosity and pore size distribution which are correlated with Db, WFPS and OM.

- pH. Soil pH is non-neutral toward denitrification with multiple direct and indirect effects (Simek and Cooper, 2002). Because of the use of the acetylene blockage technique for measuring denitrification, the influence of pH on nitrification rate and hence the supply of NO_3^- (Cheng et al., 2004; Hwang and Hanaki, 2000) is not taken into account.

– Soil depth (SD). Soil depth affects connectivity to the surface and hence influences aeration, O_2 concentration and ultimately denitrification rate.

Three records from Cosandey et al. (2003) were discarded because they unbalanced the validation process, and strongly influenced the NEMIS model calibration.

2.2 Artificial neural networks

The first mathematical representation of a neuron was introduced by McCulloch and Pitts (1943) with the perceptron. Each neuron receives input vectors (X), performs a weighted sum (α), and through an activation (also called transfer) function (G) (which may be linear or non-linear) produces a result (Y) in the form:

$$Y = G(WX + b) \quad (1)$$

where $W = (w_{i,1}, w_{i,2}, \ldots, w_{i,n})$ are the neurons weights, $X = (x_1, x_2, \ldots, x_N)$ are the vector inputs of neuron i, b is the neuron bias. $\alpha = (WX + b)$ is the input weighted sum (also called potential of neuron i) and G is the activation function. The classic non-linear activation function used is the sigmoid function:

$$G(\alpha) = (1 + e^{-\alpha})^{-1} \quad (2)$$

One or more neurons form a layer. In our study we used the common feed-forward ANN structure deriving from the perceptron, also called "multi-layer perceptron". The first neurons are forming the input layer, the lasts are forming the output layer, the others are forming one or more hidden layers (Hagan et al., 1996). The standard notation used throughout this work is [3:4:1], meaning 3 input nodes, 4 hidden and 1 output nodes (5 neurons). The number of input variables necessary for predicting the desired output variable determines the number of input nodes. The optimum number of hidden nodes and hidden layers is dependent on the complexity of the modelling problem. During training, patterns of input and corresponding output pairs are presented to the

ANN, and the learning algorithm iteratively adjusts the values of connection weights within the ANN structure. It is desirable to attain the required level of accuracy with the simplest possible ANN structure (i.e., the fewest nodes) because this minimises training time, improves network generalization and lessens over-fitting effects (Hagan et al., 1996).

2.3 Boosted regression trees

BRT are (after Elith et al., 2008) "an ensemble method for fitting statistical models that differs fundamentally from conventional techniques that aim to fit a single parsimonious model. Boosted regression trees combine the strengths of two algorithms: regression trees (models that relate a response to their predictors by recursive binary splits) and boosting (an adaptive method for combining many simple models to give improved predictive performance). The final BRT model can be understood as an additive regression model in which individual terms are simple trees, fitted in a forward, stage-wise fashion". A k-fold cross-validation (CV) is used to avoid the effect of over fitting (over training) and assess the prediction performance. The training process is stochastic: it includes a random or probabilistic component (for example, sub-samples for CV are chosen randomly). This means that, unless a random seed is set initially, final models will be subtly different each time they are calibrated. BRT models can be fitted to a variety of response types (Gaussian, Poisson, binomial, etc.). The method is insensitive to outliers, and can accommodate missing data in predictor variables by using surrogates (Breiman et al., 1984). The final number of trees is controlled by two important factors: the learning rate (or shrinkage parameter) and the tree complexity.

One of the interesting features of BRT is the assessment of variable relative influences, based on the number of times a variable is selected for splitting, weighted by the squared improvement to the model as a result of each split, and averaged over all trees (Friedman and Meulman, 2003). The relative influence (or contribution) of each variable is scaled so that the sum adds to 100, with higher numbers indicating stronger influence on the response. For a detailed example, see the working guide from Elith

et al. (2008). Because interpreting variable interactions is much easier with BRT, it was used to analyse the relationships and the relative influences of input variables toward total denitrification.

2.4 NEMIS model

The NEMIS model uses a common formalism (Heinen, 2006b; Johnsson et al., 1987, 1991; Sogbedi et al., 2001):

$$Da = Dp \cdot f_N \cdot f_S \cdot f_T \tag{3}$$

with

$$f_S = \frac{N}{K+N} \tag{4}$$

$$f_N = \left(\frac{S - S_t}{S_m - S_t}\right)^w \tag{5}$$

$$f_T = Q_{10}^{\frac{T-Tr}{10}} \tag{6}$$

Da is the denitrification rate (mg N kg^{-1} soil d^{-1}) and Dp is the potential denitrification (mg N kg^{-1} soil d^{-1}). The denitrification potential can be either a LDP or a DEA. f_N is a nitrate dimensionless function, where N is the actual nitrate soil content (mg N kg^{-1} soil) and K is the nitrate soil content (mg N kg^{-1} soil) when $f_N = 0.5 \cdot f_S$ is a dimensionless function of water saturation, where S is the WFPS, S_t the WFPS threshold below which denitrification does not occur and S_m the maximal WFPS (in our case $S_m = 1$). f_T is a dimensionless function of the soil temperature T (°C), Tr is the reference temperature when the potential denitrification Dp was determined, and Q_{10} is the increase factor for a temperature increase of 10 °C. This function has a specific form in NEMIS, where two different Q_{10} are used for two ranges of temperature:

$$f_T = fT\text{ref} \times Q_{10}^{\frac{T-Tr}{10}} \tag{7}$$

BGD
7, 2313–2360, 2010

if $T \geq 10$, $Q_{10} = 2$, $Tr = 20$, fTref=1 otherwise $Q_{10} = 50$, $Tr = 10$, fTref=0.5. Temperatures are in °C.

2.5 Model development

The ANN, BRT and NEMIS models were calibrated on the same subsets. The performance assessment and the validation of the model was done using the same approach:

- The modelling performance was evaluated using the conservative independent validation: the dataset was randomly subsampled into a calibration and a test subset.
- We used a resampling technique (the random subsampling and calibration is repeated many times) to estimate the distribution of the performance criterion and its median. These estimated distributions have been used to compare different models/approaches.
- We selected the models displaying a median performance (independent validation) as representatives. These median models were used for graphical comparison and the response shape analysis.

The feed-forward ANN calibration was done using a classic method (using a training and a validation subset to control overfitting). To select the simplest ANN structure (with the fewest hidden nodes), we started with the number of nodes in the hidden layer equal to twice the number of input variables. We then decreased the number of nodes until there was a significant decrease in model performance (independent validation). A number of different ANN models have been developed. We always retained three base variables: temperature, WFPS and NO_3^- which previous studies have shown to be important. The nomenclature used is ANNn(X,Y) where n is the number of input variables ($n \geq 3$) and X, Y are the independent variables included in addition to the 3 base variables. The suffix G denotes that the model was trained on the whole (global)

dataset. Thus ANN5(OM,pH)G denotes a model using the 3 base input variables plus 2 others (OM and pH) which was trained on the whole dataset.

The BRT training was done using the methodology and the R code from Elith et al. (2008). The nomenclature used is the same as for ANN. The BRT was specifically used to analyse the variable relationships. Different combination of input variables were tested, starting from a model using all the available variables, and then discarding the variables of lower importance until model performance significantly decreased.

The NEMIS model (using DEA as the denitrification potential Dp) was calibrated following a methodology adapted from Oehler et al. (2009) and Heinen (2006a). NEMIS was calibrated on the whole dataset (denoted NEMIS4G) without the Henault and Germon (2000) dataset, because it contains no DEA measurements. NEMIS was also calibrated separately on each of the Oehler et al. (2007), Cosandey et al. (2003) and Luo et al. (1999) datasets (denoted NEMIS4O, NEMIS4C and NEMIS4L).

More details about the calibration steps are available in Appendix A.

2.6 Statistics

The model performance criterion used in this study was the Normalized Root Mean Square Error (NRMSE) defined as:

$$\mathrm{RMSE} = \sqrt{\sum \frac{(s-o)^2}{n}} \tag{8}$$

$$\mathrm{NRMSE} = \frac{\mathrm{RMSE}}{\bar{o}} \tag{9}$$

where s are simulated values, o the observed values, $\bar{o}$ the average of observed values and n the size of the sample. We used the normalized criterion to enable comparisons between different sites and studies. For median comparisons we used the non-parametric Wilcoxon Rank test (w test). All the data processing, model developments and statistics were performed using the software "R" version 2.8 (2008).

3 Results

3.1 ANN results

The best results were always obtained with a [*N*:6:1] ANN topology, *N* being the number of input factors (*N* varied from 4 to 7). Using a large number of internal nodes tended to give a better fit to the training dataset, but without gains for the test dataset (i.e. independent validation).

The ANN4(OM)G model and the ANN4(DEA)G model (Fig. 2) performed equally on the Test dataset (independent validation) (w test, $p > 0.05$). This demonstrates that OM is as good a predictor for denitrification rate modelling as DEA. Figure 3 shows the effects of adding other variables to ANN4(OM)G base. Adding a 5th variable significantly improved model performance (w test, $p < 0.05$). It did not matter whether this 5th variable was *Db*, pH or SD as each of them resulted in a similar improvement. Adding a 6th or 7th variable did not improve model performance (w test, $p > 0.05$), but displayed more over-training effects. Potentially, different techniques could be used to "open" the ANN and try to understand the variable relationships. Given the limitations of such techniques (Olden et al., 2004) we preferred to use the BRT approach.

3.2 BRT results

The BRT was constructed using a number of trees varying from 1000 to 1500, a learning rate of 0.01 and a tree complexity of 5. More complex structures were not found to increase prediction performance.

The BRT8(OM,pH,Db,DEA,SD)G model performance (NRMSE, independent validation) was 1.08. Figure 4 shows the relative influence of the different variables on the response. The variables are sorted from the most influencing: OM, WFPS, *T*, NO_3^-, pH, *Db*, DEA, SD. Scores for *T* and WFPS were not significantly different (w test, $p > 0.05$). Simplification of the model down to 5 variables was done without significant loss of performance. Figure 5 shows the hierarchy of the variables for the BRT5(OM,pH)G

model, which did not differ from BRT8(OM,pH,Db,DEA,SD)G. The relative influence of DEA did not change with or without Henault and Germon (2000) records. *Db*, DEA and SD accounted for less than 10% of the influence, less than the influence of pH (12.4%). Reducing the BRT topology to 5 inputs did not shift the influence carried by the 3 discarded variables to a particular one. Discarding pH from the model did not increase the influence of Db, DEA and SD. Apparently, the effects of *Db* and pH are independent, or at least treated as such in the BRT approach. Also, the importance of NO_3^- is evident. As envisaged, the BRT approach successfully lessened dataset autocorrelation effects.

3.3 Choosing an ANN model using combined ANN and BRT results

One of the main results from the ANN approach is the possible replacement of DEA in favour of OM. Looking at the BRT analysis, OM appears to be a better candidate. As already stated, the 3 supplement variables *Db*, pH and SD are correlated (Table 2), and adding one of these to the model improves significantly its predictive performance, in both ANN and BRT approaches. Adding another one seems to add unnecessary complexity to the model, without performance gains. In our final model we decided to add pH to the base factors and OM. We rejected SD because the mechanism by which soil depth influences denitrification is unclear, and this variable has clearly a low influence in BRT approaches. Looking at the ANN results only, there was no clear trend toward the choice of *Db* or pH. But looking at the BRT results, pH helped explaining variability more than *Db*. The cause-effect relationship between pH and denitrification remains unclear, even though a through review (Simek and Cooper, 2002) has clearly shown that such an effect is indeed present and should be accounted for. pH might also be important when estimating N_2O emissions because pH affects the N_2/N_2O ratio (Firestone et al., 1980).

3.4 The ANN5(OM,pH)G model

The final ANN model denoted ANN5(OM, pH)G has a [5:6:1] topology. Figure 6 shows the performances (independent validation with the test subset) of the ANN5(OM, pH)G model, the BRT5(OM, pH)G model and the NEMIS4G model. Removing Henault and Germon (2000) records does not change the median performance of the ANN and BRT models. The median tests NRMSE are respectively 1.10, 1.11 and 1.91 clearly indicating that ANN5(OM, pH)G and the BRT5(OM, pH)G model result in better predictions (w test, $p < 0.05$, computed using Oehler et al. (2007), Cosandey et al. (2003) and Luo et al. (1999) but without Henault and Germon (2000) records because this record does not include DEA which is a required input in NEMIS). Heinen (2006a) already pointed out that NEMIS-like models can perform quite well when calibrated for a specific site, but that they do not perform well when applied over a range of different soil types with the same parameter set. The site-specific calibrated NEMIS on the Oehler et al. (2007), Cosandey et al. (2003) and Luo et al. (1999) datasets (NEMIS4O, NEMIS4C and NEMIS4L) showed that model coefficients (notably those relating denitrification rate to WFPS) varied significantly among datasets. Consequently, when NEMIS was calibrated using all 3 datasets (NEMIS4G) it did not perform particularly well. There is no clear difference in prediction performance (w test, $p > 0.05$) between the ANN5(OM, pH)G and the BRT5(OM, pH)G models. Looking at the range of the Test NRMSE (roughly between 0.5 and 1.9 for ANN, 0.7 and 2.2 for BRT, and 1.2 and 2.8 for NEMIS), there seems to be a rather high influence of the sub-sampling process on the independent validation process. This can be due to the lack of data (55 records for the Test dataset) coupled to the presence of few extreme *Da* values that can have a relatively large impact on the independent validation process. In contrast, the Training NRMSE min and max values are relatively low for both BRT and ANN (between 0.70 and 1.00 for a median of around 0.80).

Figure 7 presents the performance of the chosen ANN5(OM, pH)G, BRT5(OM, pH)G and NEMIS4G. ANN NRMSE for each dataset from Oehler et al. (2007), Henault and

Germon (2000), Luo et al. (1999) and Cosandey et al. (2003) are respectively 1.20, 0.64, 0.67 and 0.48. ANN5(OM, pH)G and BRT5(OM, pH)G display a comparable behaviour. As expected, when looking at the NRMSE numbers, Fig. 7 highlights the meaning of the differences in performance (from NRMSE of 1.90 to 1.10) between the median NEMIS4G model and the median ANN5(OM, pH)G model. The ANN model clearly outperforms the NEMIS model. Site specific calibration of NEMIS gives median NRMSE (computed on the whole dataset, not independent ones) for Oehler et al. (2007), Luo et al. (1999) and Cosandey et al. (2003) of 1.55, 0.66 and 1.03, to be compared with 1.20, 0.67 and 0.48 for ANN5(OM, Db)G. Overall, the ANN5(OM, pH)G model seems to be at least as good as (w test, $p > 0.05$ for Luo et al., 1999), if not better (w test, $p < 0.05$ for Oehler et al., 2007 and Cosandey et al., 2003), than the site specific NEMIS models.

3.5 ANN model response shapes

The power of ANN is in their capability to capture more than the sum of separated effects, i.e. the variable interactions. In Figs. 8 and 9 we show univariate and covariate response shapes. For Fig. 8 the fixed values were chosen to represent the conditions of a classic cultivated silty loam soil with medium OM%. Temperature and WFPS are set to be non-limiting. For Fig. 9 the fixed values were chosen to represent a silty clay riparian soil with plausible occurrence of a Temperature of 16 °C and full water saturation (WFPS=100%). Beside Temperature and WFPS, the fixed values were around the median of the dataset. In particular, the Fig. 9 presents the *Da* rate response to variations of different factors, at a fixed temperature of 16 °C. At 10 °C and down to 5 °C, response shapes, and trends are globally similar. The ANN used for these graphs and the algorithm to allow for its reconstruction are detailed in Appendix B. Intuitively the level of trust of the model is related to the data density in each part of the data space. Denser data points are needed to represent fast gradient change area. Without external knowledge, we cannot know if the gradients are well represented: indeed, they are built with the data (training). The assessment of the performance (or trust) is made at a

global level (the NRMSE). The prediction performance of the ANN outside the training (calibration) dataset space is not assessed and it can display physically unrealistic behaviour. Dataset boundaries are shown in Fig. 8 and Fig. 9. As guidelines to evaluate data point distribution, scatter plots of the combinations of Da, NO_3^-, WFPS, OM, pH and T are available in the Appendix.

3.5.1 Influences of T and WFPS

The response shapes of T and WFPS (Fig. 8a and Fig. 8c) are similar to the description proposed by NEMIS types of models. Whatever the values of other variables, there are clear threshold values: below 11 °C and a WFPS of 40%, predicted Da rate is very low, and nearly null for WFPS<20%. At around a WFPS of 75%, 50% of the maximal Da rate is achieved. The response to T below 5 °C seems odd, slightly increasing toward 0 and with a non 0 response when $T = 0$ °C. This may show some of the limits of the datasets which included no records with very low or null T. WFPS needs to be greater than 60% to be less limiting than $[NO_3^-]$ or OM % (Fig. 9a and d). This figure is shown also for lower T. Overall, WFPS is predicted to be a far more important driver than T when T is below 11 °C (conditions which can be found during winter in temperate areas).

3.5.2 Effects of the substrates NO_3^- and OM

Da response to $[NO_3^-]$ variation (Fig. 8b) is (also) similar to NEMIS. Very low NO_3^- concentrations still induce a relatively high Da. Also, above a $[NO_3^-]$ of 200, and up to 800 mg N kg^{-1} soil, the response is not a straight plateau line, but quickly decreases. Again, this shows some of the limits of the dataset which included no records with very low or null $[NO_3^-]$, and data are very scarce for $[NO_3^-]$>200 mg N kg^{-1} soil (Fig. 1, top middle panel) especially around 20 °C.

The response to OM is nearly linear with Da rate increasing with OM % (Fig. 8d). Looking at Fig. 9b, NO_3^- and OM effects seem rather independent and additive. The

behaviour of the model near values of 0 may seem odd. As we are modelling a non-dynamic *Da* rate, beside an artefact effect due to a lack of data (particularly true for low OM), the model simply predicts that denitrification starts at very low levels of $[NO_3^-]$ and OM. Also some low NO_3^- supply from nitrification process might have also occured, even if measurements were performed using the C_2H_2 blockage technique which in principle should inhibit nitrification. In practice, when included in a dynamic model the overall denitrified N will be very low in such conditions, with NO_3^- being quickly depleted.

3.5.3 Influence of pH

The shape of the response to pH is a skewed bell, with a maximum for a pH of around 7.1 (Fig. 8e). The decrease after this maximum value toward alkaline condition is supported by few records. Figure 9c and f shows an independent pH response with NO_3^- and OM, with again a maximum around a pH of 7. This value is coherent with what has been found in the literature (see Simek and Cooper, 2002). However, looking at Fig. 9e, pH impact seems to be a function of WFPS (or rather the other way round, as it is unlikely to see fast variations of pH in soils), with maximum values going down from WFPS/pH of 100%/7 to 40%/5. This result is the same at 10 °C, and is also present in the BRT5(OM,pH)G model. As pH represents different types of soils, this might be the expression of different micro-organism populations, with different sensitivity to low $[O_2]$, hence to low WFPS. This effect might be also related to the soil structure: we know that pH is positively correlated to Db in our dataset and the ML methodologies might have captured some indirect Db effects. Finally, the degree of water saturation of the soil micro-porosity for a given WFPS changes for different soil Db and might be more important than saturation of the macro-porosity in controlling O_2 diffusion. Here the effect of WFPS on limiting O_2 diffusion may simply be higher in low *Db* soils (which exhibit low pH), because of a higher macro/micro porosity ratio.

4 Discussion

Our work was first motivated by concerns on the predictive performances of widely used simplified model such as NEMIS. The use of ML methods not only provided far more generalized denitrification models with better performances, but may have also shed new light on denitrification processes, especially with respect to the relative influence of factors and how they interact.

The ANN calibration method aimed at reducing the effect of initial conditions by repeating training and sub-sampling, and by carefully assessing the prediction performance. This is relatively time-consuming but is essential. The BRT method is based on more rational design decisions, has very good predictive performance, results are easier to interpret and the training is faster. But we can see some issues for efficient uses as a predictive model: the model response is not smooth, and because of portability and mainly computing time issues, it cannot be easily and efficiently implemented in field or larger scale models.

The two ML models perform better than NEMIS on our extended database (which includes data from uplands and wetlands in intensive and less intensive agrosystems but with relatively uniform (loamy) soil types). This is also true for the ANN4(DEA)G (NRMSE=1.35) which uses the same inputs as NEMIS. To be fair with the NEMIS model, it is to be noted that NEMIS was originally designed to use a LDP, which is obviously quite a different method to evaluate denitrification potential than DEA. However, the model has been successfully used with DEA measurements (see Heinen references for more details).To check that the conjugate gradient method used for NEMIS optimisation was not underperforming, we also tried other techniques such as differential evolution, but results did not improve significantly.

Overall, performances are still around an error of 100%. The performance seems to be mainly impaired by the Oehler et al. (2007) dataset. The main characteristics differentiating this dataset from the others are the presence of different soils and that measurements have been obtained in natural conditions (low temperatures), exhibiting

the lowest *Da*. There might be a real effect not captured by the ML algorithms or not contained in the tested input variables. It is also possible that the relatively high measurement errors associated with low gas concentrations could be the cause of discrepancies in the prediction performance or reflect a limitation of the C_2H_2 blockage technique, especially on these low drainage soils.

At first glance, the ANN results agree with the mathematical representation of NEMIS like models, which were already capturing the main effects (beside pH). In details, contrary to ANN, NEMIS does not take into account variable interactions, such as temperature impact on each factor, which are more subtle than a linear effect. This is apparently needed to efficiently simulate denitrification in real world conditions, where input variables are not at the higher end of their range (e.g. 20 °C, 100% WFPS, 200 mg N kg^{-1} soil, 10% OM, pH 7) as often explored in laboratory-controlled experiments (e.g. the ones used to build NEMIS). ML are less sensitive to data noise, and this might also explain why they perform better on the low Da range. Especially because measurements tend to be less precise as we measure low Da rates, and measurement biases and errors tend to be more impacting (e.g. limit of the sensor sensitivity, leaks or contaminations becoming more important, impact of nitrification inhibition if low $[NO_3^-]$). Overall, we think the main significance of this contribution is methodological: with ML approaches (or other modelling approaches like the generalized linear models or the additive linear models) different experimental design (other than controlled laboratory experiments) could be used to understand processes, especially at larger scales (e.g. catchment). The better representation of small Da rates may also have an impact on our understanding of the N cycle dynamic at the catchment scale, mainly because the unsaturated areas can represent the vast majority of the total surface. As the problem is non-linear and spatial interactions are crucial, this would have to be thoroughly tested combining ML with spatially distributed models.

ML approaches are interesting tools to study single variable effects, and, if enough data is available, they may not need measurements from experiment specifically designed to study the impact of separate factors. They are particularly useful to analyze

and design models with data from surveys based on stratified experiment approaches (i.e. gradients sampling). As such, when using ML as an analysis tool the main objective when collecting data is to capture gradients, the most variability possible in all the variable spaces. To develop a NEMIS like model, a classic laboratory controlled experiment where all the variables are fixed but one was used. Generally, more measurements are needed for a ML analysis. However data can be obtained from surveys and not only from manipulative experiments, and may be more representative of the studied process in his "non-disturbed" environment. Moreover, interactions are more likely to be captured. After a ML based analysis, if the process and variable relationships are better understood, a simpler mathematical representation can be formuled.

The BRT analysis reaffirms the importance of temperature, WFPS and NO_3^-, and highlights the importance of OM and pH. Our results and other works (Cosandey et al., 2003; Simek et al., 2000, 2002) indicate that the relationship between DEA and Da is unclear. We successfully used OM instead of DEA without performance loss. This is consistent with the findings of Cosandey et al. (2003), who suggested that the proximal factors, available OM, O_2 and NO_3^-, exert a stronger control on denitrification rates than the size of the denitrifying enzyme pool. As we used the Cosandey et al. (2003) dataset, we checked separately its impact on BRT results. It appears that without the Cosandey et al. (2003) data, OM and DEA have the same relative influences. The Cosandey et al. (2003) dataset shows the widest OM range and the highest OM values in our database. As the sampling of the gradients is not uniform across the datasets, this particular dataset might have biased the results while representing only 13% of the records. Also DEA measurements may be less precise than OM measurements. This might have led the BRT analysis to favour OM, even if it is relatively resistant to data noise. However Cosandey et al. (2003) dataset presents a larger range of values and there is no clear trend in favour of DEA without this dataset. More recently, apart from the Cosandey et al. (2003) conclusions, Miller et al. (2008) suggested that Da is decoupled from the denitrifier community abundance. Overall, DEA does not seem to be a better indicator of Da rate than OM, especially in agrosystems where supply of

NO_3^- is frequent and denitrifier communities are already adapted to their environment. An interesting implication is the integration of a feedback loop from soil organic carbon long term dynamic.

Another relevant result is related to the effect of pH. This factor may be the one which has to be taken into account to differentiate soils, and give the "generic" quality of the ANN model. Despite genuine reported effects of pH on denitrification (see Simek and Cooper, 2002), this would have to be confirmed with a larger dataset, as the separation of pH and Db may not be sufficient. We will need to widen the database to other more contrasted type of soils (with more clay notably) and more records to fill gaps in the gradients and lessen dataset effects. This will improve prediction accuracy and increase model generality.

ANN might be a promising approach for $N_2O/(N_2+N_2O)$ modelling as well (soil N_2O emission modelling with ANN has already been successfully performed, but not specifically from denitrification (Ryan et al., 2004). The next obvious step will be coupling the ANN model to a catchment scale model.

Appendix A

Detailed calibration routines

A1 ANN calibration

The ANN training was done using the package AMORE (Pernía-Espinoza et al., 2005). The methodology used is outlined in the following steps:

1. scaling the input variable:

$$x_{i,\text{scaled}} = \frac{x_i - \min_x}{\max_x - \min_x} \tag{A1}$$

Scaling is not mandatory for the input variables, but can ease further analysis of the trained ANN weights and biases.

2. Scaling the output variable: a simple linear scaling (between 0 and 1), a log transformation and an arcsine transformation of the response variable were tested and all resulted in similar prediction performance. The arcsine transformation was finally chosen because it exhibited a more normal distribution of the residuals and was not over fitted on the highest values. Specifically, the arcsine transformation implies:

$$x_{i,\text{transformed}} = \text{arcsine}\left(\sqrt{\frac{x_i - \min_x}{\max_x - \min_x}}\right) \times \frac{360}{2\pi \times 100} \quad \text{(A2)}$$

The principal characteristic of the arcsine transformation is to stretch the low and high values, and condense the medium range values. The scaling between 0 and 1 for the response variable is mandatory for ANN, as the output of the ANN is between these values (sigmoid function).

3. Randomly sub-sampling the dataset to give 3 subsets: Training, Validation, Testing (in the proportion 6/8, 1/8, 1/8 of samples). The Training and Validation subsets are used for the training (calibration) phase, and the Testing subset is used for independent validation

4. Training/Validation of a feed-forward ANN. The learning algorithm used was the adaptive gradient descend with momentum, using the robust Least Mean Log Square criterion (Liano, 1996). The ANN was initialized with random weights and bias. Over-training was controlled by the validation subset.

5. Step 4 was repeated 22 times with different initial conditions of weights and biases. This is necessary because the initial conditions of the ANN weights and biases are not neutral and can affect the prediction accuracy of the algorithm.

Specifically, assuming an a priori normal distribution, we have used the following approach: (a) we want to be in the 10% best cases (P) (b) we want to be in that case with a confidence of 90% (Pconf) That gives n such as:

$$(1-P)^n \leq P\text{conf} \Longrightarrow n \geq 22 \quad \text{(A3)}$$

Only the best combination of validation and training NRMSE was retained.

6. Steps 3 to 5 were repeated 33 times. The number of times a step was repeated results from a trade off between statistical significance and computing time. This resampling method enabled us to estimate the distribution of the performance criterion, provided confidence intervals for the calibration and prediction process and allowed for statistical model comparisons.

7. As a representative ANN, we selected the one displaying median Test performance.

The outputs are used after being transformed back and scaled back to the original data space (the performance is evaluated in the original data space).

A2 BRT calibration

The BRT training was done using the R code from Elith et al. (2008), which uses the package gbm (Ridgeway, 2007). The methodology used is outlined in the following steps:

1. Scaling of input and output variables using the same procedure as for the ANN calibration.

2. Randomly sub-sampling the dataset to give 2 subsets: Training and Testing (in the proportion 7/8 and 1/8 of samples). The Training subset is used for the training (calibration) phase, and the Testing subset is used for independent validation.

3. Training/validation of a BRT using a Gaussian response.

4. Steps 2 to 3 were repeated 33 times. This resampling method enabled us to estimate the distribution of the performance criterion, provided confidence intervals for the calibration and prediction process and allowed for statistical model comparisons.

5. As a representative BRT, we selected the one displaying median Test performance.

The outputs are used after being transformed back and scaled back to the original data space (the performance is evaluated in the original data space).

A3 NEMIS calibration

The NEMIS model (using DEA as the denitrification potential Dp) was calibrated following a methodology adapted from Oehler et al. (2009) and Heinen (2006a):

1. Randomly sub-sampling the dataset to give 2 subsets: Calibration and Testing (in the proportion 7/8 and 1/8 of samples). The Calibration subset is used for the calibration phase, and the Testing subset is used for independent validation.

2. Calibration of a NEMIS model (minimising the RMSE with a gradient descent algorithm).

3. Steps 1 to 2 were repeated 33 times. This resampling method enabled us to estimate the distribution of the performance criterion, provided confidence intervals for the calibration and prediction process and allowed for statistical model comparisons.

4. As a representative NEMIS model, we selected the one displaying median Test performance.

Appendix B

ANN model equation

The sygmoid transfer function:

$$G(\alpha) = (1 + e^{-\alpha})^{-1} \tag{B1}$$

The full ANN equation:

$$Da = \sin\left(\frac{G(W_o \times G(W_h \times X + b_h) + b_o)}{\frac{360}{2\pi \times 100}}\right)^2 \times 1.20057 \tag{B2}$$

with the input vector

$$X = \begin{bmatrix} \frac{NO_3^- - 0.34}{759.66} \\ \frac{WFPS - 17.985}{82.015} \\ \frac{Temperature - 2.5}{18.545} \\ \frac{OM - 1}{11.2} \\ \frac{pH - 5.1}{2.9} \end{bmatrix} \tag{B3}$$

the weight matrix of the hidden layer

$$W_h = \begin{bmatrix} -0.0692 & -0.4172 & 1.4045 & -2.8685 & 1.92593 \\ 0.71683 & -2.8787 & 2.86418 & -0.4788 & -2.4463 \\ 4.66622 & 5.10768 & -7.2039 & 5.70122 & -1.5005 \\ -5.9564 & -0.6698 & -3.8146 & -2.0052 & -4.8317 \\ 9.29211 & 6.0770 & 1.27236 & 1.02784 & 7.24817 \\ -17.676 & 0.24308 & 0.37321 & -3.0420 & 1.04624 \end{bmatrix} \tag{B4}$$

BGD
7, 2313–2360, 2010

and its bias

$$b_h = \begin{bmatrix} -0.0203 \\ 0.62884 \\ -0.6357 \\ 7.28462 \\ -2.5042 \\ -1.97921 \end{bmatrix} \quad \text{(B5)}$$

the weight matrix of the output layer

$$W_o = [\ -2.0436\ \ -2.4251\ \ -2.9519\ \ -3.2196\ \ -3.3679\ \ -10.289\ ] \quad \text{(B6)}$$

and its bias

$$b_o = [5.35984] \quad \text{(B7)}$$

In the above, Da is denitrification rate (mg N kg^{-1} soil d^{-1}), NO_3^- is nitrate soil concentration (mg N kg^{-1} soil), temperature is in (°C), OM is in organic matter % (g OM g^{-1} soil). Figure D represent the topology of this ANN.

Appendix C

List of the abbreviations

ANN	Artificial Neural Network
BRT	Boosted Regression Trees
C_2H_2	acetylene
CV	Cross-Validation
Da	actual Denitrification
Db	Bulk density
DEA	Denitrifying Enzyme Activity

LDP	long term Denitrification Potential
ML	Machine Learning
N	Nitrogen
N_2	Di-Nitrogen
N_2O	Nitrous Oxyde
NO_2^-	Nitrite
NO_3^-	Nitrate
NRMSE	Normalized Root Mean Squared Error
O_2	Di-oxygen
OM	Organic Matter
SD	Soil Depth
T	Temperature
WFPS	Water Filled Pore Space
w test	Wilcoxon rank test

Appendix D

Data point distribution in the 5 chosen factors and response data space

Figure D1 represents the scatterplots of the combination of *Da*, NO_3^-, WFPS, OM, pH and *T*. These can be used as guidelines to evaluate the domain of validity of the ANN5(OM,pH)G model.

Acknowledgements. We would like to thank John Leathwick for his kind support regarding BRT. The work was founded by the National Institute of Water & Atmospheric Research Capability Funds.

References

Arnold, J. G. and Fohrer, N.: SWAT2000: current capabilities and research opportunities in applied watershed modelling, Hydrol. Process., 19, 563–572, 2005. 2315

Basset-Mens, C., Anibar, L., Durand, P., and van der Werf, H. M. G.: Spatialised fate factors for nitrate in catchments: Modelling approach and implication for LCA results, Sci. Total Environ., 367, 367–382, 2006. 2315

Beaujouan, V., Durand, P., and Ruiz, L.: Modelling the effect of the spatial distribution of agricultural practices on nitrogen fluxes in rural catchments, Ecol. Modell., 137, 93–105, 2001. 2315

Beven, K.: Prophecy, Reality and Uncertainty in Distributed Hydrological Modeling, Adv. Water Res., 16, 41–51, 1993. 2316

Breiman, L., Friedman, J., Olshen, R., and Stone, C.: Classification and Regression Trees, Wadsworth International Group, Belmont, CA, USA, 1984. 2322

Cheng, W., Tsuruta, H., Chen, G., and Yagi, K.: N2O and NO production in various Chinese agricultural soils by nitrification, Soil Biology & Biochemistry, 36, 953–963, 2004. 2320

Cicerone, R.: Changes in stratospheric ozone, Science, 237, 35–42, 1987. 2314

Cosandey, A. C., Maitre, V., and Guenat, C.: Temporal denitrification patterns in different horizons of two riparian soils, Eur. J. Soil Sci., 54, 25–37, 2003. 2317, 2318, 2319, 2321, 2325, 2328, 2329, 2334, 2348

Cote, M., Grandjean, B. P. A., Lessard, P., and Thibault, J.: Dynamic modelling of the activated sludge process: Improving prediction using neural networks, Water Res., 29, 995–1004, 1995. 2316

Crutzen, P. J., Mosier, A. R., Smith, K. A., and Winiwarter, W.: N_2O release from agro-biofuel production negates global warming reduction by replacing fossil fuels, Atmos. Chem. Phys. Discuss., 7, 11191–11205, 2007, http://www.atmos-chem-phys-discuss.net/7/11191/2007/. 2314

Cybenko, G.: Approximation by superpositions of a sigmoidal function, Mathematics of Control, Signals, and Systems, 2(4), 303–314, 1989. 2316

Elith, J., Leathwick, J. R., and Hastie, T.: A working guide to boosted regression trees, J. Animal Ecol., 77, 802–813, 2008. 2317, 2322, 2325, 2337

Faraggi, D. and Simon, R.: A neural network model for survival data, Statistics in Medicine, 14, 73–82, 1995. 2316

Firestone, M., Firestone, R., and Tiedje, J.: Nitrous oxide from soil denitrification: factors controlling its biological production, Science, 208, 749–751, 1980. 2327
Fisher, J. and Acreman, M. C.: Wetland nutrient removal: a review of the evidence, Hydrol. Earth Syst. Sci., 8, 673–685, 2004, http://www.hydrol-earth-syst-sci.net/8/673/2004/. 2315
Friedman, H. and Meulman, J.: Multiple additive regression trees with application in epidemiology, Statistics in Medicine, 22, 1365–1381, 2003. 2322
Groffman, P. M., Altabet, M. A., Bohlke, J. K., Butterbach-Bahl, K., David, M. B., Firestone, M. K., Giblin, A. E., Kana, T. M., Nielsen, L. P., and Voytek, M. A.: Methods for measuring denitrification: Diverse approaches to a difficult problem, Ecol. Appl., 16, 2091–2122, 2006. 2315, 2316, 2318
Haag, D. and Kaupenjohann, M.: Landscape fate of nitrate fluxes and emissions in Central Europe - A critical review of concepts, data, and models for transport and retention, Agr. Ecosyst. Environ., 86, 1–21, 2001. 2315
Hagan, M. T., Demuth, H. B., and Beale, M.: Neural network design, PWS Publishing Company, Boston, Massachusetts, 1996. 2321, 2322
Hansen, S., Jensen, H. E., Nielsen, N. E., and Svendsen, H.: Simulation of nitrogen dynamics and biomass production in winter wheat using the Danish simulation model DAISY, Nutr. Cycl. Agroecosyst., 27, 245–259, 1991. 2320
Heinen, M.: Application of a widely used denitrification model to Dutch data sets, Geoderma, 133, 464–473, 2006a. 2316, 2320, 2325, 2328, 2338
Heinen, M.: Simplified denitrification models: Overview and properties, Geoderma, 133, 444–463, 2006b. 2316, 2323
Henault, C. and Germon, J. C.: NEMIS, a predictive model of denitrification on the field scale, Eur. J. Soil Sci., 51, 257–270, 2000. 2315, 2317, 2318, 2319, 2320, 2325, 2327, 2328, 2348, 2351, 2355, 2356
Henault, C., Bizouard, F., Laville, P., Gabrielle, B., Nicoullaud, B., Germon, J. C., and Cellier, P.: Predicting in situ soil N_2O emission using NOE algorithm and soil database, Glob. Change Biol., 11, 115–127, 2005. 2316
Hornik, K., Stinchcombe, M., and White, H.: Multilayer feedforward networks are universal approximators, Neural Networks, 2, 359–366, 1989. 2316
Hwang, S. J. and Hanaki, K.: Effects of oxygen concentration and moisture content of refuse on nitrification, denitrification and nitrous oxide production, Bioresour. Technol., 71, 159–165, 2000. 2320

IPCC: Guidelines for National Greenhouse Gas Inventories, online available at: http://www.ipcc-nggip.iges.or.jp, verified the 20th of October 2009, 2006. 2314
Irie, B. and Miyake, S.: Capabilities of three-layered perceptrons, Proceedings of the IEEE Second International Conference on Neural Networks (San Diego), 1, 641–647, 1988. 2316
Jarvis, S. C., Hatch, D. J., and Lovell, R. D.: An improved soil core incubation method for the field measurement of denitrification and net mineralization using acetylene inhibition, Nutr. Cycling Agroecosyst., 59, 219–225, 2001. 2348
Johnsson, H., Bergström, L., Jansson, P., and Paustian, K.: Simulated nitrogen dynamics and losses in a layered agricultural soil, Agr. Ecosyst. Environ., 18, 333–356, 1987. 2320, 2323
Johnsson, H., Klemedtsson, L., Nilsson, A., and Svensson, B.: Simulation of field scale denitrification losses from soils under grass ley and barley, Plant Soil, 138, 287–302, 1991. 2323
Knisel, W. G.: GLEAMS : Groundwater Loading Effects of Agricultural Management Systems, version 2.10, University of Georgia, Costal Plain Experiment Station, Biological and Agricultural Engineering Department, 1993. 2315
Kralisch, S., Fink, M., Flugel, W. A., and Beckstein, C.: A neural network approach for the optimisation of watershed management, Environ. Modell. Softw., 18, 815–823, 2003. 2316
Lehuger, S., Gabrielle, B., Oijen, M. v., Makowski, D., Germon, J. C., Morvan, T., and Henault, C.: Bayesian calibration of the nitrous oxide emission module of an agro-ecosystem model, Agr. Ecosyst. Environ., 133, 208–222, 2009. 2316
Leip, A., Marchi, G., Koeble, R., Kempen, M., Britz, W., and Li, C.: Linking an economic model for European agriculture with a mechanistic model to estimate nitrogen and carbon losses from arable soils in Europe, Biogeosciences, 5, 73–94, 2008, http://www.biogeosciences.net/5/73/2008/. 2315
Lek, S., Guiresse, M., and Giraudel, J. L.: Predicting stream nitrogen concentration from watershed features using neural networks, Water Res., 33, 3469–3478, 1999. 2316
Li, C., Frolking, S., and Frolking, T.: A model of nitrous oxide evolution from soil driven by rainfall events: 1. Model structure and sensitivity, J. Geophys. Res., 97, 9759–9776, 1992. 2315
Liano, K.: Robust error measure for supervised neural network learning with outliers, IEEE Transactions on Neural Networks, 7(1), 246–250, 1996. 2336
Lischeid, G.: Investigating trends of hydrochemical time series of small catchments by artificial neural networks, Phys. Chem. Earth Part B, 26, 15–18, 2001. 2316
Luo, J., White, R. E., Ball, P. R., and Tillman, R. W.: Measuring denitrification activity in soils under pasture: Optimizing conditions for the short-term denitrification enzyme assay and

effects of soil storage on denitrification activity, Soil Biol. Biochem., 28, 409–417, 1996. 2348
Luo, J., Tillman, R. W., and Ball, P. R.: Grazing effects on denitrification in a soil under pasture during two contrasting seasons, Soil Biol. Biochem., 31, 903–912, 1999. 2317, 2318, 2325, 2328, 2329, 2348
Martin, T. L., Kaushik, N. K., Trevors, J. T., and Whiteley, H. R.: Review: Denitrification in temperate climate riparian zones, Water Air Soil Pollut., 111, 171–186, 1999. 2315
McCulloch, W. S. and Pitts, W.: A logical calculus of the ideas imminent in nervous activity, Bulletin of Mathematical Biophysics, 5, 115–133, 1943. 2321
Miller, M. N., Zebarth, B. J., Dandie, C. E., Burton, D. L., Goyer, C., and Trevors, J. T.: Crop residue influence on denitrification, N_2O emissions and denitrifier community abundance in soil, Soil Biol. Biochem., 40, 2553–2562, 2008. 2334
Moldrup, P., Olesen, T. R., Yoshikawa, S., Komatsu, T., and Rolston, D. E.: Predictive-descriptive models for gas and solute diffusion coefficients in variably saturated porous media coupled to pore-size distribution: II. Gas diffusivity in undisturbed soil, Soil Sci., 170, 854–866, 2005. 2320
Molenat, J. and Gascuel-Odoux, C.: Modelling flow and nitrate transport in groundwater for the prediction of water travel times and of consequences of land use evolution on water quality, Hydrol. Process., 16, 479–492, 2002. 2315
Mosier, A. and Kroeze, C.: Potential impact on the global atmospheric N2O budget of the increased nitrogen input required to meet future global food demands, Chemosphere – Global Change Science, 2, 465–473, 2000. 2314
Nevison, C.: Review of the IPCC methodology for estimating nitrous oxide emissions associated with agricultural leaching and runoff, Chemosphere – Global Change Science, 2, 493–500, 2000. 2314
Oehler, F., Bordenave, P., and Durand, P.: Variations of denitritication in a farming catchment area, Agr. Ecosys. Environ., 120, 313–324, 2007. 2315, 2317, 2318, 2320, 2325, 2328, 2329, 2332, 2348
Oehler, F., Durand, P., Bordenave, P., Saadi, Z., and Salmon-Monviola, J.: Modelling denitrification at the catchment scale, Sci. Total Environ., 407, 1726–1737, 2009. 2316, 2325, 2338
Olden, J. D., Joy, M. K., and Death, R. G.: An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data, Ecol. Modell., 178, 389–397, 2004. 2326

Pernía-Espinoza, A. V., Ordieres-Meré, J. B., Martínez-de Pisón, F. J., and González-Marcos, A.: TAO-robust backpropagation learning algorithm, Neural Networks, 18, 191–204, 2005. 2335

Petersen, S. O., Schjonning, P., Thomsen, I. K., and Christensen, B. T.: Nitrous oxide evolution from structurally intact soil as influenced by tillage and soil water content, Soil Biol. Biochem., 40, 967–977, 2008. 2320

R Development Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, online available at: http://www.R-project.org, ISBN 3-900051-07-0, 2008. 2325

Ridgeway, G.: gbm: Generalized Boosted Regression Models, online available at: http://www.i-pensieri.com/gregr/gbm.shtml, 2007. 2337

Ruiz, L., Abiven, S., Durand, P., Martin, C., Vertès, F., and Beaujouan, V.: Effect on nitrate concentration in stream water of agricultural practices in small catchments in Brittany: I. Annual nitrogen budgets, Hydrol. Earth Syst. Sci., 6, 497–506, 2002, http://www.hydrol-earth-syst-sci.net/6/497/2002/. 2315

Ryan, M., Müller, C., Di, H. J., and Cameron, K. C.: The use of artificial neural networks (ANNs) to simulate N2O emissions from a temperate grassland ecosystem, Ecol. Modell., 175, 189–194, 2004. 2335

Ryden, J., Skinner, J., and Nixon, D.: Soil core incubation system for the field measurement of denitrification using acetylen-inhibition, Soil Biol. Biochem., 19, 753–757, 1987. 2317, 2348

Schapire, R.: The boosting approach to machine learning an overview, MSRI Workshop on Nonlinear Estimation and Classification, 2002, Springer, New York, 2003. 2317

Sebilo, M., Billen, G., Grably, M., and Mariotti, A.: Isotopic composition of nitrate-nitrogen as a marker of riparian and benthic denitrification at the scale of the whole Seine River system, Biogeochemistry, 63, 35–51, 2003. 2315

Simek, M. and Cooper, J. E.: The influence of soil pH on denitrification: progress towards the understanding of this interaction over the last 50 years, Eur. J. Soil Sci., 53, 345–354, 2002. 2320, 2327, 2331, 2335

Simek, M., Cooper, J. E., Picek, T., and Santruckova, H.: Denitrification in arable soils in relation to their physico-chemical properties and fertilization practice, Soil Biol. Biochem., 32, 101–110, 2000. 2320, 2334

Simek, M., Jisova, L., and Hopkins, D. W.: What is the so-called optimum pH for denitrification in soil?, Soil Biol. Biochem., 34, 1227–1234, 2002. 2334

Smith, M. and Tiedje, J.: Phases of denitrification following oxygen depletion in soil, Soil Biol. Biochem., 11, 261–167, 1979. 2348

Smits, J. R. M., Breedveld, L. W., Derksen, M. W. J., Kateman, G., Balfoort, H. W., Snoek, J., and Hofstraat, J. W.: Pattern classification with artificial neural networks: classification of algae, based upon flow cytometer data, Anal. Chim. Acta, 258, 11–25, 1992. 2316

Sogbedi, J., van Es, H., and Huton, J.: N fate and transport under variable cropping history and fertilizer rate on loamy sand and clay loam soils: I. Calibration of the LEACHM model, Plant Soil, 229, 57–70, 2001. 2323

Suen, J. P. and Eheart, J. W.: Evaluation of neural networks for modeling nitrate concentrations in rivers, Journal of Water Resources Planning and Management-Asce, 129, 505–510, 2003. 2316

Telszewski, M., Chazottes, A., Schuster, U., Watson, A. J., Moulin, C., Bakker, D. C. E., González-Dávila, M., Johannessen, T., Körtzinger, A., Lüger, H., Olsen, A., Omar, A., Padin, X. A., Ríos, A. F., Steinhoff, T., Santana-Casiano, M., Wallace, D. W. R., and Wanninkhof, R.: Estimating the monthly pCO2 distribution in the North Atlantic using a self-organizing neural network, Biogeosciences, 6, 1405–1421, 2009, http://www.biogeosciences.net/6/1405/2009/. 2316

Tiedje, J. M., Simkins, S., and Groffman, P. M.: Perspectives on measurement of denitrification in the field including recommended protocols for acetylene based methods, Plant Soil, 115, 261–284, 1989. 2348

Turpin, N., Bontems, P., Rotillon, G., Brlund, I., Kaljonen, M., Tattari, S., Feichtinger, F., Strauss, P., Haverkamp, R., Garnier, M., Porto, A. L., Benigni, G., Leone, A., Ripa, M. N., Eklo, O.-M., Romstad, E., Bioteau, T., Birgand, F., Bordenave, P., Laplana, R., Lescot, J.-M., Piet, L., and Zahm, F.: AgriBMPWater: systems approach to environmentally acceptable farming, Environ. Modell. Softw., 20, 187–196, policies and Tools for Sustainable Water Management in the European Union, 2005. 2315

Vinten, A. J. A., Castle, K., and Arah, J. R. M.: Field evaluation of models of denitrification linked to nitrate leaching for aggregated soil, Eur. J. Soil Sci., 47, 305–317, 1996. 2315

Yoshinari, T. and Knowles, R.: Acetylene inhibition of nitrous oxide reduction by denitrifying bacteria, Biochemical and Biophysical Research Communications, 69, 705–710, 1976. 2348

Yoshinari, T., Hynes, R., and Knowles, R.: Acetylene inhibition of nitrous oxyde reduction and measurement of denitrification and nitrogen fixation in soil, Soil Biol. Biochem., 9, 177–183, 1977. 2317

Table 1. The different measurement methods of Da and DEA in the database.

Source	Measure	Method variant	Incubation time	Incubation temperature
Cosandey et al. (2003)	DEA	Smith and Tiedje (1979), flasks, mixed	4 h	20 ˚C
	Da	Yoshinari and Knowles (1976), flasks, disturbed soil samples	4 h	20 ˚C
Luo et al. (1999)	DEA	Luo et al. (1996), flasks, mixed	5 h	20 ˚C
	Da	Ryden et al. (1987), soil cores, slightly disturbed	24 h	daily soil temperature variation
Oehler et al. (2007)	DEA	Luo et al. (1996), flasks, mixed	5 h 30 min	20 ˚C
	Da	Adaptation of Jarvis et al. (2001), Soil cores, undisturbed	24 h	daily soil temperature variation
Henault and Germon (2000)	*Da*	Adaptation of Tiedje et al. (1989), Soil cores, undisturbed	3 h to days	20 ˚C

Table 2. Pearson *r* correlation coefficient. Statistical significance indicated by * ($p < 0.05$)

Pearson *r*	Da	DEA	NO_3^-	WFPS	*T*	pH	OM	*Db*	SD
Da	–	0.27*	−0.03	0.43*	0.42*	0.41*	0.39*	0.12*	0.13*
DEA		–	−0.27*	0.10*	0.07	0.13*	0.66*	−0.37*	−0.34*
NO_3^-			–	−0.15*	0.16*	0.13*	−0.27*	0.24*	0.04
WFPS				–	0.07	0.20*	0.09	0.26*	0.16*
T					–	0.42*	0.03	0.30*	0.24*
pH						–	0.01	0.63*	0.24*
OM							–	−0.42*	−0.12*
Db								–	0.57*
SD									–

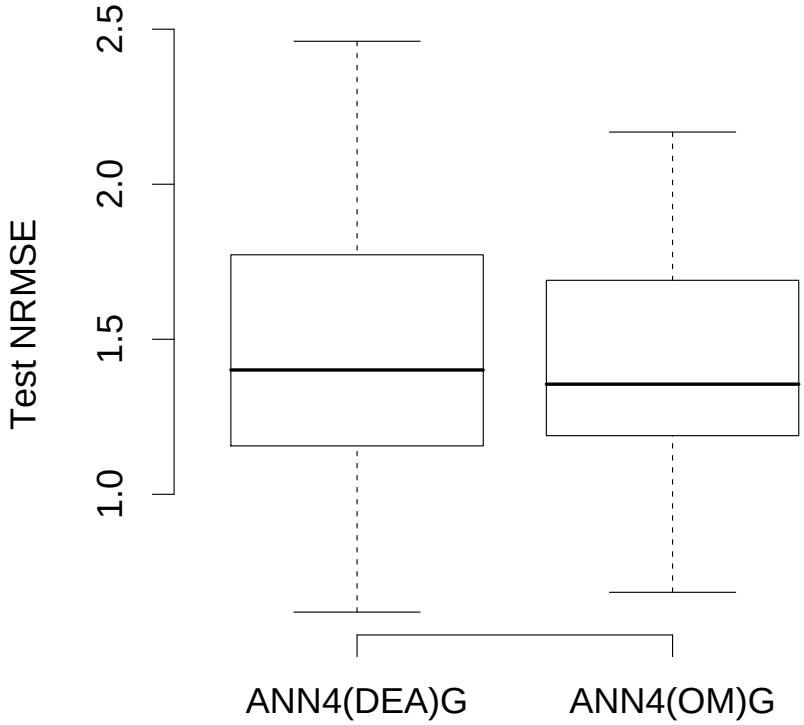

Fig. 1. Distribution of the variables in the database (449 records).

BGD
7, 2313–2360, 2010

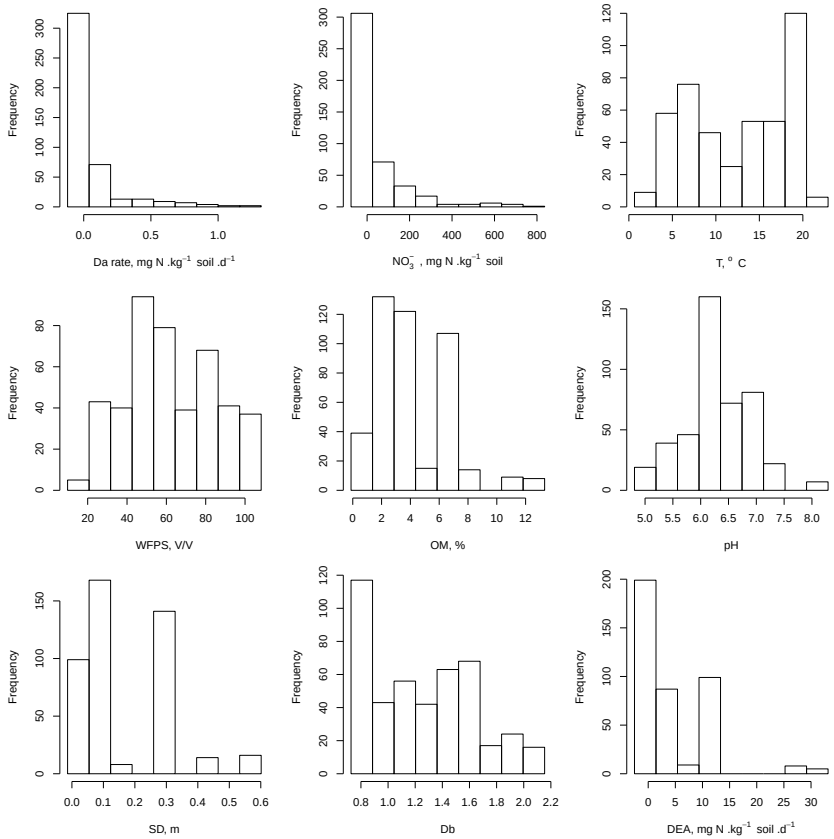

Fig. 2. Boxplots (description of quartiles with maximum at 1.5 interquartile range) of performance (independent validation) of ANN4(DEA)G and ANN4(OM)G, both without Henault and Germon (2000) records.

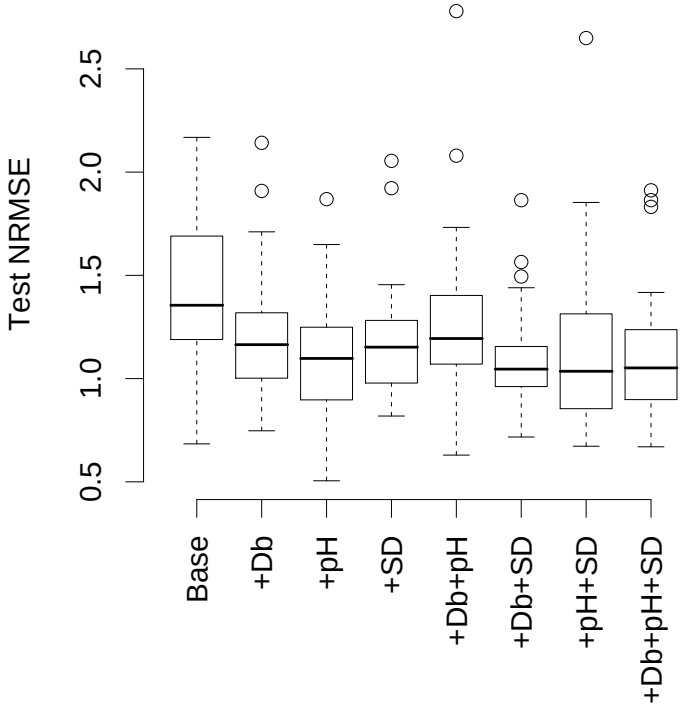

Fig. 3. Boxplots (description of quartiles with maximum at 1.5 interquartile range) of the ANN performance (independent validation) with different input combinations. Circles are outlier candidates. Base=ANN4(OM)G, +Db=ANN5(OM, Db)G, +pH=ANN5(OM, pH)G, +SD=ANN5(OM, SD)G, +Db+pH=ANN6(OM, Db, pH)G, +Db+SD=ANN6(OM, Db, SD)G, +pH+SD=ANN6(OM, pH, SD)G, +Db+pH+SD=ANN7(OM, Db, pH, SD)G.

BGD
7, 2313–2360, 2010

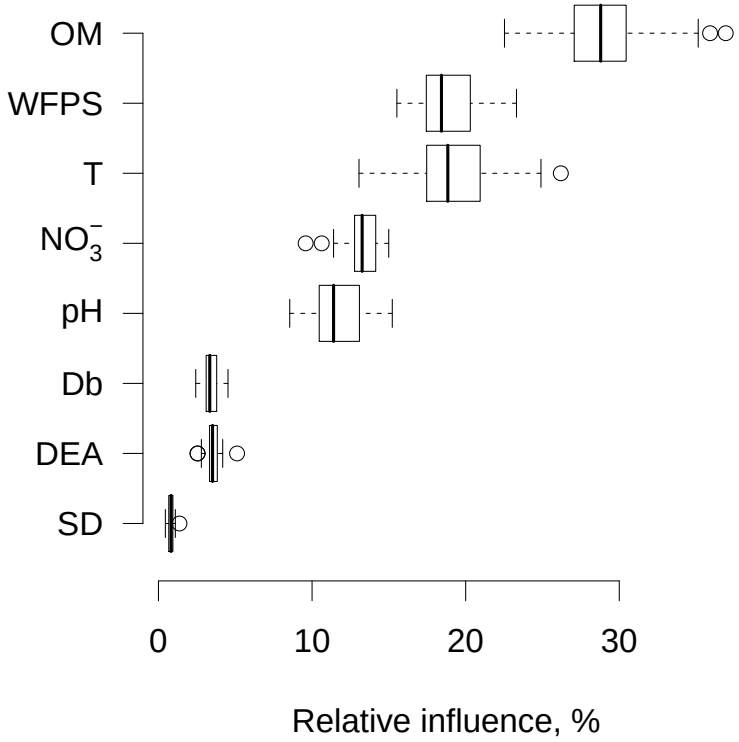

Fig. 4. Boxplots (description of quartiles with maximum at 1.5 interquartile range) of relative influences of the 8 tested input variables, as revealed by the BRT8(OM,pH,Db,DEA,SD)G model. Circles are outlier candidates.

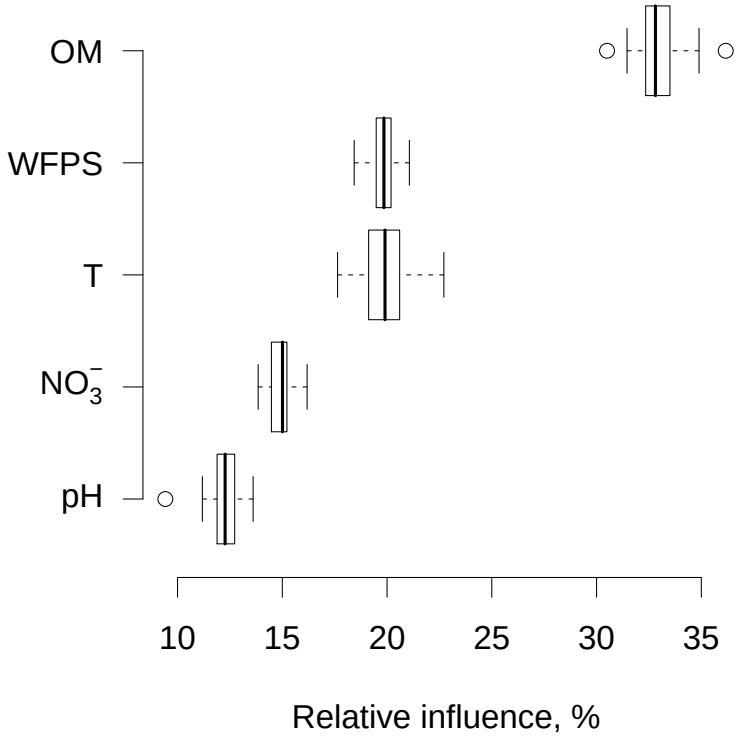

Fig. 5. Boxplot (description of quartiles with maximum at 1.5 interquartile range) of relative influences of the 5 input variables, as revealed by the BRT5(OM,pH)G model. Circles are outlier candidates.

BGD
7, 2313–2360, 2010

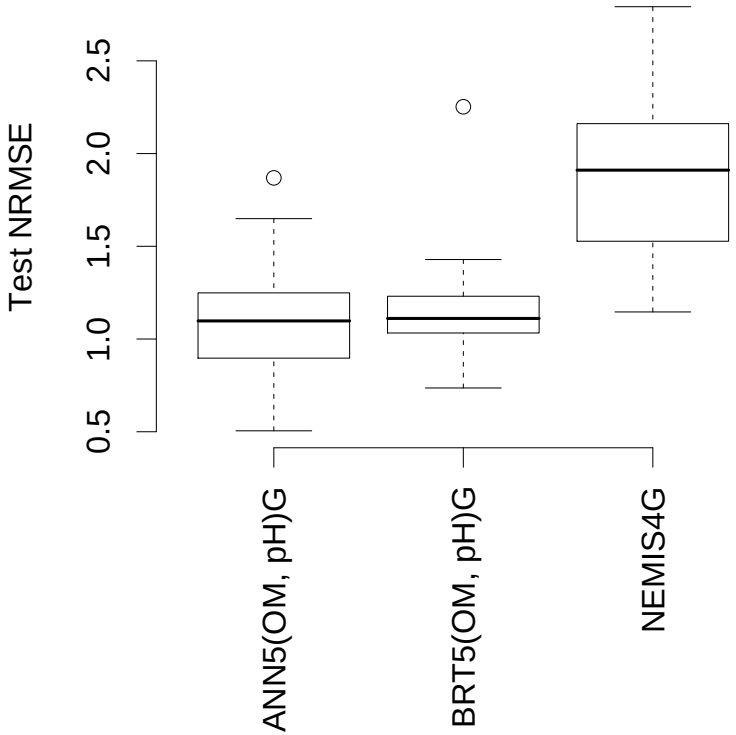

Fig. 6. Boxplot (description of quartiles with maximum at 1.5 interquartile range) of prediction performance (independent validation) of ANN5(OM, pH)G, BRT5(OM, pH)G and NEMIS4G (without Henault and Germon (2000) records). Circles are outlier candidates.

BGD
7, 2313–2360, 2010

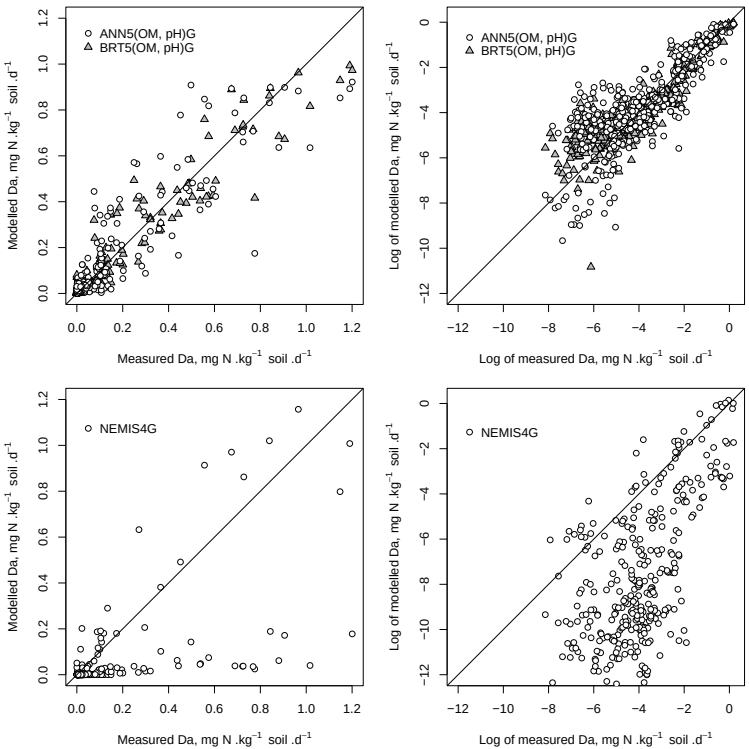

Fig. 7. Comparison of ANN5(OM, pH)G, BRT5(OM, pH)G and NEMIS4G (whole dataset without Henault and Germon (2000) records) Da prediction performance (independent validation).

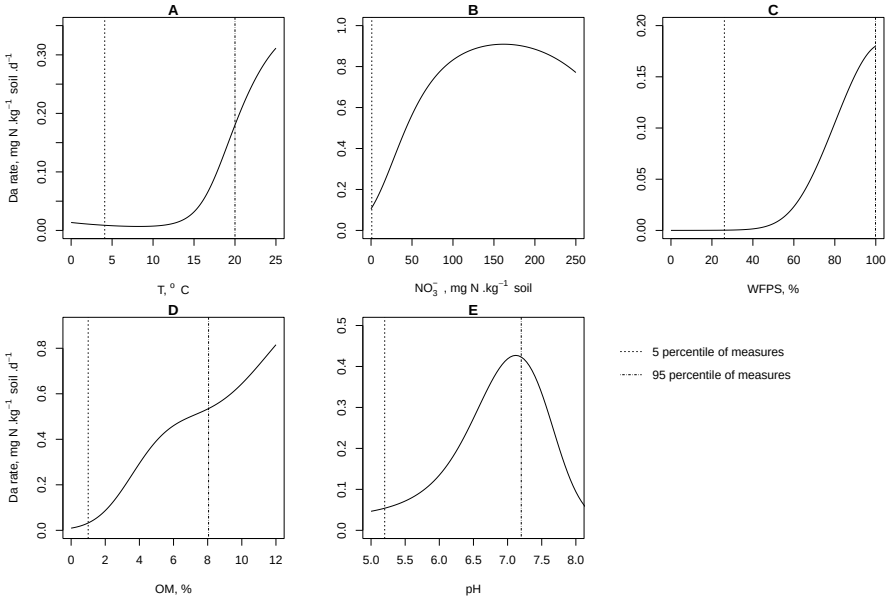

Fig. 8. Univariate responses (Da rates) of ANN5(OM,pH)G model. Fixed values: pH=6.2, OM=3%, $[NO_3^-]$=10 mg N kg^{-1} soil, Temperature=20 °C, WFPS=100%.

BGD
7, 2313–2360, 2010

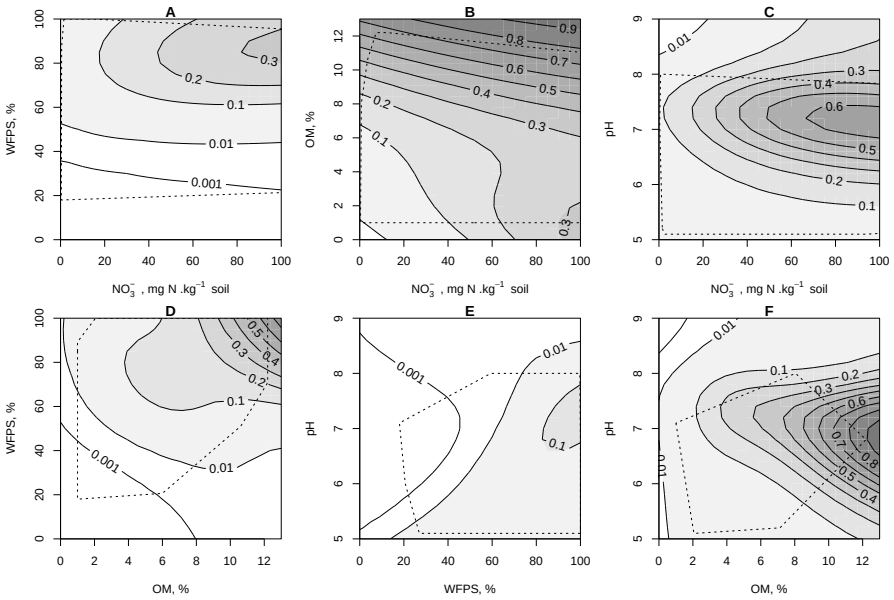

Fig. 9. *Da* rate (mg N kg^{-1} soil d^{-1}) in function of different couples of variables. The broken lines delimit the training dataset. Fixed values: pH=6.5, OM=6%, [NO_3^-]=20 mg N kg^{-1} soil, Temperature=16 °C, WFPS=100%.

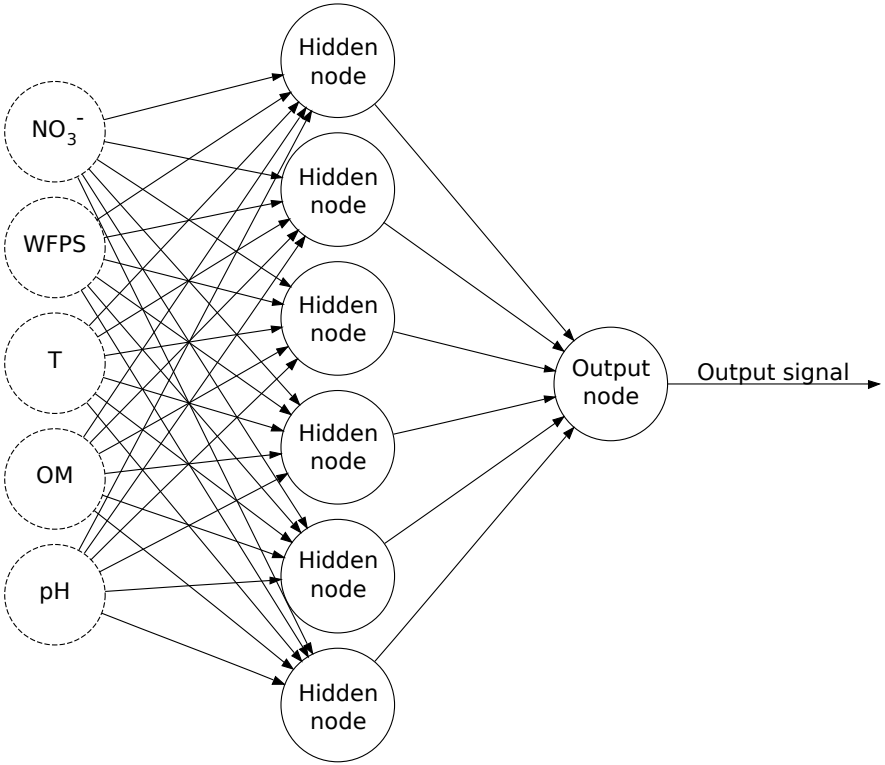

Fig. A1. ANN5(OM,pH)G model [5:6:1] topology. Solid circles represent neurons and dashed circles represent the inputs.

BGD
7, 2313–2360, 2010

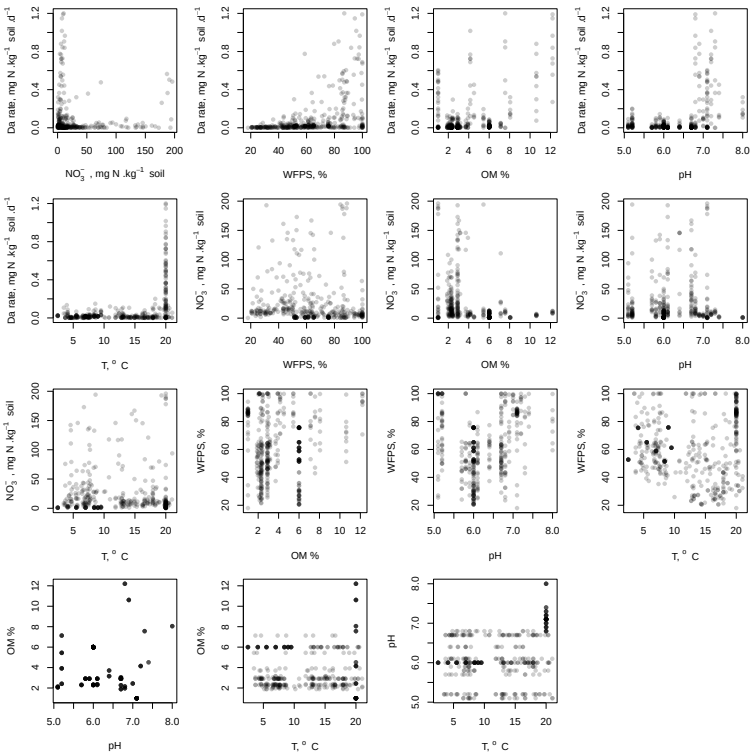

Fig. D1. Scatterplots of the combination of *Da*, NO_3^-, WFPS, OM, pH and *T*. A light grey dot represents one data point. Dots get darker as data points overlap. NO_3^- range is limited to 200 mg N kg^{-1} soil as there is few data between 200 and 800 mg N kg^{-1} soil.