



Estimating nitrogen fluxes at the European scale by upscaling INTEGRATOR model outputs from selected sites

G. J. Reinds¹, G. B. M. Heuvelink², T. Hoogland¹, J. Kros¹, and W. de Vries¹

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

²ISRIC World Soil Information P.O. Box 353, 6700 AJ Wageningen, The Netherlands and Wageningen University, The Netherlands

Correspondence to: G. J. Reinds (gertjan.reinds@wur.nl)

Received: 26 April 2012 – Published in Biogeosciences Discuss.: 4 June 2012

Revised: 14 October 2012 – Accepted: 16 October 2012 – Published: 16 November 2012

Abstract. A comparison was made between upscaled model results of nitrogen (N) fluxes to air and water from 450 sites within the EU-27 and results derived for the entire EU-27 area using the model INTEGRATOR. The 450 sites were selected using stratified random sampling, dividing the EU-27 into 150 strata and selecting three sites at random within each stratum. The strata were based on important environmental factors influencing N fluxes. Hierarchical divisive cluster analysis was used to reduce the numerous combinations of environmental factors to the required total of 150, such that the heterogeneity of environmental factors within strata was as small as possible. Modelled NH_3 , N_2O and NO_x emissions and N leaching/runoff obtained were scaled up from the 450 sites to the entire EU-27 and were within 10 % of results obtained by running the model for the whole of the EU-27 using about 36 500 sites. This implies that a reliable estimate of N fluxes for EU-27 can be made by upscaling results of the 450 selected sites suggesting that dramatic reduction in computation time can be achieved without substantial deterioration of results

as (i) decreased biodiversity of terrestrial and aquatic ecosystems caused by eutrophication and acidification, (ii) too high NO_3 concentrations in drinking water, caused by elevated nitrate (NO_3) leaching to groundwater and (iii) global climate change, induced by emissions of nitrous-oxide (Galloway et al., 2003, 2008; Erisman, 2004).

To gain insight in the risks associated with N use at the European scale, the model INTEGRATOR (De Vries et al., 2011a, c; Kros et al., 2012) has been developed that assesses N output fluxes from agricultural and terrestrial ecosystems in Europe to:

1. air, i.e. emissions of ammonia (NH_3), nitrous-oxide (N_2O), nitrogen oxides (NO_x) and dinitrogen (N_2);
2. water, i.e. N leaching to groundwater and N runoff to surface water, in response to N inputs.

INTEGRATOR calculates these output fluxes for 36 466 calculation units (NCUs: NitroEurope Calculation Units), being unique combinations of soil type, administrative region, slope class and altitude class with a mean size of 199 km^2 . To gain insight in the uncertainty of the N fluxes, a comparison of model results was foreseen within the NitroEurope project (www.nitroeuropa.eu) of different complex dynamic models that predict soil N fluxes on a European scale, such as DayCent (Del Grosso et al., 2005), DNDC (e.g. Li et al., 2000) and EPIC (Bouraoui and Aloe, 2007; Van der Velde et al., 2009), thereby performing an uncertainty analysis for each model. However, as this requires a Monte Carlo type of uncertainty quantification/uncertainty analysis (UQ/UA) of these complex models, this cannot be done for all 36 466 mapping units (NCUs). This would be too computationally

1 Introduction

The intensification of agricultural production by enhanced nitrogen use over the past few decades has strongly increased global food production, but at high environmental costs (e.g. Smil, 1999; Tilman et al., 2001). Environmental effects of the intensification in agriculture are manifested through the loss of nitrogen (N) to the atmosphere and hydrosphere, which causes a number of ecological and human health effects, such

demanding, given that this involves hundreds of Monte Carlo runs for each NCU. Therefore it was decided to limit the analysis to a sample of 450 sites within Europe, i.e. a realistic number to be processed in an UQ/UA with the complex dynamic models. The UQ/UA results can be scaled up to the whole of Europe with statistical quantification of the upscaling error provided the sample of sites is chosen using probability sampling (De Gruijter et al., 2006).

In this paper, a comparison is made between upscaled model results from 450 selected sites within Europe and the results based on simulations for the whole of Europe (i.e. all NCUs), using the model INTEGRATOR. To enable statistical upscaling of results to the entire EU, sites were selected using stratified random sampling. By using appropriate statistical inference, this yields not only an estimate but also the estimation error in statistical terms (i.e. as a probability distribution). Deterministic quantification of the upscaling error (i.e. the actual value of the error) is achieved by a direct comparison between results from the 450 sites and those obtained for the whole of EU-27. This is feasible because a single run of INTEGRATOR takes much less computing time than those of the more complex dynamic models.

The aim of the study was to assess whether a reliable estimate of the total European greenhouse gas emission can be made by upscaling results for the selected sites. First the model INTEGRATOR is shortly described, followed by a description of the stratified random sampling plot selection and of the statistical methodology used to upscale results to the European and country scale. Next the results of the plot selection are presented, followed by a description of the results on N fluxes using both methods. Results focus on the comparability of N inputs by fertilizer and manure, N uptake and N surplus, NH_3 , and N_2O emissions and N leaching/runoff. Based on the comparison, the reliability of the sampling and upscaling procedure is discussed.

2 Methodological approaches

2.1 The model INTEGRATOR

INTEGRATOR has been developed to assess responses of N and greenhouse gas (GHG: N_2O , CH_4 and CO_2) emissions to European-scale changes in land use, land management and climate at a high spatial resolution, both in the past and in the future, focusing on changes in the period 1970–2030 (De Vries et al., 2011a). It covers major agricultural and non-agricultural ecosystems (grassland, arable land, forests, and short vegetation) and includes interactions between agricultural land and non-agricultural land via gaseous emissions of NH_3 (and NO_x) and the resulting N deposition. INTEGRATOR includes complete N balances, and can be applied at EU level (EU-27). To achieve all these aims, INTEGRATOR (i) uses relatively simple and transparent model calculations based on the use and adaptation of available modelling

approaches, (ii) includes empirical modelling approaches, using statistical relationships between model outputs and environmental variables and (iii) focuses on the derivation of high-resolution spatially explicit input data. The INTEGRATOR concept is based on an appropriate balance between model complexity and data availability. For N fluxes, which is the focus of this paper, INTEGRATOR includes sub-models to predict

- NH_3 , NO_x , N_2O and CH_4 emissions from housing and manure storage systems and agricultural soils, based on the adapted MITERRA-Europe model (Velthof et al., 2009);
- NO_x and N_2O emissions from non-agricultural terrestrial systems, based on a meta-model of DNDC (Kesik et al., 2005; De Vries et al., 2007) and empirical relationships relating N_2O and NO_x emissions to georeferenced N inputs and stand/site characteristics, including meteorological parameters and soil characteristics (Bloemerts and de Vries, 2009);
- N_2O emissions from groundwater and surface waters according to a hole-in-the-pipe-model (Keuskamp et al., 2012);
- N deposition (an emission-deposition matrix for NH_3 and NO_x), accounting for the interaction between agricultural and non-agricultural soils (EMEP, 2009).

To derive a complete N budget, background emissions and energy emissions are also included in INTEGRATOR, based on literature data (Simpson et al., 1999) and IMAGE (Bouwman et al., 2006) model calculations. Since this paper focuses on results from agricultural systems, obtained from an adapted version of MITERRA-EUROPE, this model is described briefly below. MITERRA-Europe (Velthof et al., 2009) is a deterministic and static N cycling model that calculates N emissions on an annual basis, using N emission factors and N leaching fractions distinguishing various manure types and manure application practices. The model can be used to assess the effects of measures and policies aiming at emission reduction of ammonia, nitrous oxide (N_2O), N oxides (NO_x), and methane (CH_4) to the atmosphere, leaching of N (including nitrate) to groundwater and surface waters, and on the phosphorus (P) budget at EU-27 level, country level, and regional (NUTS-2) level. INTEGRATOR includes an adapted version of MITERRA, including more detail in N_2O and NO_x emission factors, based on a literature study (Lesschen and Velthof, 2009). Furthermore, INTEGRATOR calculates at the level of homogeneous combinations of land use, soil type etc., whereas MITERRA calculates at NUTS level. More information on the various differences is given in De Vries et al. (2011a).

2.2 Clustering and stratified random sampling

The set of 450 sites in Europe for which INTEGRATOR was applied was obtained using stratified simple random sampling. Since land use is considered a dominant factor, we required that each stratum had to be homogeneous with respect to land use. The EU was divided into 150 geographical strata, such that the strata were homogeneous with respect to the other environmental factors that control N fluxes. Next three sites were randomly selected within each stratum. Simple random sampling of three sites from each of the 150 clusters was achieved by using the “spsample” function from the gstat package for R (Bivand et al., 2008). Sampling was based on digital representations of the NCU clusters in the Lambert Azimuthal Equal Area projection (Geographic Coordinate System GCS ETRS 1989). The sampling procedure is explained in detail below.

2.2.1 Principle of stratified random sampling

Stratified random sampling starts by dividing the total population of M NCUs into K strata, ideally such that the variation of the target variable (i.e. greenhouse gas emission) within strata is small compared to the variation between strata. Each of the K strata contains M_k NCUs ($k=1,\dots,K$). Obviously, the sum of all M_k equals M . Within each stratum, a simple random sample of m_k sites is chosen using simple random sampling. In our case $M = 36\,466$, and we used $K = 150$ and $m_k = 3$. This yields a total of 450 sites. The important advantage of using $m_k > 1$ is that variability within strata can also be assessed. Using multiple sites from the same stratum allows comparison of model results at different sites with similar environmental conditions.

The approach taken to define the K strata (or groups) conforms with conventional clustering techniques from multivariate statistics; see Kaufman and Rousseeuw (1990) or Davis (2002) for details.

2.2.2 Factors controlling clustering

The M NCUs were clustered into groups that are homogeneous with respect to the expected model output, in this case annual N emission. Clustering was therefore based on important environmental factors that are known to influence emission, i.e. (i) land use, divided in arable land, grassland (including rough grazing) and nature (including forests but excluding wetlands), (ii) N manure application and N grazing, (iii) N fertilizer application, (iv) mean annual temperature, (v) total annual precipitation and (vi) soil type/texture, divided in sand, clay and peat. Loamy soils were added to the clay soils because of comparable hydrological characteristics. Wetlands were excluded (220 NCUs), and so were 2 NCUs for which INTEGRATOR calculated N Manure inputs greater than 400 kg ha^{-1} , because these values were considered unrealistic. As a result the total number of NCUs

was reduced from 36466 to 36244. Data for precipitation and temperature for each NCU were derived from Mitchell et al. (2004). The assessment of N inputs for each NCU is described in detail in De Vries et al. (2011b). Soil texture is an attribute for each NCU, obtained from the European Soil Database polygon map (Panagos, 2006).

Next the subsequent cluster analysis was applied to each of the three initial clusters separately. From the 36244 NCUs, there were 11375 NCUs comprising arable land (1.763 million km^2), 9789 NCUs comprising managed grass and rough grazing (grassland in pastoral use, 0.866 million km^2), and 15080 nature NCUs (1.923 million km^2). Differences in inclusion probabilities were taken into account when the results of the 450 point sites were scaled up to the entire EU (see Sect. 2.3).

2.2.3 Initial clustering per land use type by rounding

To reduce the number of initial clusters (i.e. NCUs) for each of the three land use types, we grouped NCUs with similar values of continuous numerical environmental factors using classes for these factors. For each NCU, annual precipitation less than 50 mm was rounded to multiples of 10 mm and annual precipitation greater than 50 mm was rounded to multiples of 50 mm. Mean annual temperature was rounded to multiples of $0.5\text{ }^\circ\text{C}$.

Applications of N through chemical fertilizer and animal manure less than $50\text{ kg ha}^{-1}\text{ yr}^{-1}$ were rounded to multiples of $1\text{ kg ha}^{-1}\text{ yr}^{-1}$, and applications greater than $50\text{ kg ha}^{-1}\text{ yr}^{-1}$ were rounded to multiples of $10\text{ kg ha}^{-1}\text{ yr}^{-1}$. After rounding, the original set of 36244 NCUs was reduced to 11629 unique combinations. Computation-wise, this proved to be an important advantage.

2.2.4 Data transformation

The relationship between annual N emission and annual temperature and precipitation is known to be non-linear. Therefore, annual temperature and precipitation were transformed prior to entering the cluster analysis. For temperature, a square root transform and, for annual precipitation, a sigmoid transformation were used:

$$T_{\text{tr}} = \sqrt{T + T_0} \quad (1)$$

$$P_{\text{tr}} = 1 - \frac{1}{1 + e^{c(P-h)}} \quad (2)$$

where T_{tr} and P_{tr} are the transformed mean annual temperature (T) and precipitation (P), respectively. The coefficients were taken as $T_0 = 7\text{ }^\circ\text{C}$, $c = \ln(9)/500\text{ mm}^{-1}$, and $h = 1000\text{ mm}$. T_0 was chosen such that no NCU had a negative value for $T + T_0$; the sigmoid transformation parameters were chosen such that precipitation values of 500, 1000 and 1500 mm yielded transformed values of 0.10, 0.50 and 0.90, respectively, assuming that precipitation amounts above 10000 mm will lead to soils being wet for large parts of the year with associated high N_2O emissions.

2.2.5 Calculating dissimilarities

Most clustering techniques make use of so-called dissimilarity matrices, which measure the dissimilarity or “distance” between objects. The dissimilarity is calculated from a weighted comparison of the differences between the factor values of a pair of objects. Thus, the dissimilarity between two objects p and q (i.e. NCUs) may be computed as

$$D(p, q) = \sum_{f=1}^F \alpha_f D_f(p, q) \quad (3)$$

where F is the number of factors, the D_f are individual factor distances and the α_f are weights. In the case of NCU clustering, there are five factors: N manure application, N fertilizer application, transformed annual temperature and transformed annual precipitation and soil texture class (recall that land use has already been dealt with). For soil texture, we used $D_{\text{soil}}(\text{sand, clay}) = 0.6$, $D_{\text{soil}}(\text{sand, peat}) = 0.8$, and $D_{\text{soil}}(\text{clay, peat}) = 0.4$; we thus assumed the largest dissimilarity between sand and peat soils. Soil texture dissimilarities between two NCUs with the same soil texture are zero. The distances D_f of the continuous numerical factors were computed by taking the absolute differences of the standardised values of the factors. Standardisation of factor values was achieved by subtracting the mean factor value among all NCUs from each observation and dividing by their standard deviation. The weights were chosen as $\alpha_{\text{soil}} = 0.20$, $\alpha_{\text{Nmanure}} = 0.40$, $\alpha_{\text{Nfert}} = 0.25$, $\alpha_{\text{Temp}} = 0.10$ and $\alpha_{\text{Prec}} = 0.05$. We thus assumed that e.g. differences in the amount of manure between two sites are more important than differences in temperature of precipitation. The weights were assigned based on the relative importance of environmental factors on N emissions, based on literature data and expert knowledge.

2.2.6 Comparison of three cluster techniques

The dissimilarities D were used to group objects into clusters. The general principle is that the dissimilarity between objects of the same clusters should be as small as possible. Different techniques were evaluated: (i) hierarchical agglomerative clustering (HAC), (ii) hierarchical divisive clustering (HDC) and (iii) k -means clustering (KMC). The most satisfactory technique was used to define the strata used for sampling.

HAC techniques (Kaufman and Rousseeuw, 1990) start with as many clusters as there are objects. At first, each object is a small cluster by itself. Clusters are merged until only one large cluster remains which contains all the objects. At each stage the two *nearest* clusters are combined to form one larger cluster. This requires a definition of what “nearest” means. The “Agnes” algorithm from the R-cluster package (Maechler et al., 2009) distinguishes several options. In the most common approach, the distance between two clusters is the mean of the dissimilarities between the objects in

one cluster and the objects in the other cluster. The clustering may be stopped when the required number of clusters has been reached. In our case we used 50 final clusters per land use type.

HDC algorithms construct a hierarchy of clusters starting with one large cluster containing all objects (i.e. all NCUs of a given land use type). We used the “Diana” algorithm from the R-cluster package (Maechler et al., 2009). Clusters are split until each cluster contains only a single object. At each stage, the cluster with the largest diameter was selected and split. The diameter of a cluster was defined as the largest dissimilarity between any two of its objects. To divide the selected cluster, the algorithm first looks for its most disparate object (i.e. the object with the largest mean dissimilarity to the other objects of the selected cluster). This object initiates the “splinter group”. In subsequent steps, the algorithm re-assigns objects that are closer to the “splinter group” than to the “old party”. The result is a branching tree that can be cut at any desired level of number of final clusters (i.e. 50 per land use class).

The KMC algorithm (Hartigan and Wong, 1979) is not hierarchical and therefore neither divisive nor agglomerative. In this case the algorithm starts with (arbitrary chosen) centroids of the number of desired clusters. Next all objects are assigned to their nearest cluster. Once all objects have been assigned to clusters, the centroids are computed again. In case of a categorical variable such as soil texture type, the centroid takes the dominant value. Next allocation of all objects based on the new centroids is done, and the process is repeated iteratively until no more changes occur. We used the “PAM” algorithm that is fully described Kaufman and Rousseeuw (1990).

The Agnes, Diana and PAM algorithms were implemented in the “cluster” library of the R statistical software package (<http://www.r-project.org/>). The dissimilarity matrices for arable, grass and nature were calculated and used to assess the performance of the three algorithms. To evaluate and compare the performance of the algorithms, the homogeneity of the 150 clusters was calculated and the cluster sizes (i.e. the number of NCUs grouped into a cluster) were calculated. Homogeneity was evaluated by calculating the summed within-cluster variances for each of the four continuous-numerical factors and the summed within-cluster entropy for soil texture (Brus et al., 2008).

2.3 Spatial aggregation of model outputs to the European scale

Since the selected 450 sites are a stratified simple random sample from the entire EU-27, with known inclusion probabilities, upscaling of the results at these sites to the entire EU is relatively straightforward and also allows the assessment of the associated estimation error. The mathematical-statistical procedure used to obtain estimates of averaged model outputs for the entire EU-27 and for countries within

EU-27, including the sampling error associated with the estimates, is described below.

2.3.1 Upscaling to EU-27

The mean of model outputs across the EU-27, μ , was estimated by (De Gruijter et al., 2006)

$$\hat{\mu} = \sum_{i=1}^K a_i \cdot \hat{y}_i \quad (4)$$

where

$$a_i = \frac{A_i}{A} \quad (5)$$

$$A = \sum_{i=1}^K A_i \quad (6)$$

$$\hat{y}_i = \frac{1}{m} \sum_{j=1}^m y_{ij} \quad (7)$$

with K the number of strata, a_i the relative area of stratum i , A_i the area of stratum i ($i = 1 \dots K$), A the total area of EU-27, m the number of sites within each stratum ($m = 3$), y_{ij} the model output at site j of stratum i ($j = 1 \dots m$; $i = 1 \dots K$) and \hat{y}_i the estimated mean of model output in stratum i . The estimate is unbiased, implying that

$$E[\mu - \hat{\mu}] = 0. \quad (8)$$

The variance of the estimation error is estimated by

$$V(\mu - \hat{\mu}) = \sum_{i=1}^K a_i^2 \cdot V(\hat{y}_i) \quad (9)$$

where $V(\hat{y}_i)$ is the variance of \hat{y}_i , which is estimated by

$$\hat{V}(\hat{y}_i) = \frac{1}{m \cdot (m - 1)} \sum_{j=1}^m (y_{ij} - \hat{y}_i)^2. \quad (10)$$

Assuming a normal distribution of the estimation error, the lower and upper limit of the symmetric 95 % confidence interval for the mean μ are estimated by $\hat{\mu} - 1.96 \cdot \sqrt{\hat{V}(\mu - \hat{\mu})}$ and $\hat{\mu} + 1.96 \cdot \sqrt{\hat{V}(\mu - \hat{\mu})}$.

2.3.2 Upscaling to countries

Upscaling to countries instead of the EU is easy when all K strata lie either entirely inside or entirely outside a country. In that case the equations above can be applied to the subset of strata that lie inside the country (i.e. sum only over the strata inside the country). When some of the strata lie partly inside and partly outside the country, then only that part of

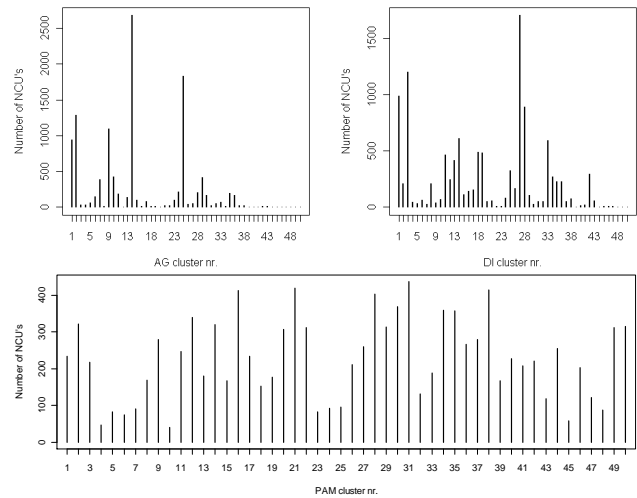


Fig. 1. Number of NCU's within each of the 50 clusters for arable land obtained with agglomerative (AG), divisive (DI) and k -means (PAM) clustering.

the stratum that lies in the country was included in the analysis and replaces the original stratum (note that this affects the stratum area A_i). The equations presented above were used as before, but note that in this case m no longer needs to be constant and equal to three, but may be smaller than three for strata that are partly outside the country. In fact, m was then replaced by the number of stratum sites that are located inside the country. This created a problem when m is smaller than two. When $m = 1$, the stratum variance can no longer be estimated, and when $m = 0$ neither can the mean value. For a stratum with $m = 1$, the variance could be estimated by merging the stratum with a neighbouring stratum (collapsed strata method, Cochran, 1977, p. 139), but in this study we applied the approach only for strata with $m > 1$.

3 Results

3.1 Clustering and stratified random sampling

Dissimilarity matrices for arable, grass and nature used to assess the performance of the agglomerative (AG), divisive (DI) and the k -means partitioning around medoids (PAM) clustering methods are presented in Table 1. Medoids are objects in a cluster with minimal mean dissimilarity to all other objects. The table shows the summed within-cluster variance for the various factors affecting N fluxes. Within-cluster variances of temperature and within-cluster entropies for soil texture are much smaller for nature than for grassland and arable land, which means that the 50 nature clusters are much more homogeneous with respect to mean annual temperature and soil texture. This can be explained by the fact that clustering for nature is not influenced by the factors manure and fertilizer, since these are zero for nature (hence zero within-cluster

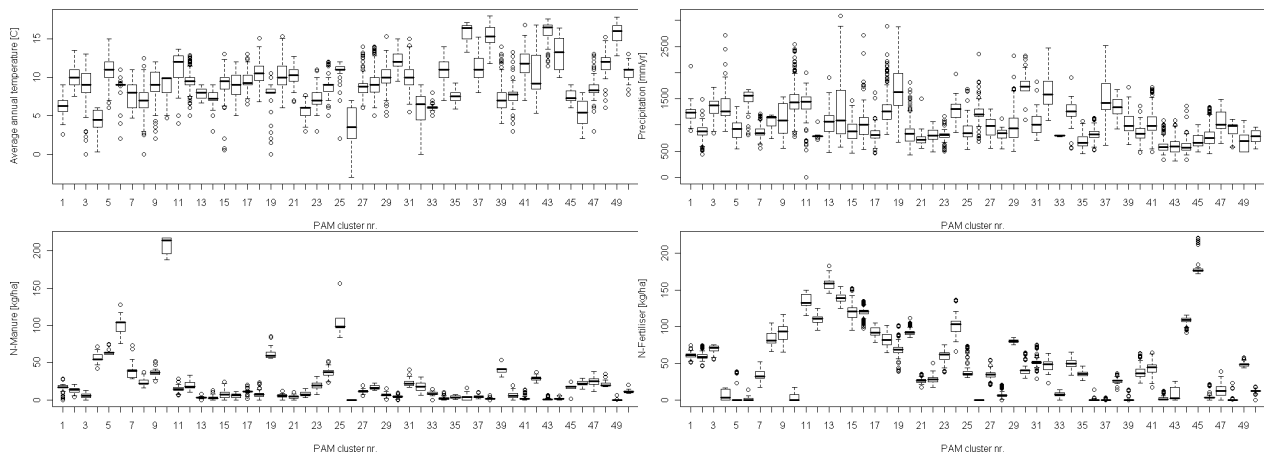


Fig. 2. Box plots of the median continuous-numerical environmental factors for each of the 50 arable clusters obtained using the *k*-means (PAM) algorithm.

Table 1. Summed within-cluster variance for mean annual temperature (°C), annual precipitation (mm), N manure ($\text{kg N ha}^{-1} \text{ yr}^{-1}$), fertilizer ($\text{kg N ha}^{-1} \text{ yr}^{-1}$) and summed within-cluster entropy for soil texture using agglomerative (AG), divisive (DI) and *k*-means (PAM) clustering.

Land Use	Method	Temp	Prec	Manure	Fertilizer	Texture
Nature	AG	31.03	1 405 440	0	0	0.00
	DI	34.82	1 195 942	0	0	0.00
	PAM	25.71	1 191 540	0	0	0.07
Grass	AG	174.23	2 018 085	1244	4277	24.99
	DI	199.61	2 353 377	1332	5314	26.73
	PAM	177.93	2 709 979	1859	6694	22.47
Arable	AG	122.61	1 125 550	995	2040	16.74
	DI	128.98	1 485 726	960	2151	23.80
	PAM	128.50	1 292 348	1388	2610	12.46

variances). Note also that the differences between precipitation variances for the three land uses are much less because precipitation has a much smaller weight. For nature, PAM has smaller variances than DI and AG, whereas for grassland and arable land AG has the smallest variances for the continuous-numerical factors (except for manure in arable land). PAM also has the smallest within-cluster entropy for soil texture. Overall, differences between the heterogeneity measures of the algorithms are not very large.

Figure 1 gives the number of NCUs within each of the 50 clusters for arable land using agglomerative (AG), divisive (DI) and *k*-means (PAM) clustering. PAM proved to be most successful in creating clusters of uniform size, while maintaining acceptable summed within-cluster variances (Table 1). The same results were obtained for grassland and nature (not shown). It was therefore decided to derive 50 clusters for each of the three land use types using *k*-means clustering with the PAM algorithm. Below, resulting maps and cluster properties are only presented for arable land.

Box plots of the values of the continuous-numerical factors for the 50 arable clusters are shown in Fig. 2. The largest

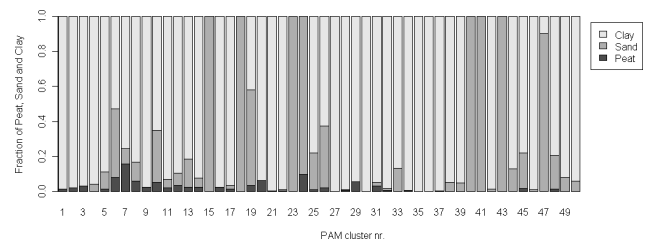


Fig. 3. Proportion of soil texture for each of the 50 arable clusters obtained using the *k*-means (PAM) algorithm.

differentiation in factor values between clusters is in annual application of N-manure and N-fertilizer. This is in agreement with the large weights assigned to these factors. Some clusters such as clusters 6, 10, 19 and 25 (all located in North-West Europe) have large values for N-manure, whereas many have small values. Cluster 45 has the largest value for N-fertilizer, whereas cluster 26 has no N-fertilizer. Annual precipitation and temperature have much more overlap between clusters, indicating that clustering was not very successful to differentiate these factors (partly caused by the low weight assigned to these factors). The proportion of the three soil texture classes, peat, sand and clay over the clusters (Fig. 3) shows that the majority of clusters are not homogeneous with respect to soil texture; in almost all cases one of the soil texture classes is clearly dominant.

The majority of clusters are also geographically concentrated (Fig. 4). For instance, note that the Czech Republic is almost entirely grouped into a single cluster. The differentiation of the various clusters is mainly associated with distinct differences in the N input by fertilizer and manure (Fig. 4). Simple random sampling of three sites within each of the 150 clusters yielded the 450 sites given in Fig. 5.

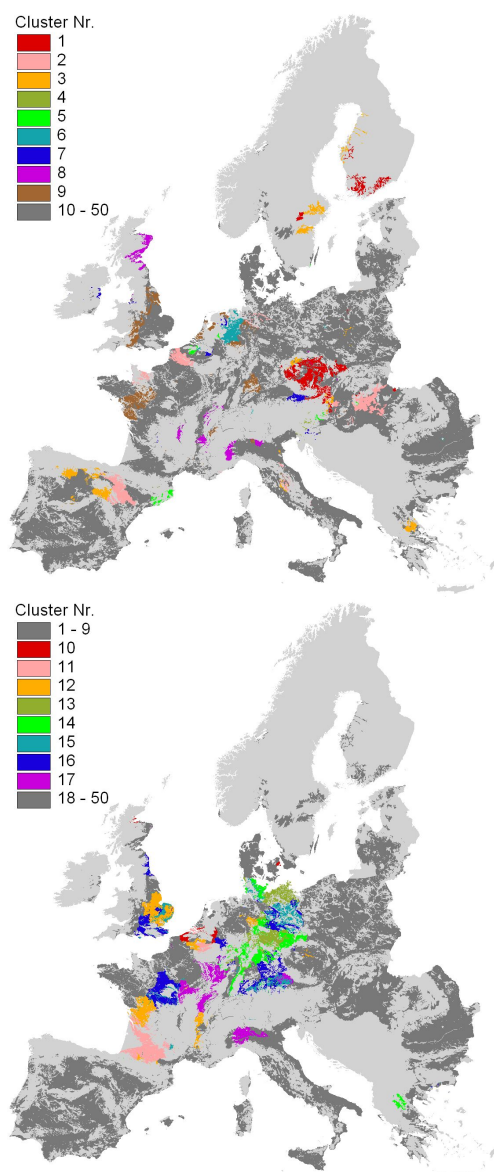


Fig. 4. Spatial distribution of the arable clusters 1–9 (left) and 10–17 (right) obtained with the PAM algorithm.

3.2 Comparison of modelled nitrogen budgets obtained by upscaling from selected areas and with full aerial support

3.2.1 Results at European level

For arable land, the mean total N inputs, crop N uptake, emission of N_2O and NH_3 and NO_x to the atmosphere and losses of N to ground water and surface water based on the 450 selected points compare well with the mean computed with full aerial support (Table 2). The overall difference is always less than 10 %, and the 95 % confidence interval includes the value obtained by running INTEGRATOR for the whole of

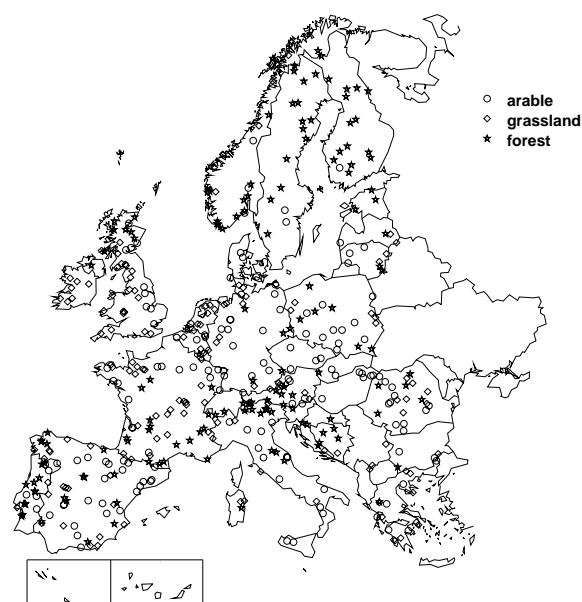


Fig. 5. Stratified simple random sample of 450 sites across Europe (EU-27) used to estimate the total emission for EU-27.

Table 2. Estimated annual N budget of arable land in EU-27 in 2000 based on INTEGRATOR using full aerial support and based on up-scaling from selected areas.

Source	N budget ($kg\ N\ yr^{-1}$)		
	Full aerial support	Upscaled means	Upscaled confidence interval
Inputs			
Manure input	34.5	33.2	28.6–37.9
N fertilizer	58.9	59.0	54.0–63.9
Deposition	11.2	12.4	11.8–13.0
Biological N fixation	5.8	6.0	5.6–6.3
Total	110.4	110.6	103.0–118.2
Outputs			
Net crop uptake	70.2	72.4	70.0–74.7
N accumulation	–31.2	–34.7	–54.4–17.4
NH_3	12.2	12.4	11.3–13.5
N_2O	1.6	1.7	1.1–2.4
N_2O , indirect	0.23	0.23	0.21–0.25
NO_x	0.82	0.96	0.52–1.4
N_2	36.6	37.9	25.9–49.9
N leaching + runoff	19.8	19.7	17.9–21.6
Total	110.4	110.6	102.3–116.7

Europe. Noteworthy is also the difference in the computed confidence intervals: for most balance terms, the 450 points yield a narrow interval, but for the emissions of N_2O and NO_x the uncertainty is larger. This is most likely due to the many factors that influence these emissions, such as rainfall, temperature and soil characteristics that have a large spatial variability.

Table 3. Estimated annual N budget of nature in EU-27 in 2000 based on INTEGRATOR using full aerial support and based on up-scaling from selected areas.

Source	N budget (kg N yr ⁻¹)		
	Full aerial support	Upscaled means	Upscaled standard deviation
Inputs			
Biological fixation	2.0	1.9	1.9–1.9
Atmospheric deposition	10.9	9.3	8.4–10.1
Total	12.9	11.2	10.3–12.0
Output from land			
N uptake	2.3	2.1	1.9–2.3
N accumulation	5.4	4.8	4.2–5.5
NH ₃	0.18	0.13	0.11–0.15
N ₂ O	0.37	0.37	0.33–0.41
NO _x	0.12	0.12	0.08–0.15
N ₂	2.1	1.9	1.8–2.1
N leaching + runoff	2.3	1.9	1.6–2.2
Total	12.9	11.2	10.6–12.2

For nature, the mean N inputs and N losses from both methods are also mostly within 10 %. The emission of N₂O is even identical for both methods (Table 3). The NH₃ flux from full areal support, however, is approximately 30 % larger than that from the 450 points and lies outside the 450-point confidence interval. This is most likely due to the influence of N deposition: with full area support the mean N deposition is higher than the mean of the 450 points. This could be due to the way the points are sampled or to the non-normal distribution of the N deposition. In INTEGRATOR NH₃ emissions increase with higher N deposition; including NCUs with higher N deposition thus causes higher mean emission. The same holds for other balance terms influenced by N deposition such as N uptake and N leaching. The results for agricultural land based on full aerial support are also presented and discussed in detail in De Vries et al. (2011b) and those for non-agricultural land in De Vries et al. (2011c). Results for grassland are not discussed because emission from grassland in INTEGRATOR is computed with MITERRA for managed grass and with the DNDC meta-model for unmanaged grass. Since both methods are used already for arable and nature for which the results are extensively described, adding grassland would not provide new insights in how well the method performs. Secondly, since grassland consists of both managed and unmanaged grass, results are not easily interpretable as a mix of two models is used to obtain the results.

3.2.2 Results at country level

Results for countries are presented only for those countries where, for a given land use type (nature and arable land), at least two points fall within the country area, so that a mean and standard deviation can be computed. As an example, results are presented for the N₂O emission. The calculated

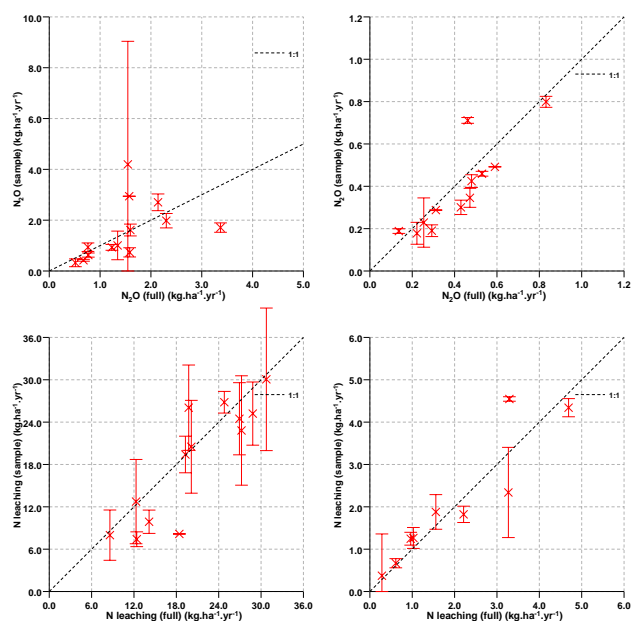


Fig. 6. N₂O emission for arable land (upper left) and for nature (upper right), N leaching for arable land (lower left) and for nature (lower right), for the full areal support (x-axis) and the selected sites (y-axis) for 13 countries that contain at least two points. Error bars indicate the range between minimum and maximum values for individual countries.

mean N₂O emission from arable land for the countries based on the sample plots within the country yields different results compared to results based on the full areal support (Fig. 6, upper left). For most countries the absolute difference is < 0.5 kg ha⁻¹ yr⁻¹, but for Austria, UK and Hungary the difference is > 1 kg ha⁻¹ yr⁻¹. With the exception of Hungary, the variance in the computed emissions at the selected plots within a country is small, indicated by the small range indicated by the error bars in the graphs. For Hungary the plots in the sample yield very different N₂O emissions, leading to a large within-country variation. For arable land the calculated up-scaled N₂O emissions in selected plots are mostly lower than the N₂O emissions computed with full aerial support, in line with the results for EU-27. For nature (Fig. 6 upper right), the uncertainty in modelled N₂O emissions is much lower because in the absence of N inputs by manure and fertilizers, the N input consists of N deposition, which has a much lower spatial variability. There is a reasonable agreement between the results from the two methods, and no systematic bias. Uncertainties in computed N leaching for arable soils are high (Fig. 6, lower left), and the results based on the sample plots mostly fall within the confidence interval based on full areal support. Again for nature, the uncertainty is lower (for the same reason as for the N₂O emission), and for a few countries the estimates based on the sample plots fall outside the confidence interval using full support (Fig. 6 lower right).

4 Conclusions

Results of this study show that an accurate estimate (within 10 %) of the total N inputs, crop N uptake, emission of N₂O and NH₃ and NO_x to the atmosphere and losses of N to ground water and surface water from arable land at EU-27 level can be obtained by running the model for a stratified random sample of 150 points. This shows that the method and data used to stratify spatial units into clusters contained the most important environmental factors that are known to influence emission. For nature, estimated N₂O and NO_x emissions were very accurate, but the procedure was less successful to predict NH₃ emissions. This may be caused by the fact that N deposition was not included in the clustering procedure, while this parameter strongly influences the NH₃ emissions from these ecosystems, or because the distribution of N deposition over Europe is skewed and we missed the high deposition areas in our set of selected sites. This study thus shows that, with only a few exceptions, a reliable estimate of the total N fluxes for EU-27 can be made by upscaling results of the 450 selected sites. If we assume that this also holds for other, more detailed, process-based N emission models, an enormous reduction in computation time can be achieved with this clustering method without substantial deterioration of results. This is specifically useful when an uncertainty analysis is carried out at EU-27 scale, because for a Monte-Carlo type uncertainty analysis a large number of model runs per site is required. It is, however, advisable to test the methods used in this study also for these type of models. INTEGRATOR computes N emissions from N inputs using mostly linear functions, and it is yet unclear if calculations on the selected sites with more complex non-linear models such as DNDC or DayCent would also be in line with Europe-wide assessments of these models.

This study also shows that the sample of 150 points per land use class for the whole of EU-27 is too small to make reliable N₂O emission estimates for arable land for most individual countries. In 14 out of the 27 countries, there were less than two points for a given land use type (nature and arable land) within the country area. For the remaining 13 countries, reasonable estimates on a country basis were only made for model outputs that depend on spatially homogenous input, such as NH₃ emissions from arable land and nature. In general, the number of points per country is too small to accurately represent the spatial variability of the factors influencing for example N₂O emissions. It is therefore advised to restrict the upscaling procedure either to EU-27 or to large regions within Europe, such as Scandinavia, the Mediterranean or Central Europe.

Acknowledgements. The authors gratefully acknowledge funding by the European Commission DG Research for the NitroEurope Integrated Project (grand no. 017841). The research is co-financed by the Dutch Ministry of Agriculture, Nature and Food Quality as part of the strategic research programmes “Sustainable spatial

development of ecosystems, landscapes, seas and regions” and “Climate change”, carried out by Wageningen University and Research Centre.

Edited by: S. Reis

References

- Bivand, R. S., Pebesma, E. J., and Gómez-Rubio, V.: Applied Spatial Data Analysis with R, Springer, New York, USA, 2008.
- Bloemerts, M. and de Vries, W.: Relationships between nitrous oxide emissions from natural ecosystems and environmental factors, Alterra Wageningen UR, Wageningen, The Netherlands, 112 pp., 2009.
- Bourauoi, F. and Aloe, A.: European Agrochemicals Geospatial Loss Estimator: Model Development and Application, Office for Official Publications of the European Communities, Luxembourg/EUR Report 22690 EN, 118 pp., 2007.
- Bouwman, A. F., Kram, T., Klein Goldewijk, K. (Eds.): Integrated Modelling of Global Environmental Change. An Overview of IMAGE 2.4, Netherlands Environmental Assessment Agency (MNP), Bilthoven, The Netherlands, 2006.
- Brus, D. J., Bogaert, P., and Heuvelink, G. B. M.: Bayesian Maximum Entropy prediction of soil categories using a traditional soil map as soft information, *Eur. J. Soil Sci.*, 59, 166–177, doi:10.1111/j.1365-2389.2007.00981.x, 2008.
- Cochran, W. G.: Sampling Techniques, 3rd Edn., Wiley, 1977.
- Davis, J. C.: Statistics and Data Analysis in Geology, 3rd Edn., Wiley, 2002.
- De Gruijter, J. J., Brus, D. J., Bierkens, M. F. P., and Knotters, M.: Sampling for Natural Resource Monitoring, Springer Verlag, New York, 2006.
- De Vries, W., Butterbach Bahl, K., Denier van der Gon, H. A. C., and Oenema, O.: The impact of atmospheric nitrogen deposition on the exchange of carbon dioxide, nitrous oxide and methane from European forests, in: Greenhouse gas sinks, edited by: Reay, D. S., Hewitt, C. N., Smith, K. A., and Grace, J., 249–283, 2007.
- De Vries, W., Leip, A., Reinds, G. J., Kros, J., Lesschen, J. P., and Bouwman, A. F.: Comparison of land nitrogen budgets for European agriculture by various modeling approaches, *Environ. Poll.*, 159, 3254–3268, doi:10.1016/j.envpol.2011.03.038, 2011a.
- De Vries, W., Leip, A., Reinds, G. J., Kros, J., Lesschen, J. P., Bouwman, A. F., Grizzetti, B., Bourauoi, F., Butterbach Bahl, K., Bergamaschi, P., and Winiwarer, W.: Geographic variation in nitrogen budgets in agricultural and other terrestrial ecosystems over Europe, in: The European Nitrogen Assessment, edited by: Sutton, M. A., Howard, C. M., Erisman, J. W., Billen, G., Bleeker, A., Grennfelt, P., van Grinsven, H., and Grizzetti, B., Cambridge University Press, Chapter 15, Cambridge, UK, 317–344, 2011b.
- De Vries, W., Leip, A., Reinds, G. J., Kros, J., Lesschen, J. P., and Bouwman, A. F.: Comparison of land nitrogen budgets for European agriculture by various modeling approaches, *Environ. Poll.*, 159, 3254–3268, doi:10.1016/j.envpol.2011.03.038, 2011c.
- Del Grosso, S. J., Mosiera, A. R., Parton, W. J., and Ojima, D. S.: DAYCENT model analysis of past and contemporary soil N₂O and net greenhouse gas flux for major crops in the USA, *Soil Till. Res.*, 83, 9–24, doi:10.1016/j.still.2005.02.007, 2005.

- EMEP: Status Report 1/09 Transboundary acidification, eutrophication and ground level ozone in Europe in 2007. Joint MSC-W & CCC & CEIP Report, Norwegian Meteorological Institute, Oslo, Norway, 2009.
- Erismann, J. W.: The Nanjing Declaration on Management of Reactive Nitrogen, *BioScience*, 54, 286–287, 2004.
- Galloway, J. N., Aber, J. D., Erismann, J. W., Seitzinger, S. P., Howarth, R. W., Cowling, E. B., and Cosby, B. J.: The Nitrogen Cascade, *BioScience*, 53, 341–356, 2003.
- Galloway, J. N., Townsend, A. R., Erismann, J. W., Bekunda, M., Cai, Z., Freney, J. R., Martinelli, L. A., Seitzinger, S. P., and Sutton, M. A.: Transformation of the Nitrogen Cycle: Recent Trends, Questions, and Potential Solutions, *Science*, 320, 889–892, 2008.
- Hartigan, J. A. and Wong, M. A.: A K-means clustering algorithm, *Appl. Stat.*, 28, 100–108, 1979.
- Kaufman, L. and Rousseeuw, P. J.: *Finding Groups in Data: An Introduction to Cluster Analysis*, Wiley, New York, 1990.
- Kesik, M., Ambus, P., Baritz, R., Brüggemann, N., Butterbach-Bahl, K., Damm, M., Duyzer, J., Horváth, L., Kiese, R., Kitzler, B., Leip, A., Li, C., Pihlatie, M., Pilegaard, K., Seufert, S., Simpson, D., Skiba, U., Smiatek, G., Vesala, T., and Zechmeister-Boltenstern, S.: Inventories of N₂O and NO emissions from European forest soils, *Biogeosciences*, 2, 353–375, doi:10.5194/bg-2-353-2005, 2005.
- Keuskamp, J., van Drecht, G., and Bouwman, A. F.: European scale modelling of groundwater denitrification and associated N₂O production, *Environ. Pollut.*, 165, 67–76, 2012.
- Kros, J., Heuvelink, G. B. M., Reinds, G. J., Lesschen, J. P., Ioannidi, V., and de Vries, W.: Uncertainties in model predictions of nitrogen fluxes from agro-ecosystems in Europe, *Biogeosciences Discuss.*, 9, 6051–6094, doi:10.5194/bgd-9-6051-2012, 2012.
- Lesschen, J. P. and Velthof, G. L.: Estimation of N₂O emission factors for soils depending on environmental conditions and crop management, in: *Proceedings of the Fifth International Symposium on Non-CO₂ Greenhouse Gases (NCGG-5)*, Wageningen, The Netherlands, 30 June–3 July 2009, Wageningen, the Netherlands, 2009.
- Li, C., Aber, J., Stange, F., Butterbach-Bahl, K., and Papen, H.: A process oriented model of N₂O and NO emissions from forest soils: I. Model development, *J. Geophys. Res.*, 105, 4369–4384, 2000.
- Maechler, M., Rousseeuw, P., Struyf, A., and Hubert, M.: *Cluster Analysis Basics and Extensions*. R package version 1.12.1, available at: <http://CRAN.R-project.org/package=cluster> (last access: 1 July 2012), 2009.
- Mitchell, T. D., Carter, T. R., Jones, P. D., Hulme, M., and New, M.: A comprehensive set of high-resolution grids of monthly climate for Europe and the globe: the observed record (1901–2000) and 16 scenarios (2010–2100), Tyndall Centre for Climate Change Research, Working paper 55., 19 pp., 2004.
- Panagos, P.: The European soil database, *GEOConnexion*, 5, 32–33, 2006.
- Simpson, D., Winiwarter, W., Börjesson, G., Cinderby, S., Ferreira, A., Guenther, A., Hewitt, C. N., Janson, R., Khalil, M. A. K., Owen, S., Pierce, T. E., Puxbaum, H., Shearer, M., Skiba, U., Steinbrecher, R., Tarrasón, L., and Öquist, M. G.: Inventorying emissions from nature in Europe, *J. Geophys. Res.*, 104, 8113–8152, 1999.
- Smil, V.: Nitrogen in crop production: An account of global flows, *Global Biogeochem. Cy.*, 13, 647–662, 1999.
- Tilman, D., Fargione, J., Wolff, B., D'Antonio, C., Dobson, A., Howarth, R., Schindler, D., Schlesinger, W. H., Simberloff, D., and Schwackhamer, D.: Forecasting agriculturally driven global environmental change, *Science*, 292, 281–284, 2001.
- Van der Velde, M., Bouraoui, F., and Aloe, A.: Pan-European regional-scale modelling of water and N efficiencies of rapeseed cultivation for biodiesel production, *Glob. Change Biol.*, 15, 24–37, doi:10.1111/j.1365-2486.2008.01706.x, 2009.
- Velthof, G. L., Oudendag, D. A., Witzke, H. P., Asman, W. A. H., Klimont, Z., and Oenema, O.: Integrated assessment of nitrogen emission losses from agriculture in EU-27 using MITERRA-EUROPE, *J. Environ. Qual.*, 38, 1–16, doi:10.2134/jeq2008.0108, 2009.