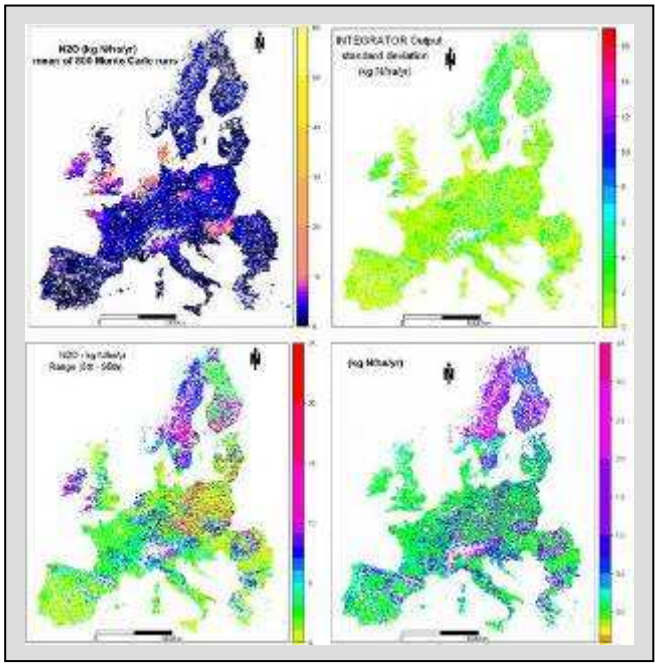
Centre for Geo-Information

Thesis Report GIRS-2009-03

Uncertainty Analysis of Predicting Terrestrial Nitrous-Oxide Emissions in Europe with INTEGRATOR

Truong Ngoc Phuong

March, 2009



WAGENINGEN UNIVERSITY
WAGENINGEN UR

Uncertainty Analysis of Predicting Terrestrial Nitrous-Oxide Emissions in Europe with INTEGRATOR

Truong Ngoc Phuong

Registration number 81 06 20 843 110

Supervisors

Dr. ir. Gerard Heuvelink
Dr. ir. Sytze de Bruin

A thesis submitted in partial fulfilment of the degree of Master of Science
at Wageningen University and Research Centre,
The Netherlands.

Date: 31 March, 2009
Wageningen, the Netherlands

Thesis code number: GRS-80436
Thesis Report: GIRS-2009-03
Wageningen University and Research Centre
Laboratory of Geo-Information Science and Remote Sensing

Acknowledgements

After more than haft year of hard working, I am now happy to present you my thesis report about uncertainty analysis of terrestrial nitrous-oxide emissions in Europe predicted with the INTEGRATOR model. I am greatly grateful to my two supervisors Gerard Heuvelink and Sytze de Bruin for their excellent instruction and their great support during the time. They both showed great enthusiasm about supervising. I am also grateful to Gert Jan Reinds for his great support. Without his help I could not successfully have carried out my work with the INTEGRATOR model and achieved that nice results for my thesis. I would like to thank Wim de Vries and Hans Kros for their explanations and providing data related to the INTEGRATOR model. I would like to thank Willy ten Haaf, study advisor for his advice during the time. Finally, I would like to thank MGI fellow students for letting me use their computers due to my extensive computational work and thank Antoinette Stoffers, GRS secretary for helping arrange the working place and colloquium.

Trương Ngọc Phương

Wageningen, 2009

Summary

*The **INTEGRATOR** model (**I**ntegrated **N**itrogen **T**ool across **E**urope for **GR**een House Gas and **A**mmonia **T**argeted to **O**perational **R**esponses) is the integrated, GIS-based, simple multi-sector, multi-component modelling tool for estimating Greenhouse Gas (GHG) emission at the European scale. It is used for tracking changes over time for assessment of the success of policy measures and the impact of land use changes on the overall European N and GHG budgets. It comprises of multiple sub-models to predict GHG emission including terrestrial nitrous oxide (N_2O), which is the target of the uncertainty analysis reported in this study. To calculate terrestrial N_2O emission from natural and agricultural land, the INTEGRATOR model uses two different models which contain different sources of error. Uncertainty about the model inputs (soil pH and soil organic carbon), model parameters (emission factors and coefficients in the relationship describing N_2O exchange rate from soil) and model structure (random residual error in the empirical model) are quantified based on defining characteristics of their probability distribution. Uncertainty propagation to the output is quantified following Monte Carlo approach. Analysis of spatial aggregation effects on uncertainty in estimated N_2O emissions are carried out by aggregating the outputs from point scale of 10 by 10 kilometres spatial interval to national scale. The main results of this study indicate that: (1)The developed statistical models are sufficient to quantify uncertainty about model inputs, model parameters and model structure; (2) Uncertainty about N_2O estimate is expressed by the standard deviation which varies from very low at 0.007 kg N/ha/yr to around 16.58 kg N/ha/yr over all examined locations; (3) Error in emission factors and the empirical model are the most important sources of error in calculating terrestrial N_2O emission by the INTEGRATOR model; (4) Spatial aggregation from point scale to national scale reduces much of the uncertainty in the estimated terrestrial N_2O emission; (5) The degree of reduction is proportional to the extent of the size of the area.*

This report consists of eight sections:

*- Section 1 (pp 5-6) is the **Introduction** part in which the context, problem definition, research objective and detailed research questions are introduced.*

*- Section 2 (pp 6-9) is about **the INTEGRATOR model**, its sub-models and the spatial scale of its input and output. Detailed explanations of how terrestrial N_2O emissions from natural and agricultural areas are calculated by INTEGRATOR are given as well.*

*- Section 3 (pp 10-16) presents the **Methodology** of quantifying model input, parameter and model structure uncertainty; quantifying uncertainty propagation; and assessment of the spatial aggregation effect on output uncertainty*

*- Section 4 (pp 17-26) presents all the **Results** of this study*

*- Section 5 (pp 26-30): **Discussions** about results are given*

*- Section 6 (pp 30-31): **Conclusions***

*- Section 7 (pp 32): **Bibliography***

*- Section 8 (pp 32-55): **Appendices** include all information related to this study, some extensive results which were not convenient to put in the Results section and all developed R code for this study.*

Table of contents

List of figures

List of tables

1 Introduction

1.1 Context and background

Today, climate change is a problem of global concern. Human activities such as industrial production and land use change cause the amount of greenhouse gases (GHG) such as carbon dioxide, methane, and nitrous oxide to increase (IPCC). This enhanced greenhouse effect is the main issue of the climate change problem. In an attempt to prevent anthropogenic climate change, the Kyoto Protocol was adopted in 1997 and entered into force in 2005 (UNFCCC 1998). The major feature of the Kyoto Protocol is that it sets binding targets for 37 industrialized countries and the European community for reducing GHG emissions (UNFCCC 1998). As stated in Article 7 of Kyoto Protocol, all signatories are required to submit annual GHG inventories on anthropogenic GHG emissions and any step taken to improve the estimates. Therefore, much research was carried out to better estimate GHG emission in Europe, among others using advanced computational models. Models for predicting GHG emissions such as the DNDC model (http://www.dndc.sr.unh.edu/index.html) and the IIASA RAINS/GAINS model (http://gains.iiasa.ac.at/gains/index.html) were already built and used in Europe and other countries.

As part of the common purpose of predicting GHG emission, the Integrated Project NitroEurope (NEU) focuses on the nitrogen cycle and its influence on the European GHG balance. The NEU project started in 2006 to address the effect of reactive nitrogen (Nr) supply on the net GHG budget for Europe. Reactive nitrogen, which comprises of ammonia, ammonium, nitrate, nitrite, nitrous oxide and others molecules is another large source of GHG, besides carbon dioxide (ifa 2007). The NEU project aims to advance fundamental understanding of C-N interactions at different scales and deliver: process-based models, landscape-level assessments, European maps of C-N pools, Nr fluxes and net GHG exchange and independent verification of GHG inventories, as required under the Kyoto Protocol (NitroEurope 2006). There are six components in the NEU project, named (1) Flux network, (2) Ecosystem manipulation, (3) Plot scale modelling, (4) Landscape analysis, (5) European Integration and (6) Verification. Amongst these components, the fifth component develops and applies GIS-based integrated agro-ecological assessment tools for the European scale to assess present nitrogen and GHG emissions, sources, sinks and their interactions with terrestrial systems; assess interactions between C and N and between agricultural and non-agricultural systems; predict past, present and future N and GHG emissions/sinks in response to various scenarios (NitroEurope 2006).

The integrated, GIS-based, simple multi-sector, multi-component modelling tool for estimating GHG emission at European scale developed within NEU is called INTEGRATOR (Integrated Nitrogen Tool across Europe for GHG and Ammonia Targeted to Operational Responses). INTEGRATOR consist of various modules, calculating Nr and GHG emissions from various sources including agricultural areas and non-agricultural areas (De Vries et al. 2008). INTEGRATOR uses detailed input data, combinations of detailed GIS based environmental data (e.g. land use, soil type, climate), farming data (farm type and agricultural management) and detailed land use changes (De Vries et al. 2008). In the NEU project, all modellers need to quantify the uncertainties of their model results and submit the uncertainty reports to NEU (Oijen 2006). Therefore, uncertainty quantification of estimated Nr and GHG emission is included in the INTEGRATOR model as an integral part. The requirement of uncertainty analysis is to ensure that INTEGRATOR provides accuracy measures together with the estimates of GHG emission. Uncertainty assessment is also an important element of component six of the NEU project.

1.2 Problem definition

INTEGRATOR was built in order to gain highly detailed inventories of GHG emission over Europe. However, the reliability of the predictions of GHG emission is equally important information when the results of the model are submitted under the Kyoto Protocol requirement or used by decision or policy makers. Scientists nowadays pay more consideration to the uncertainty of GHG emission estimates because uncertainty estimates are necessary to achieve the complete GHG emission inventories (IPCC 2000). The knowledge about model uncertainties reflects the reliability of GHG emission inventories. Therefore, one of the objectives of the NEU project is to assess the uncertainty of the INTEGRATOR model results and use these together with independent measurement/inverse modelling approaches for verification of European GHG inventories (NitroEurope 2006).

Being a model, INTEGRATOR model cannot produce absolutely certain results. Different sources of uncertainties can be categorized into four main sources: (1) input uncertainty; (2) model parameter uncertainty; (3) model structure uncertainty and (4) model solution uncertainty (Heuvelink 2008). In the context of this research, uncertainty is defined as an interval around a value such that any repetition of estimating this value will produce a new result that mainly lies within this interval.

The input data used in INTEGRATOR comprise of various attribute data and geographic data in order to enhance the quality of GHG emission predictions at European scale (De Vries et al. 2008). These input data were obtained from different sources, at different levels of detail and scale over Europe. INTEGRATOR uses soil pH and soil organic carbon (OC) as inputs to predict N_2O emission from natural and agricultural areas. The problem is that mapping soil properties is not perfect and always contains errors because these maps are typically based on limited knowledge and limited information (Brus and Heuvelink 2007). Uncertainty assessment for soil data is increasingly necessary not only for better soil data management but also as the preliminary step of further applications (Goovaerts 2001; Brus and Heuvelink 2007). Continuous soil properties, such as pH and OC always show spatial variation even within perfectly pure soil map units (Webb and Lilburne 2005; Brus and Heuvelink 2007). If values of pH and OC in input maps are assigned constant values to map units, they cannot correctly represent the realistically spatial distribution of pH and OC values over continuous space (Brus and Heuvelink 2007). Thus, when soil pH and OC are interpolated over Europe to obtain spatially distributed maps of pH and OC, these pH and OC maps are never completely free of error because of interpolation error.

Moreover, uncertainty of the model itself is also an important issue. Uncertainty of INTEGRATOR outputs can be caused by the uncertainty in the model itself. INTEGRATOR uses two different approaches to estimate N_2O emission from agricultural and natural areas: a simple model using emission factors for estimating emission from agricultural areas and a linear regression model for emission from natural areas. The uncertainty can then be caused by: 1.uncertainty of parameters used, here these are uncertain emission factors and uncertain regression coefficients, 2. uncertainty of model structure, here represented by the regression model residual. Model input, parameter and model structure uncertainties can all be important contributory sources to model output uncertainties (Heuvelink 2008).

For this reasons, quantification of INTEGRATOR inputs, model parameters and model structure uncertainties and their propagation is essential to provide information on the reliability of INTEGRATOR results and possible improvement of the model estimation accuracy. In addition, the magnitude of estimation uncertainty of N_2O soil emission also varies with scale (Heuvelink 1998). It is interesting to evaluate spatial aggregation effects on prediction uncertainty.

In this research, these questions of how to quantify model input, model parameter and model structure uncertainties and their propagation through INTEGRATOR; how spatial aggregation affects on prediction results and whether any improvement can be possible to enhance the accuracy of prediction results were addressed. We restrict ourselves to the prediction accuracy of N_2O soil emission from agricultural and natural area.

1.3 Research objective and research questions

Objective

The objective of this research is to assess uncertainty about terrestrial N_2O emission estimated by the INTEGRATOR model from agricultural and natural land over Europe, by quantifying uncertainty of model inputs, model parameters and model structure, and analyse their uncertainty propagation. The focus is on uncertainty assessment of N_2O emission factors and environmental factors (soil pH and OC), uncertainty of the empirical model and analysis of spatial aggregation effects on uncertainty.

Research questions

1. How can uncertainty about INTEGRATOR model inputs, model parameters and model structure including soil pH and OC, N_2O emission factors, empirical model for estimating N_2O from natural areas be quantified, taking spatial correlation and cross-correlation into account?
2. To what extent do the uncertainty about these inputs and model propagate through INTEGRATOR and does the magnitude of uncertainty propagation vary with different inputs, model parameters and model structure?
3. How do the uncertainty propagation results depend on the spatial aggregation level of the output employed in the analysis?
4. Uncertainty magnitude of which inputs, model parameters or model structure should best be reduced to substantially reduce uncertainty of the predicted N_2O emission from agricultural and natural area over Europe?

2 Estimating European N_2O soil emission with INTEGRATOR

INTEGRATOR is defined as a modelling tool for European-wide assessments of nitrogen and GHG fluxes in response to changes in land use cover, land management and climate in the past, present and future (De Vries et al. 2008). It consists of simplified process-oriented and empirical modules calculating N and GHG (NH_3, NO_x, N_2O, CO_2 and CH_4) emissions, as illustrated in Figure 2-1.

These modules are:

(i) The adapted MITERRA-Europe model: estimating NH_3, NO_x, N_2O and CH_4 emissions from housing and manure storage systems and agricultural soils.

(ii) Empirical model using empirical relationships for estimating NO_x, N_2O and CH_4 emissions from non-agricultural terrestrial systems.

(iii) The YASSO soil model: estimating CO_2 emissions from agricultural and terrestrial systems, in combination with EFISCEN for forest, MITERRA-Europe for agriculture and empirical relationships for peat lands.

(iv) Estimating N deposition using an emission-deposition matrix for NH_3 and NO_x, accounting for the interaction between agricultural and non-agricultural soils.

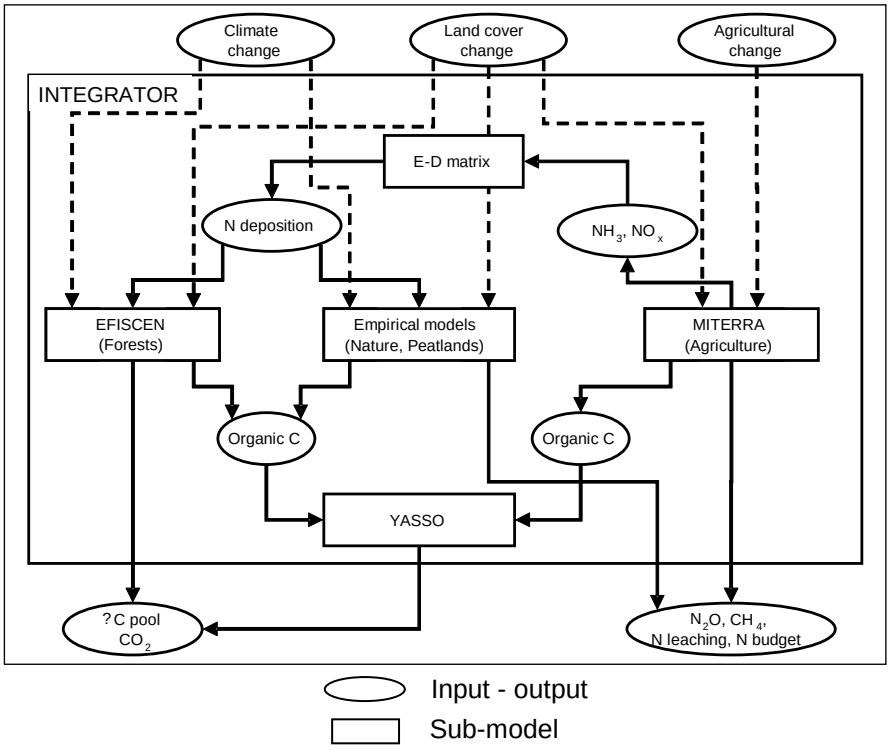


Figure 2-1: The INTEGRATOR model: its drivers, main modules and main outputs
Source: (De Vries et al. 2008)

INTEGRATOR was developed in such a way that its sub-models are modular and exchangeable. This means that any model component can be simultaneously applied for uncertainty assessment (De Vries et al. 2008).

Only sub-models of INTEGRATOR which were used for estimating N_2O soil emission from agricultural and natural areas were included in this study.

2.1 How does INTEGRATOR estimate N_2O emission from soil?

As mentioned above, this research only focused on sub-models of INTEGRATOR which were used for estimating N_2O soil emission from agricultural and natural areas. Table 2-1 indicates which land use/cover types were considered as natural and agricultural areas.

Table 2-1: Land use/cover of natural and agricultural areas

Area	Land use type/cover
Natural areas	Forest, heather and moorlands
Agricultural areas	Arable land – including: arable land (irrigated and non-irrigated), permanent crops, bio-fuel crops Grass land – including: pasture, recently abandoned arable land, natural grassland, (semi-) natural vegetation

A map of agricultural and natural areas is given in Figure 2-2.

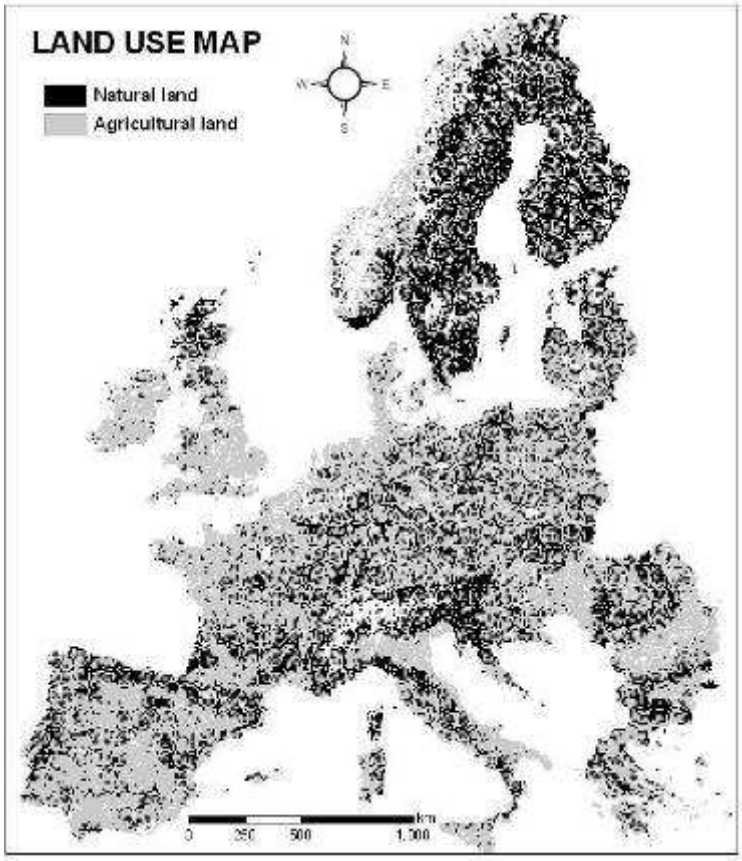


Figure 2-2: Map of natural and agricultural areas

Different approaches were used to predict N_2O soil emission from agricultural and natural areas. The following part will explain in detail these approaches.

For agricultural area: N_2O emission from soil was calculated by a simple empirical emission factor (EF) approach, accounting for major environmental factors controlling N_2O emission (De Vries et al. 2008). N_2O emission was calculated as a fraction of the amount of nitrogen input to soil as equation 2-1:

$$N_2O \text{ soil emission} = \text{Emission Factor (EF\%)} \times \text{Total N-input} \qquad 2\text{-}1$$

Values of EFs vary amongst different types of nitrogen input sources and are affected by environmental factors, such as pH, precipitation, temperature, soil texture/soil type. In the case of pH, values of EFs for pH ≥ 5 were different from those for pH < 5. Four different reference situations were used to set up the average EFs as four reference average EFs (x%). The four reference situations were based on the sources of nitrogen input to soil, including fertilizer, manure, soil organic nitrogen and crop residuals. The difference in types of fertilizer, types of crop, manure, Nitrogen fixation process and the impact of environmental factors controlling N_2O emission on the emission factors were estimated by multiplying the emission factor with a certain fraction (α). The factor α can be larger (larger EFs than the reference) or smaller (smaller EFs than the reference) than one (Velthof 2008). Values of EFs were estimated values based on literature study and

expert knowledge. The average values were used in INTEGRATOR, but there is a large variation in these values themselves and over space (Velthof 2008).

For natural area: a linear regression model was used to calculate N_2O as a function of many factors as in equation (2-2):

$$Log\ N2O = \beta 0 + \beta 1 * P + \beta 2 *(Fraction\ T<0) + \beta 3 *log(\ Ndep) + \beta 4 * (pH) + \beta 5 * (Organic\ C) + \beta_6 * (vegetation: deciduous) \quad 2\text{-}2$$

Where:

- β_i are regression coefficients
- Climate variables:

  P is mean monthly precipitation over the measuring period (mm)

  Fraction T<0 = Fraction of months with minimum T<0°C over the measuring period
- Soil variables as derived from WISE & SPADE databases, averaged over the layer 0-20 cm

  pH is pH H_2O (all different pH values are converted to pH H_2O using regressions)

  Organic C is OC content (g/kg soil)
- Deposition variable as derived from European Monitoring and Evaluation Programme (EMEP)

  Ndep is Nitrogen deposition (kg N ha^{-1} yr^{-1})
- Vegetation types include

dec	deciduous forest
con	coniferous forest
sv	short vegetation (incl. heath and grass)
mix	Mixed woodland (con+dec)

INTEGRATOR can predict N_2O emission for different scenarios/policies which take into account the effects of land use change, climate change and emission reduction methods. In developing state, INTEGRATOR can be implemented in two scenarios, A1 which is for changing at global scale and B2 for changing at regional scale (De Vries et al. 2008). N_2O soil emission can be estimated from 1970 to 2030 with 10 year intervals. Because INTEGRATOR is still in development, the study was only carried out on analysing uncertainty of the model results for the year 2000 and for scenarios A1.

2.2 Characterizing spatial scale of INTEGRATOR inputs and outputs

The INTEGRATOR model indirectly operates on NEU calculation units (NCUs) (De Vries et al. 2008). NCUs resulted from the combined partitioning of a number of input maps, based on the requirement of processes included in INTEGRATOR and their dependence on site and soil characteristics. As such, NCUs consist of unique combinations of land use, soil type, agricultural structure.

Table 2-2: INTEGRATOR input and output scale

Emissions from soil	Units
N_2O emissions	kg N ha^{-1} $year^{-1}$
Spatial resolution	NCU (ca 10 X 10 km^2)
Temporal extent	1970-2030 (10 year interval)
Spatial extent	EU27+ Norway +Switzerland +Croatia No oceans and territorial waters included

Model inputs and parameters were prepared as grid maps of 10 by 10km resolution, covering Europe. Every cell in the grid map was assigned an ID. Every cell in these grid maps were then linked to NCUs according to their location, using ID numbers. INTEGRATOR accesses each grid cell by their ID numbers and N_2O were estimated for each grid cell. Hence the output from INTEGRATOR is also a grid map of 10x10 km resolution.

3 Methodology

The Monte Carlo (MC) uncertainty analysis approach was used to analyze uncertainty of N_2O soil emission estimated by the INTEGRATOR model at specific locations, which are the nodes of 10 by 10 km grid mesh covering the whole of Europe. The idea of the MC method is to repeatedly compute results of the model, with inputs that are randomly sampled from their probability distributions (Heuvelink 1998). Inputs can be model input and/or model parameter and/or error in model structure. The model outputs form a random sample of the output probability distribution. Analyzing this sample distribution by computing its mean and its variance represents the level of uncertainty about model output, provided the sample is large enough.

As described in section 2-1, the INTEGRATOR model uses soil pH and OC as environmental factors which affect N_2O soil emission from both agricultural and natural areas. Maps of pH and OC were expected to show significant error. EFs were also expected by experts to have inaccurate values. Similarly, the regression model for estimating N_2O from natural area is also not error-free.

To quantify the uncertainty of INTEGRATOR output, the methodology consisted of three main steps, as illustrated in Figure 3-1:

- **Step 1:** quantifying model inputs (pH and OC), parameters (EFs, coefficients in relationship describing N_2O exchange rate from soil) and model structure uncertainty (random error in regression model) by defining probability distributions. Realizations were then randomly drawn from their defined distribution, taking spatial correlation and cross-correlation into account.
- **Step 2:** Monte Carlo runs for INTEGRATOR: it is first needed to determine how many MC runs are needed to obtain stable results, and next the MC runs with INTEGRATOR were carried out using the simulated values and maps of all inputs and parameters obtained in Step 1.
- **Step 3:** result from MC runs from Step 2 were analyzed to examine: (i) total uncertainty propagation to model output; (ii) contribution of model inputs, model parameters and model structure uncertainty to uncertainty of model output; (iii) spatial aggregation effect on output uncertainty.

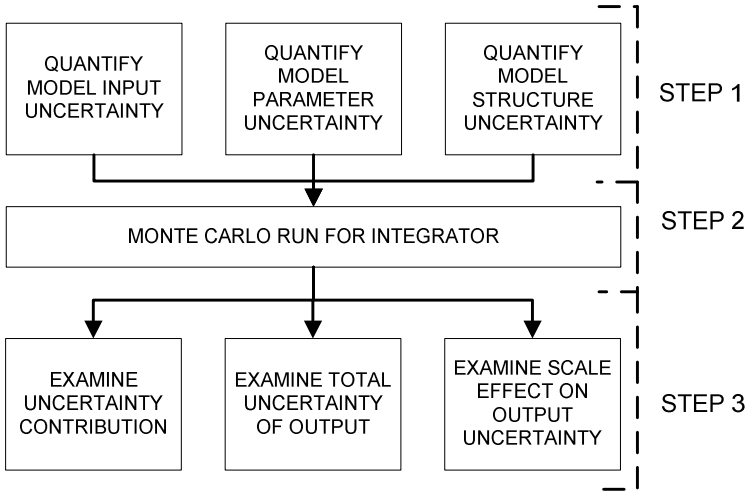


Figure 3-1: Flow chart of methodology

3.1 Input uncertainty quantification: Soil pH and OC

3.1.1 pH and OC sampled data

Soil pH and OC were obtained from two datasets: WISE 1.1 (Global Soil Profile Database - ISRIC) and SPADE 1 (Soil Profile Analytical Database for Europe v1.0) with sparse numbers of measured profiles in Europe (521 and 496 respectively) (Maree 2007). For this study, the available dataset comprised of 301 pH observations and 472 OC observations. Values of pH and OC vary in different soil types (R. J.A. Jones et al. 2004). Because of the small number of sampling points over Europe, most soil types did not have sufficient number of observations for further estimation. Therefore, soil types were grouped into clusters to ensure that every cluster has at least five observations per cluster. Clusters of sample points were formed based on soil texture (sand, loam, clay, organic, volcanic and unknown), pH class (calcareous, eutric, dystric), soil wetness

(wet or dry), organic matter content (high or low/unknown), climatic regions and land use types (Maree 2007). The three land use types that were distinguished are: arable, agricultural grass and nature. The soil types and observations were already classified into clusters and were kindly provided by the INTEGRATOR Group (for more information see (Maree 2007)). These observations were used as measured data for further spatial conditional simulation. Figure 3-2 presents a map of pH and OC sampled locations over Europe.

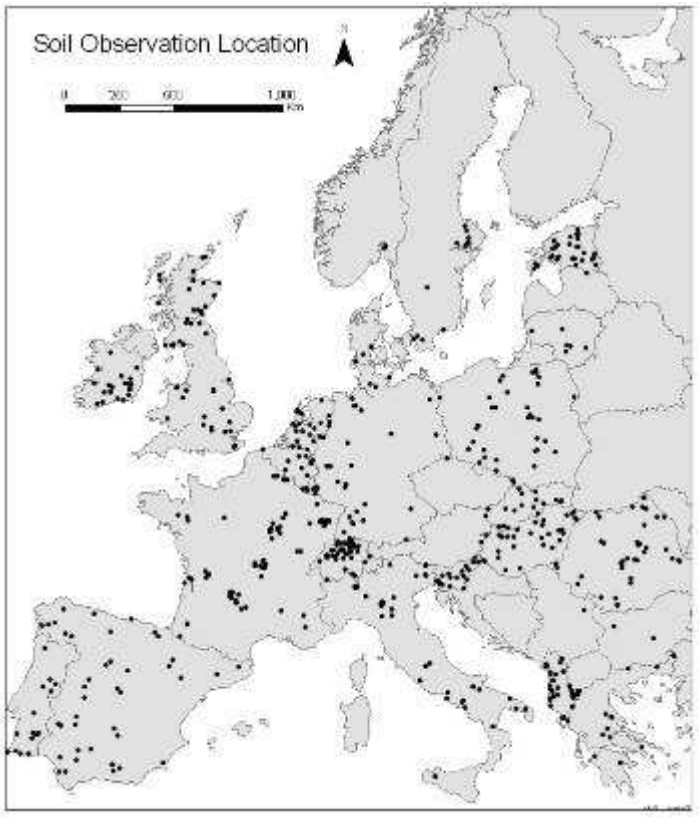


Figure 3-2: Soil Observation locations

Corresponding to each cluster of soil samples is map of cluster unit. ASCII grid files for cluster maps were created using ArcGIS, based on the 1:1 Million scale soil map of Europe in consideration of relation between the soil types on this map and the sample soil clusters**.** Nature, arable land and agricultural grass were the three main land use cover in Europe which were assessed separated. Certain numbers of clusters were assigned for each kind of land use in the way that these clusters of each land use cover whole Europe. Clustering step were already finished and kindly provided by the INTGERATOR Group. Figure 3-3, for instance, is the map of twelve clusters used for nature land.

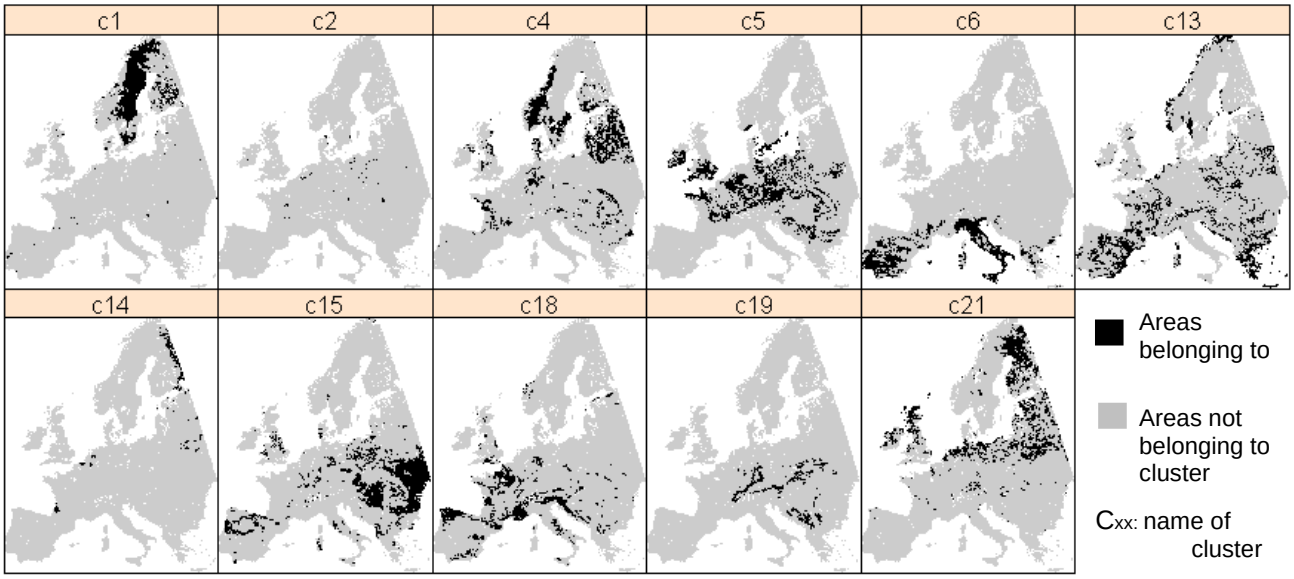


Figure 3-3: Soil cluster for nature land using to simulate OC

As mentioned above, each of land use cover type was assumed to cover whole Europe. The reason for this assumption is that INTEGRATOR assesses the effects of changes in land use cover in time on N_2O emission. Each location in Europe can have one of these three main land use types at different time. Each land use type has different pH and OC value. Thus, pH and OC were first simulated for all three different land use which cover whole Europe. Then, depending on the land use map for specific year, pH and OC values were assigned to specific location according to the land use type at that location.

3.1.2 Defining Probability distribution functions

Because the Monte Carlo method is used to analyze uncertain outputs at specific locations, multiple realizations of uncertain inputs and parameters needed to be achieved by sampling from their probability distribution. At each examined location, the exact values of pH and OC could not be known but the range of their value could be defined through their probability distributions. These probability distributions represent the uncertainty about the value at that location. The probability distribution function (pdf) of pH and OC could be achieved from observations obtained from WISE 1.1 and SPADE 1 data. Geo-statistical modelling, as explained below was used to obtain these pdfs.

The distributions of the observation data for pH and OC were examined by visualizing their histograms. The mean, quartiles and outliers (if any) of each cluster were also examined using box plots. According to previous research, soil OC is positively skewed, which has the shape of log normal distributed. Transformation to normal distribution was then carried out for OC data using logarithmic transformation.

R and R-Gstat were used to do this step.

3.1.3 Stochastic simulation of uncertain soil pH and OC

pH and OC do not have a constant mean over space but these vary with soil type. This means that the value of pH and OC would also vary between clusters. Thus, Regression Kriging (RK) (Hengl et al. 2007) may be used to simulate multiple realizations of pH and OC over European countries with cluster as independent variables of RK.

The model for estimating pH and OC value was established as:

$$Y = \mu(xi) + \sigma(xi).\varepsilon \qquad 3\text{-}1$$

where: - Y: soil pH or OC value

- μ: per-cluster mean, which are constant in each cluster but different amongst clusters

- σ: per-cluster standard deviation, which are assumed constant in each cluster but different amongst clusters. However, the variation pH (and OC) value cannot be equal for all locations within a cluster. To treat this difference, the constant standard deviation was multiplied by an error, named ε.
- ε: residual error which varied at each examined location and was assumed to have mean of zero and unit variance, and second order stationary. Assumption of second order stationary means that mean of ε is constant in space and its sermivariance only depends on distance.
- x_i: cluster i

Value of ε was simulated at each examined locations. Because its mean and its variance were known, Simple Kriging was used with known mean and variance. These steps were followed:

i. Carried out Logarithmic transformation for OC.

ii. Standardization step: The purpose of this standardization are first obtained the mean and standard deviation of each cluster and then compute the error value ε from these cluster mean and cluster standard deviation.

$$\varepsilon = \frac{Y - \mu(xi)}{\sigma(xi)} \qquad 3\text{-}2$$

iii. Analysis of spatial correlation of the error using variogram analysis: compute and fit the model to the experimental variogram of ε from observations. The experimental variogram was defined using the *variogram* function and fitted by the *fit.variogram* function with the sill value were fixed to one, using R-Gstat (Pebesma 2004).

iv. Conditional Gaussian simulation: *predict.gstat* function of R-Gstat was used to simulate ε at each node of 10 by 10 km grid mesh.

v. Using equation (3-1) to compute value of pH and OC from ε.

vi. Back transformation for OC.

Soil OC was measured in percentages. The OC value at each specific location that was randomly drawn from its lognormal distribution would have minimum value of zero but its maximum value can sometimes be larger than 100%. In such case, the simulated value was replaced by the theoretical maximum values i.e. 100%.

As mentioned above, three land use types were distinguished, namely grass land, arable land and nature land. pH and OC were simulated on the 10 by 10 km grid mesh over the whole of Europe for all three land uses. Next, the land use map of the year 2000 (Appendix 5) was used to define the land use type at each specific location; then, the value of pH and OC corresponding to this land use type was assigned to this location. Finally, we got simulated maps of pH and OC, including three different kinds of land use.

R – Gstat was used for building geostatistical modeling (See Appendix 6). The statistical model used for simulating pH for nature was already established and provided by INTEGRATOR Group.

3.2 INTEGRATOR model parameters and model structure uncertainty quantification

3.2.1 Uncertain N_2O emission factors

For agricultural land, the amount of N_2O emission was estimated as a fraction of the total nitrogen applied to soil. The average values of EFs were used in INTEGRATOR but there is large variation in these values (Velthof 2008). The data about uncertainty of EFs provided by the INTEGRATOR Group included mean values and values of 5th percentile and 95th percentile for each of the four reference EFs. These values were obtained based on previous research and literature study (Velthof 2008). According to previous research, the overall distribution of EFs was lognormal or Weibull (Dobbie and Smith 2003; RTI-International 2007). In this study, these EFs were assumed to have lognormal distribution.

Log-normal distribution can be defined by two parameters, its mean and its standard deviation. These two parameters were estimated from data available, namely the mean, 5th percentile and 95th percentile for each EF. The mean was already known. To achieve the standard deviation which should best fit for all three values above, the approach was described as below:

Call Y the emission factor, that has a lognormal distribution, so X=ln(Y) has a normal distribution, X ~N(μ,σ^2)

We need to estimate the two parameters σ and μ to get the structure of this distribution.

The parameters we already know:

5 percentile of Y = p5Y → 5 percentile of X = p5X=ln(p5Y)

95 percentile of Y= p95Y → 5 percentile of X = p95X= ln(p95Y)

$$\text{Mean of Y} = \text{mY} = \exp(\mu + \frac{\sigma^2}{2}) \rightarrow \mu = \ln(mY) - \frac{\sigma^2}{2}$$

This means that only σ had to be estimated.

We chose σ which satisfies this equation: $\int_{5pX}^{95pX} f_X(X) = 0.9$

$$\Rightarrow \int_{5pX}^{95pX} f_X(X) = F(95pX) - F(5pX)$$

$$= P(\frac{X - [\ln(mY) - \frac{\sigma^2}{2}]}{\sigma} < \frac{p95X - [\ln(mY) - \frac{\sigma^2}{2}]}{\sigma}) - P(\frac{X - [\ln(mY) - \frac{\sigma^2}{2}]}{\sigma} < \frac{p5X - [\ln(mY) - \frac{\sigma^2}{2}]}{\sigma}) = 0.9$$

The optimal σ can be obtained by trial and error or by plotting G as a function of σ,

$$G(\sigma) = \left[P(\frac{X - [\ln(mY) - \frac{\sigma^2}{2}]}{\sigma} < \frac{p95X - [\ln(mY) - \frac{\sigma^2}{2}]}{\sigma}) - P(\frac{X - [\ln(mY) - \frac{\sigma^2}{2}]}{\sigma} < \frac{p5X - [\ln(mY) - \frac{\sigma^2}{2}]}{\sigma}) \right] = 0.9 \quad \textit{(3-3)}$$

The value of σ was defined as that value for which G equal to 0.90.

After the parameters of the distribution are obtained, EF values were simulated for use in the MC runs. When a parameter was treated as uncertain, the assumption can be that they are constant over space or variable in space (Heuvelink 2008). For the EFs, we treated them as variable in space, using stochastic simulation with the assumption that there is no spatial correlation. It means the variogram was fitted with a nugget model. Simulated map of each EF were created for each of the 10 by 10 km mesh of points. These simulated maps were used by INTEGRATOR as inputs. At each location, which of the four EFs to be used was chosen by INTEGRATOR.

R – Gstat was also used in this step (See Appendix 6).

3.2.2 Uncertain empirical model for natural area

To estimate N_2O soil emission from natural areas, a multiple linear regression model was used to relate logN_2O as dependent variable Y to multiple variables as predictors X_i. Data for fitting the regression model were also provided by the INTEGRATOR Group. The data included 121 records both with measured and estimated data. The estimates were gathered from several databases and values were assigned based on location. The fitted regression model cannot perfectly represent the real relationship between these variables, thus a residual ε was used to represent the deviations of the linear model from the real relationship between the dependent Y and predictors X_i

Equation (2-2) can be briefly represented by:

$$Y = \beta X + \varepsilon$$

Where: β is the vector of regression coefficients β_i

X is the vector of predictors X_i

The uncertainty of the regression model can be separated into uncertainty in the regression coefficients and uncertainty of the regression residuals. For all locations, the value of the regression coefficients (RCs) were

assumed constant, but their values were uncertain. The distribution of β is a multivariate normal distribution. To access the uncertainty of these RCs, multiple realizations were sampled from this multivariate normal distribution. We followed a common method for drawing a random vector from the N-dimensional multivariate normal distribution with mean vector μ and covariance matrix Σ as below:

$$\beta \sim N(\mu,\Sigma)$$

i. The mean vector μ was obtained from fitting the linear multiple regression model
ii. Σ is the correlation matrix obtained from fitting the regression model. Compute the Cholesky decomposition of Σ using *Chol* function in R was carried out to find the unique lower triangular matrix A such that $AA^T = \Sigma$.
iii. Compute N independent, random, standard normal variables $Z = (z_1, \ldots, z_N)$
iv. β was then sampled from $\mu + AZ$.

The regression residual (RRes) was assumed to have a normal distribution with mean of zero and variance of the fitted residuals. These were assumed variable in space, with spatial correlation. A variogram was used to analyze their spatial correlation. The variogram was computed using the residuals obtained from fitting the regression model on 121 records. Multiple realizations of RRes were created by using stochastic simulation.

R – Gstat was also used in this step (See Appendix 6).

3.3 Monte Carlo approach for quantifying uncertainty propagation

Monte Carlo runs of INTEGRATOR were implemented on all locations of the 10 by 10 km mesh point covering Europe. Because of numerous MC runs, an R-script was written (see Appendix 6) to call INTEGRATOR in batch mode.

Table 3-1 summaries the number of model inputs and model parameters, of which the uncertainties were addressed in this step.

Table 3-1: INTEGRATOR uncertain inputs, parameters and model structure

Inputs/ Parameters	***Units***	***Use in INTEGRATOR model***
Emission factors	%	Model parameters for estimating N_2O soil emission from agricultural areas
Regression coefficients	-	Model parameters for estimating N_2O soil emission from natural areas
Regression residual	-	Error from the model used for estimating N_2O soil emission from natural areas
Soil pH	-	Input data for estimating N_2O soil emission from both agricultural and natural areas
Soil OC	%	Input data for estimating N_2O soil emission for natural areas

The strategy for MC run was established to quantify their uncertainty propagation and their uncertainty contribution to the outputs. Figure 3-4 illustrates the MC run strategy. We first carried out the so-called default run using the expected values of all uncertain items as its inputs. The results of the default run were used as a reference result to compare with results of other runs in which one or all of the inputs were sampled from their probability distributions.

Analyze contribution of each uncertain input			
DEFAULT RUN: AVERAGE VALUES OF ALL INPUTS AND PARAMETERS			
pH	OC	EMISSION FRACTIONS	REGRESSION MODEL
pH	**OC**	EMISSION FRACTIONS	REGRESSION MODEL
pH	OC	**EMISSION FRACTIONS**	REGRESSION MODEL
pH	OC	EMISSION FRACTIONS	**REGRESSION MODEL**
pH	**OC**	**EMISSION FRACTIONS**	**REGRESSION MODEL**
Analyse uncertain output			

FIXED

UNCERTAIN

Figure 3-4: Monte Carlo run strategy

Totally five MC runs were carried out. For each of the inputs and parameters, its uncertainty contribution was estimated by making each input uncertain for a MC run; while the remaining inputs and parameters were assumed certain and kept fixed. The final case is an MC run with all uncertain inputs and parameters to quantifying total uncertainty propagation to output.

To estimate the required number of MC runs, variation of the outputs were examined. 100 MC runs were first executed and then 100 more MC runs with another 100 realizations of inputs and parameters were executed. The standard deviation for each grid cells of the first 100 run outputs was plotted versus that of the second 100 run outputs using a scatter plot. The procedure was iterated for 200 and 400 MC runs. The idea of this examination is that when the number of MC runs increases, the scatter plot was expected to come to a diagonal line. This is due to the difference was less significant or the variation of outputs come to stable. The number of MC runs is sufficient when the variation of the output is stable and the scatter plot points are sufficiently close tot the diagonal line.

The following parts describe how the results of the MC runs were analyzed to address total uncertainty and relative uncertainty contribution.

a) Total uncertainty analysis

The outputs of the MC runs where all uncertain inputs and parameters were taken into account were used in this case. Summary statistics using variance and range parameters were used to show the uncertainty interval of INTEGRATOR outputs. Maps of these statistical parameters were used for visualization.

b) Relative uncertainty contribution analysis

As mentioned in section 2-1, INTEGRATOR uses different approaches to estimate N_2O emission from natural and agricultural land. Each inputs and parameters, therefore separately contributed to the output uncertainty for different land uses. Figure 3-5 shows the contributions of each input and parameter. Soil pH and EFs were used in estimating N_2O emission from agricultural land, so did their uncertainty contributions. Meanwhile, uncertain soil OC and pH and uncertainty in the regression model contributed to uncertainty of N2O emission estimated from natural areas.

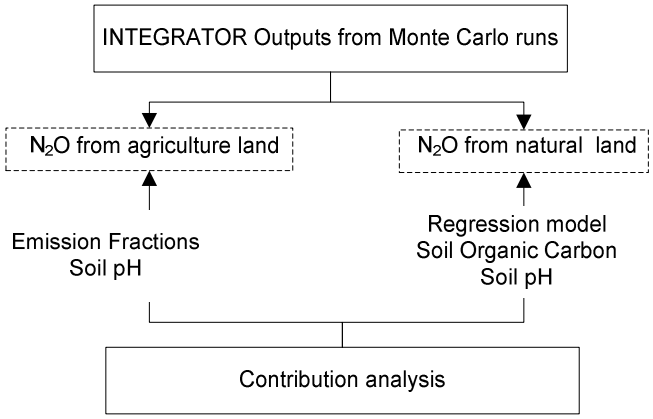


Figure 3-5: Contribution of inputs, parameters and model structure uncertainty analysis

To analyze relative contributions, the percentage ratio between the variance caused by each uncertain inputs/parameters and the total variance caused by all uncertain inputs/parameters were computed. Mapping this percentage ratio was used for visualizing the uncertainty contributions.

3.4 Assessing influence of spatial aggregation on outputs' uncertainty

In addition to the importance of uncertainty quantification for model outputs, differences in spatial scale of the outputs is also an interesting issue, because the accuracy of predictive model outputs will vary with their spatial scale (Heuvelink 1998). In this section, the question how the change of spatial aggregation level affects the uncertainty of N_2O prediction is addressed. To investigate change in output uncertainty at different spatial scales, two different scales of the outputs were examined: (1) point scale (at 10x10 km resolution) and (2) National scale.

The outputs at point scale were obtained from the MC run of INTEGRATOR as the results of section 3.3. Only the MC runs of all uncertain inputs and parameters were chosen as the output to examine the influence of spatial scale on uncertainty. Spatial aggregation were carried out for each MC run by simply computing the average of all N_2O value of all points within each country (Heuvelink 2008), follow the following formula.

$$\overline{V}_{ctr} = \frac{1}{|A|}\int_A V(x)dx$$

Where: - $\overline{V}_{ctr}$ is the desired output at national scale for each MC run

- A is the country's area
- V(x) is the outputs at point scale within each country obtained from each MC

Then, the variance of $\overline{V}_{ctr}$ computed from 800 MC runs for each country: Varctr = var ($\overline{V}_{ctr}$).

These variances were compared with the variance at point scale, with its mean: Varpnts = mean(var(V(x))).

The comparison of the two variances showed the effect of differences in the INTEGRATOR output scale on uncertainty.

Figure 3-6 summaries the analysis approach.

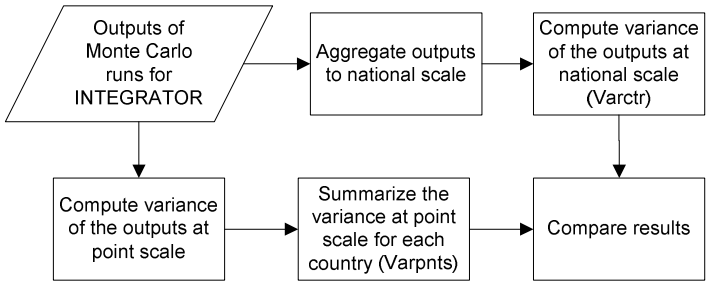


Figure 3-6 Scale effect analysis strategy

4 Results

4.1 Results of quantifying input uncertainty

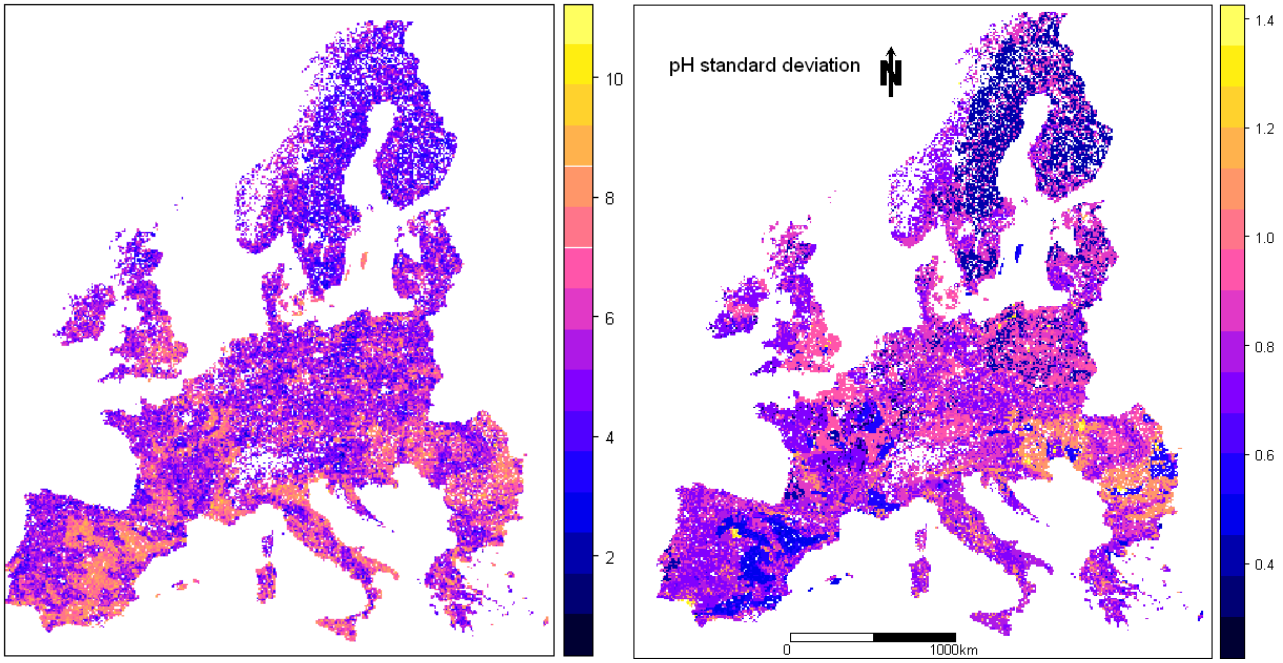
4.1.1 Soil pH uncertainty

Result of analysing pH sampled data including figures of histogram, box plot and fitted variograms of pH sampled data used for three kinds of land use cover are given in Appendix 1-2. Examining the histogram of pH sampled data shows that the histogram does not have a common distribution shape. The pH value for all clusters ranges from 4 to 8. Examining the box plot of pH value for each cluster shows that the mean value

of pH of each cluster is different. The range from lower quartile (25th percentile) to upper quartile (75th percentile) is also various.

As mentioned in section 3-1-3, instead of simulating pH value, ε was simulated. The histogram shows that the ε has normal distribution with the mean is almost equal to 0. The fitted variograms of the ε for three different kinds of land use show that all fitted variograms have the nuggets around 0.4; the ranges range from 30 to 50 kilometres; and the sills were fixed to 1.

The error was simulated at points of 10x10 km interval over Europe and was then transformed back to pH values. Figure 4-1-a is a single simulated map of pH value for the locally present land cover type. Value of simulated soil pH ranged between 1 and 11. When compared to Figure 2-2 it can be observed that pH value in natural areas (mainly forest) are typically smaller than pH value in agricultural areas (mainly arable and grass land). In natural areas, the pH value mostly ranges from 2 to 4, while in agricultural land, pH is mostly larger than 4. Most areas in Spain, France, South-East United Kingdom, Italy, Hungary, Romani, and Bulgaria show high pH value. pH in Sweden and Finland is smaller than the other areas.



(a) *(b)*

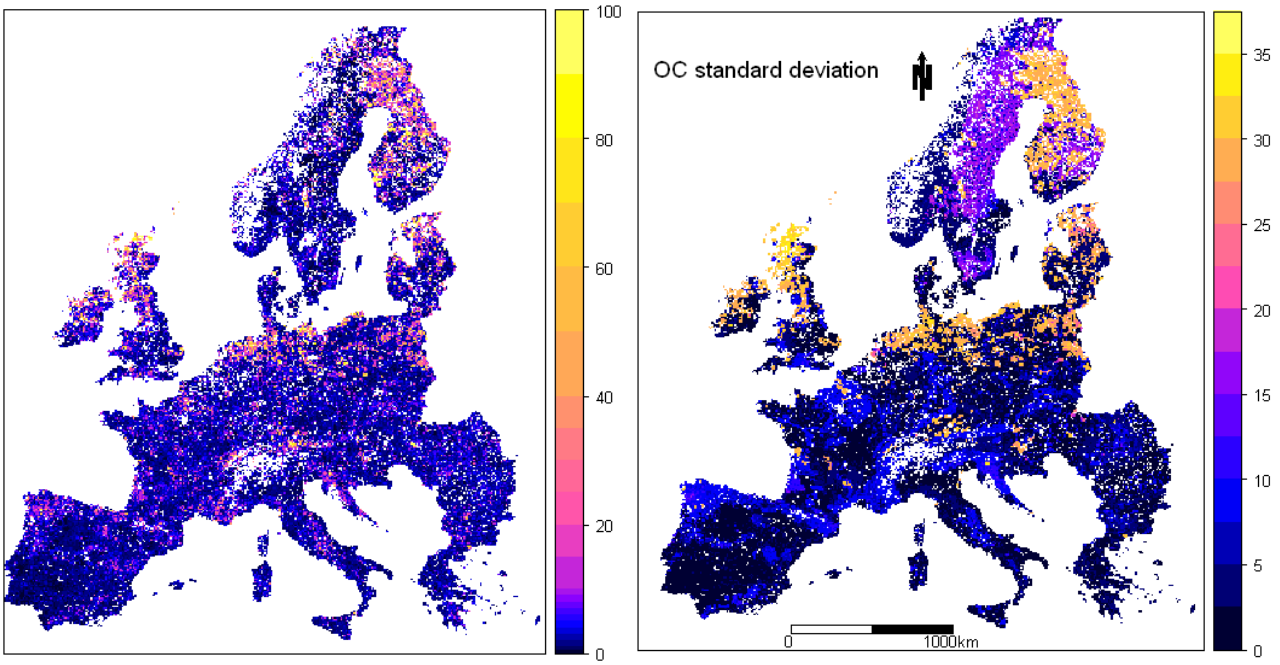
Figure 4-1: (a) Single simulated map of pH for three types of land use and (b) Map of pH standard deviation

Figure 4-1-b shows the standard deviation of simulated pH value over Europe. The standard deviation of simulated pH value over Europe is minimum at 0.3 and maximum at 1.4. When a comparison with Figure 2-2 is made, it can be seen that pH value has larger variation in agricultural area than natural area. In Finland and Sweden, pH variation is lowest. Highest variation can be seen in areas of Hungary, Romania and Bulgaria and small parts in Spain.

4.1.2 Soil OC uncertainty

Unlike pH value, the histogram of OC value shows strongly positive skewed distribution. Most of OC sampled value ranges from 0 to 5%. The box plot of OC value of each cluster also indicates the variation in OC mean value and variance amongst different clusters. The fitted variograms of the ε for three kind of land uses show that for all three land use, the variograms were fitted with the sill fixed to 1, the ranges are from 150 to 200 kilometres, and the nuggets vary from 0.3 to 0.5.

Figures of histogram, box plot and fitted variograms of OC sampled data for three kinds of land uses can be found in Appendix 1-3. Maps in Figure 4-2 show an example simulated map of OC value (a) and variation of simulated OC values for three types of land use (b).



(a) *(b)*

Figure 4-2: (a) Single simulated map of Organic carbon (%) for three types of land use and (b) Map of its standard deviation

The simulated values of OC vary between 0% and 100%. The example simulated map shows the largest OC values in areas of Finland, Sweden, North of United Kingdom, Ireland, Estonia, Latvia, Poland, North and South of Germany, some parts of Denmark and Austria. The lowest OC values were simulated in Spain and Portugal. Considering the difference in land use/cover (Figure 2-2), the map generally indicates that OC value is higher in natural than in agricultural area.

Looking at the OC standard deviation map, we can see that the variation of OC value is largest in the areas where the OC value is high, such as in Finland, Sweden, North of UK, and North of Germany … These are mostly natural areas. The area of higher variance also belongs to the common cluster of the three land use cover, cluster 21, which is peat land. In this cluster, the variation of OC value is high as we can see in the box plot of OC sampled data. The standard deviation of simulated OC value over Europe is minimum at 0.3% and maximum at 35%.

4.2 Results of quantifying N_2O EF uncertainty

As mentioned in section 2-1, there are four main EFs corresponding to four different types of nitrogen sources, including nitrogen from fertilizer, from crop residual, from soil organic nitrogen and from manure. Following the methodology in section 3-2-1, the mean, the values of 5^{th} percentile and 95^{th} percentile obtained from expert and literature study were used to estimate mean and variance parameter of their log transformed value.

Table 4-1 summarizes the result of estimating two main parameters, mean and variance for log-transformed value of the four EFs.

Table 4-1 : Distribution parameters of log transformed value of emission factors

Emission factors	Distribution parameters of log transformed value	
	Mean (μ)	Standard deviation (σ)
Emission factor of fertilizer	- 0.105	0.459
Emission factor of crop residual	- 0.588	1.084
Emission factor of manure	- 0.312	0.220
Emission factor of soil organic nitrogen	0.952	0.088

Figures of the probability distributions, which represent the uncertainty of the EFs and figures of simulated maps of EFs can be found in Appendix 2-2, 2-3. The simulated maps of EFs look like speckled noise because it was assumed that these were not spatially correlated.

Table 4-2 : Standard deviation interval of emission factor values at all examined locations over Europe

Emission factors	Standard deviation (%)	
	Minimum	Maximum
Emission factor of fertilizer	0.404	0.728
Emission factor of crop residuals	0.866	9.758
Emission factor of manure	0.147	0.190
Emission factor of soil organic nitrogen	0.205	0.254

Table 4-2 indicates the range of standard deviation of the EFs at all examined locations over Europe. As we can see, the four EFs have different level of uncertainty. Over Europe, the variation EFs at different locations also shows difference. EF of crop residuals shows highest variation of the standard deviation value over Europe. Next is the EF of fertilizer. The EF of manure and soil organic nitrogen has approximately the same lowest variation of the standard deviation value over Europe.

4.3 Results of quantifying uncertainty in empirical model used for natural areas

4.3.1 Uncertain regression coefficients

Data of 121 observations were used to fit the linear regression model describing the relationship between $LogN_2O$ and other variables (precipitation, temperature, nitrogen deposition, pH, and OC and vegetation type) as independent variables. The fitted regression model is as below:

$$\textbf{\textit{Log } }\boldsymbol{N_2O} \textbf{ =} 0.133478 \textbf{ –} 0.006914 \textbf{ * } \textbf{\textit{P}} \textbf{ +} 1.536506 \textbf{ * } \textbf{\textit{(Fraction T<0)}} \textbf{ +} 0.497303 \textbf{ *\textit{log(Ndep)}} \textbf{ –}$$

$$0.137458 \textbf{ * } \textbf{\textit{(pH)}} – 0.003439 \textbf{ * } \textbf{\textit{(Organic C)}} \textbf{ +} 0.055959 \textbf{* } \textbf{\textit{(vegetation: deciduous)}}$$

Adjusted R-squared: 0.2016

The regression coefficients are uncertain and correlated to each other. Table 4-3 indicates the estimated values of all regression coefficients and their standard deviations. Table 4-4 is the correlation matrix of these regression coefficients. The intercept and the vegetation variable have very high standard deviation in comparison with their mean.

Table 4-3: Estimated regression coefficients and parameters of their variation

Coefficients	Estimates	Std.dev from fitting	Coefficient of variation
Intercept	0.133	0.563	4.215
Precipitation	-0.007	0.003	0.429
FractionT<0	1.537	0.397	0.258
logNdep	0.497	0.290	0.584
pH	-0.137	0.062	0.448
OC	-0.003	0.001	0.292
Veg:dec	0.056	0.107	1.908

The coefficients were randomly sampled from their multivariable normal distribution for each MC run, following the method described in section 3-2-2. The histograms of randomly sampled coefficients, which illustrate their uncertainty can be found in Appendix 3.

Table 4-4: Covariance matrix of coefficients of regression model used for natural areas

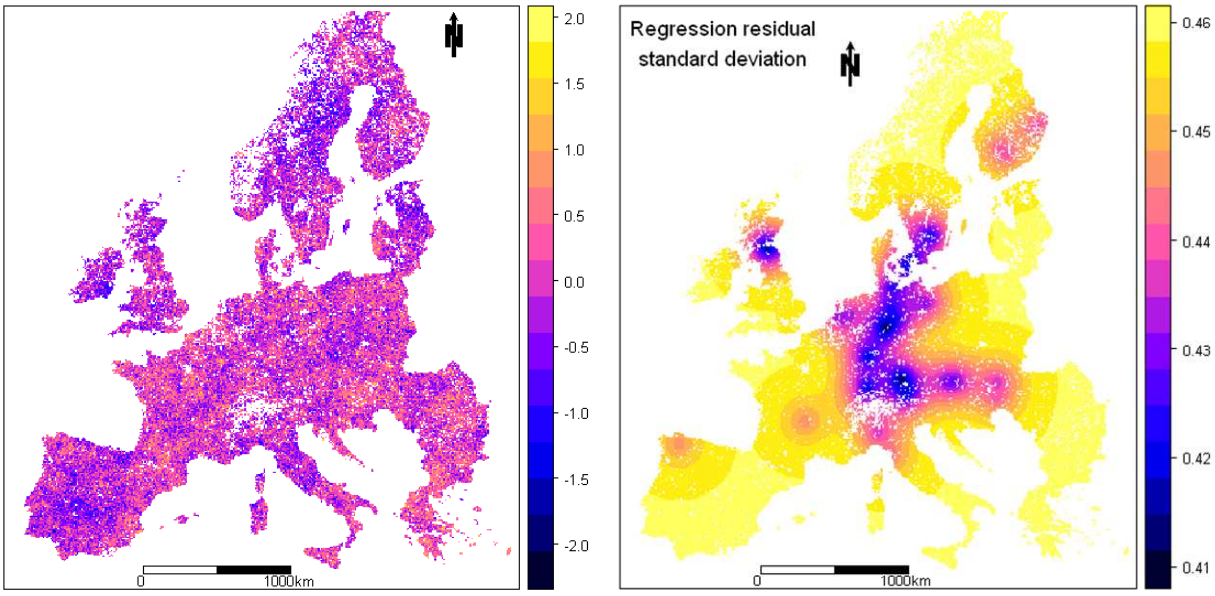
	Intercept	Prec	Fraction T<0	logNdep	pH	OC	veg:dec
Intercept	0.3165881	-0.0007828	-0.0690733	-0.1184398	-0.0127540	-0.0002627	-0.0117172
Precipitation	-0.0007828	0.0000088	-0.0001434	0.0001197	-0.0000063	0.0000007	0.0000790
Fraction T<0	-0.0690733	-0.0001434	0.1574563	0.0444586	-0.0021900	-0.0000535	-0.0072396
logNdep	-0.1184398	0.0001197	0.0444586	0.0843314	-0.0041685	0.0000984	0.0025470
pH	-0.0127540	-0.0000063	-0.0021900	-0.0041685	0.0037991	0.0000071	-0.0003727
OC	-0.0002627	0.0000007	-0.0000535	0.0000984	0.0000071	0.0000010	0.0000233
veg:dec	-0.0117172	0.0000790	-0.0072396	0.0025470	-0.0003727	0.0000233	0.0113972

To check the correctness of randomly sampled results, the variance-covariance matrix from randomly drawn values for all regression coefficients were computed and compared with variance-covariance matrix obtained from fitting the regression model. The comparison showed much similarity. The covariance matrixes are more or less equal as well. The results of these comparisons can be also found in Appendix 3.

4.3.2 Uncertain regression residual

Regression residuals which were obtained from fitting the regression model have normal distribution with mean of zero. Fitting the residual variogram shows that the nugget is 0.16; the sill is 0.21, nearly equal to the variance of the residuals (0.19) obtained from fitting the regression model; and the range is 500 kilometres. The nugget is very close to the sill.

Figures of histogram of residual value, sampled locations over Europe and its fitted variogram can be found in Appendix 4.



(a) (b)

Figure 4-3: (a) Simulated map of regression residuals (log of kg N/ha/yr) and (b) its standard deviation

One example simulated map of the regression residual is shown in Figure 4-3-a. Its value ranges from -2.0 to 2.0. The standard deviation map in Figure 4-3-b shows uncertain level of residual value over Europe. Compared to map of sampled location, we can see that the standard deviation is lower at areas where there are sampled data, and higher in other areas. The standard deviation over Europe is minimum at around 0.41 and maximum at around 0.46.

4.4 Results of quantifying uncertainty of N_2O estimate from INTEGRATOR

4.4.1 Estimating number of MC run

The first step of the MC analysis is estimating the total number of MC runs needed to reach stable results. Following the idea said in section 3-3, the results of the analysis are given in Figure 4-4. The graphs show the scatter plot of standard deviations of the first number of MC runs against those of the second number. Each point in the plots represents a grid cell on the map. It obviously shows that when the numbers of MC runs increases from 100 runs to 400 runs, the spread of the cloud of points in the scatter plot decreases. The larger the number of MC runs, the closer to the diagonal line the cloud of points comes.

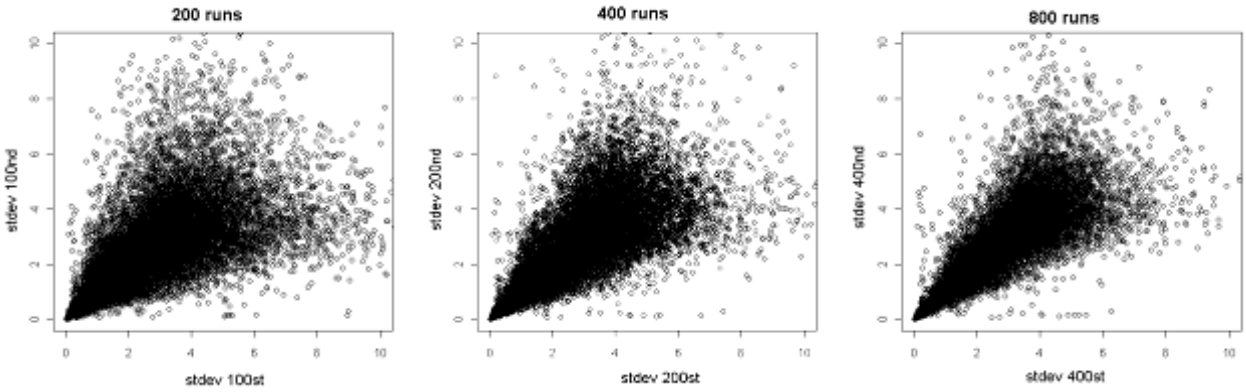
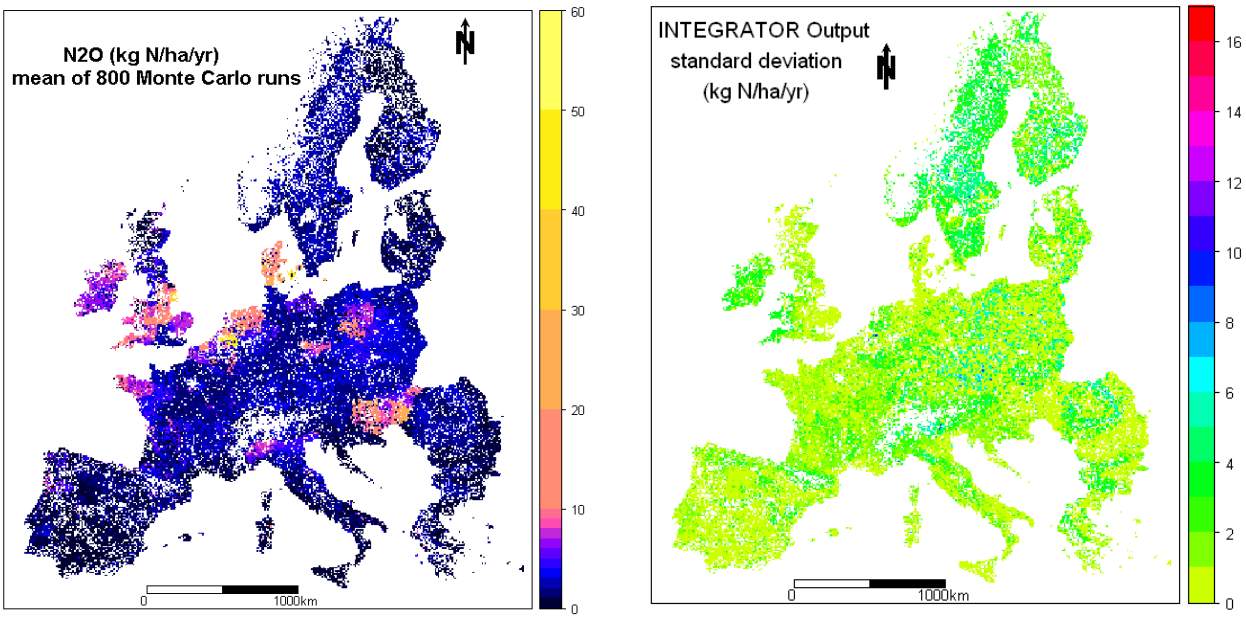


Figure 4-4: Scatter plot to examine sufficient number of Monte Carlo runs

The scatter plot for 800 runs was expected to come much closer to the diagonal line. Due to limited time of thesis work, the number of 800 iterations was decided to be the total number of MC run in this study. However, a further increase in number of MC runs is probably desirable.

4.4.2 INTEGRATOR output uncertainty

To estimate the uncertainty of INTEGRATOR outputs, MC runs of INTEGRATOR were carried out with all uncertain inputs and parameters (see section 3-3-a). A summary of all INTEGRATOR uncertain inputs, parameters and model structure used for the MC analysis can be found in Appendix 4-1.

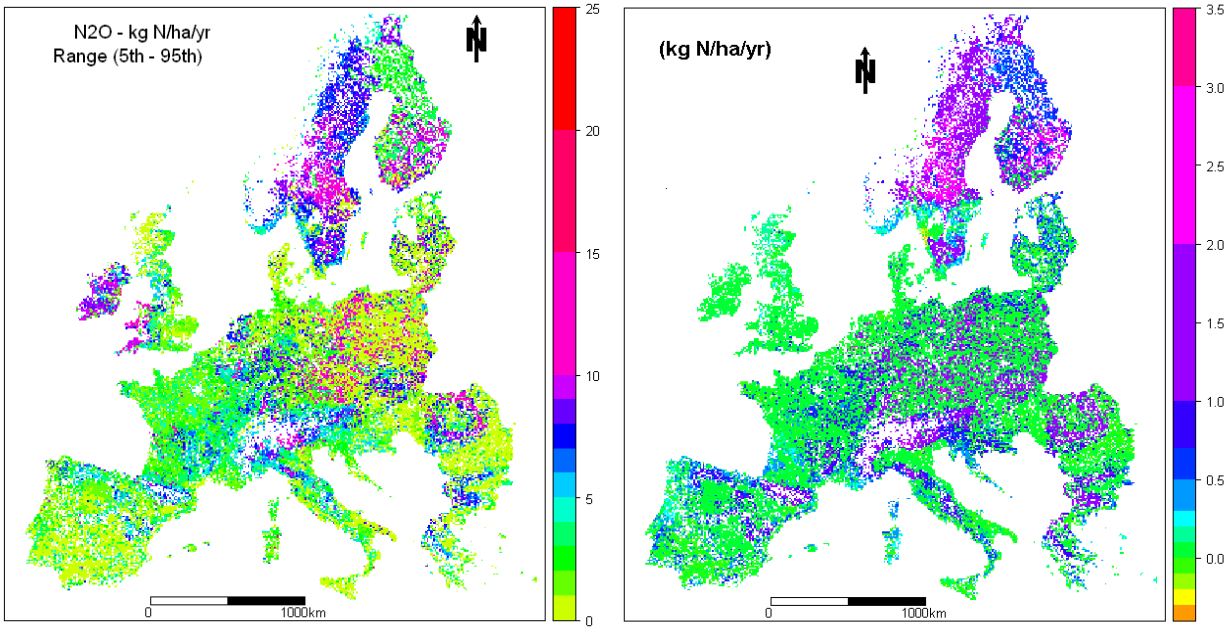


(a) (b)

Figure 4-5: (a) Estimated N_2O emission from INTEGRATOR and (b) Standard deviation of INTEGRATOR outputs

The map of INTEGRATOR output standard deviation (Figure 4-5-b) indicate the degree of dispersion of N_2O estimate outputs from its mean value (Figure 4-5-a). N_2O estimate varies in the interval of around more than 0 to around 16 kg N/ha/yr at all examined locations in Europe. In consideration of different land use cover (Figure 2-2), N_2O estimate generally has higher uncertainty in natural areas than in agricultural areas.

Another indicator of uncertainty about N_2O estimates from INTEGRATOR is the range which defined as the range of N_2O value from the value at 5^{th} percentile and at 95^{th} percentile. The range map in Figure 4-6-a is comparable to the variance map. Natural areas of Sweden, Finland, Romania, Ireland, Czech Republic and Slovakia show high range in N_2O estimate values.



(a) (b)

Figure 4-6: (a) Range of INTEGRATOR outputs
(b) Difference between output of mean uncertain inputs and mean of outputs of the same uncertain inputs

Figure 4-6-b represents the difference between output of mean uncertain inputs and mean of outputs of the same uncertain inputs. The output of mean uncertain inputs is actually the result of the so-called default run (Figure 3-4). The mean of outputs of the same uncertain inputs was obtained by calculating the mean of 800 outputs from MC. This difference is inevitable because of uncertainty propagation of model inputs, parameters and model structure to the outputs and the sensitivity of INTEGRATOR to their uncertainty. This map is also well comparable to the other two maps of variance and range. Summary of mean N_2O estimated from INTERATOR and its uncertainty interval for each European country can be completely found in Appendix 4-2.

4.4.3 Relative contribution uncertainty

Relative uncertainty contribution analysis was separately carried out for agricultural and natural areas because two different models were used to estimate N_2O for two different kinds of land use/cover, with different inputs and parameters (see section 3-3-b). Results of this analysis were also separately presented below:

- Agricultural area: uncertain input pH and uncertain parameter EFs had strongly different contribution to output uncertainty. Looking at agricultural areas in Figure 4-8-a and 4-8-b, we can clearly see the noticeably unbalanced contribution between pH and EFs. pH almost had no contribution to the output uncertainty; while more than 99% of the output uncertainty on average was contributed by EFs. Figure 4-7 indicates average percentage of contribution of pH and EFs from all examined locations in agricultural area.

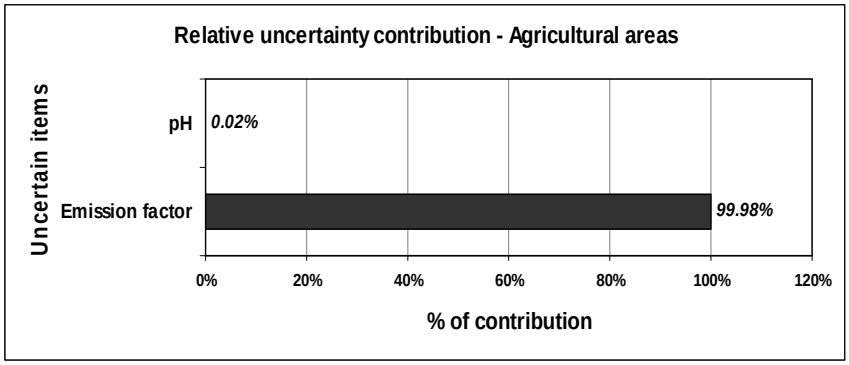
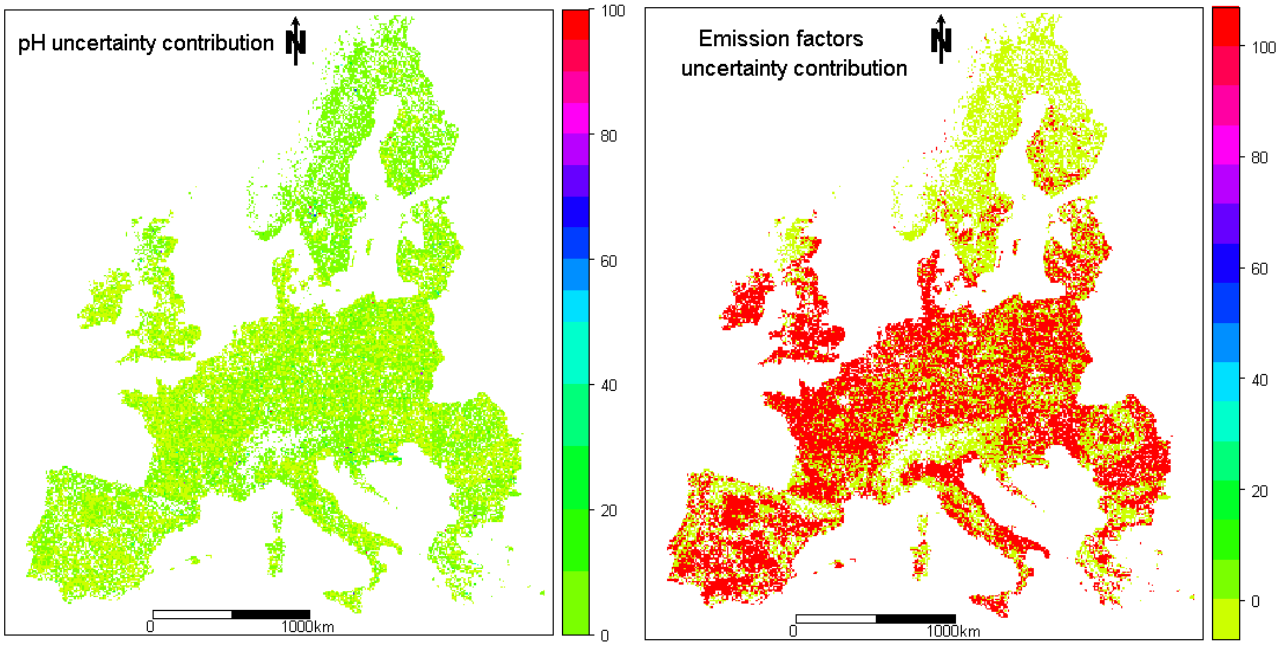


Figure 4-7: Average relative uncertainty contribution in agricultural area



(a) (b)

Figure 4-8: Relative uncertainty contribution (%) in agricultural area: (a) pH and (b) Emission factors

- Natural area: although the contributions of uncertain pH, OC and uncertain regression model were less remarkably unbalanced than the contribution of pH and EFs in agricultural areas, difference in the extent of their contributions can also be easily seen. Comparing the difference in natural area from three Figures 4-8-a, 4-9-a and 4-9-b, we can see that the regression model had the largest contribution amongst the three. Average contribution of the regression model was more than 76%. pH had the smallest contribution. OC contributed 21% to the uncertainty (Figure 4-10).

Note that uncertainty in pH contributed to output uncertainty for both land uses. The extent of contribution is different; its contribution was slightly higher in natural areas than in agricultural areas.

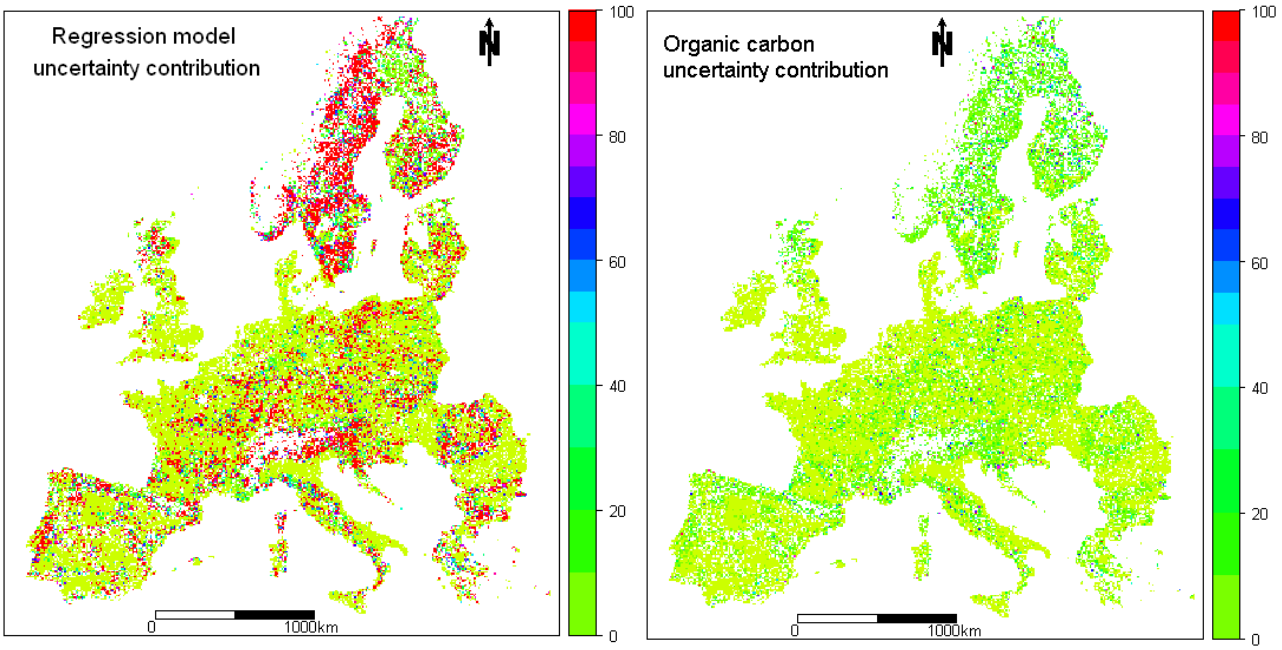


Figure 4-9: Relative uncertainty contribution (%) in natural area: (a) Regression model and (b) Organic carbon

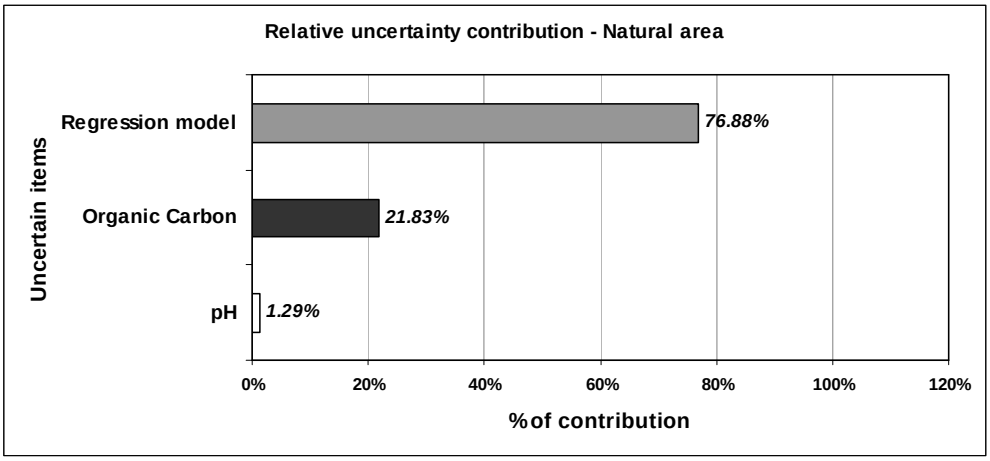


Figure 4-10: Average relative uncertainty contribution in natural area

Average uncertainty contribution of these inputs, parameters and model structure in each country were also analyzed. In the Netherlands, given as example, the relative uncertainty contributions are 16.41% for OC, 0.33% for pH, 75.57% for EFs and 7.69% for the regression model uncertainty. In Finland, relative uncertainty contribution is 39.26% for OC, 0.65% for pH, 6.80%, for EFs and 53.30% for the regression model uncertainty. The two examples show differences in average relative uncertainty contribution of uncertain items in different countries. This issue will also be discussed in section 5.3. Average relative uncertainty contribution at point scale of every European country is fully represented in Appendix 4-3.

4.5 Results of analyzing spatial aggregation effect on output uncertainty

The outputs from INTEGRATOR were at point scale of 10x10km spatial interval. Variance of N_2O estimate at points within every European country is illustrated in Figure 4-11 for each country. It shows a large variation in variance value at point scale (Varpts) within every country. These outputs at point scale were then aggregated to national scale and the variance of N_2O estimate at national scale (Varctr) was computed. The variance of N_2O estimate at national scale was also plot in the same box plot of N_2O estimate variance at point scale (red dots in Figure 4-11). The graph shows that the variance at national scale is very much lower and out of the interquartile range of the variance at point scale.

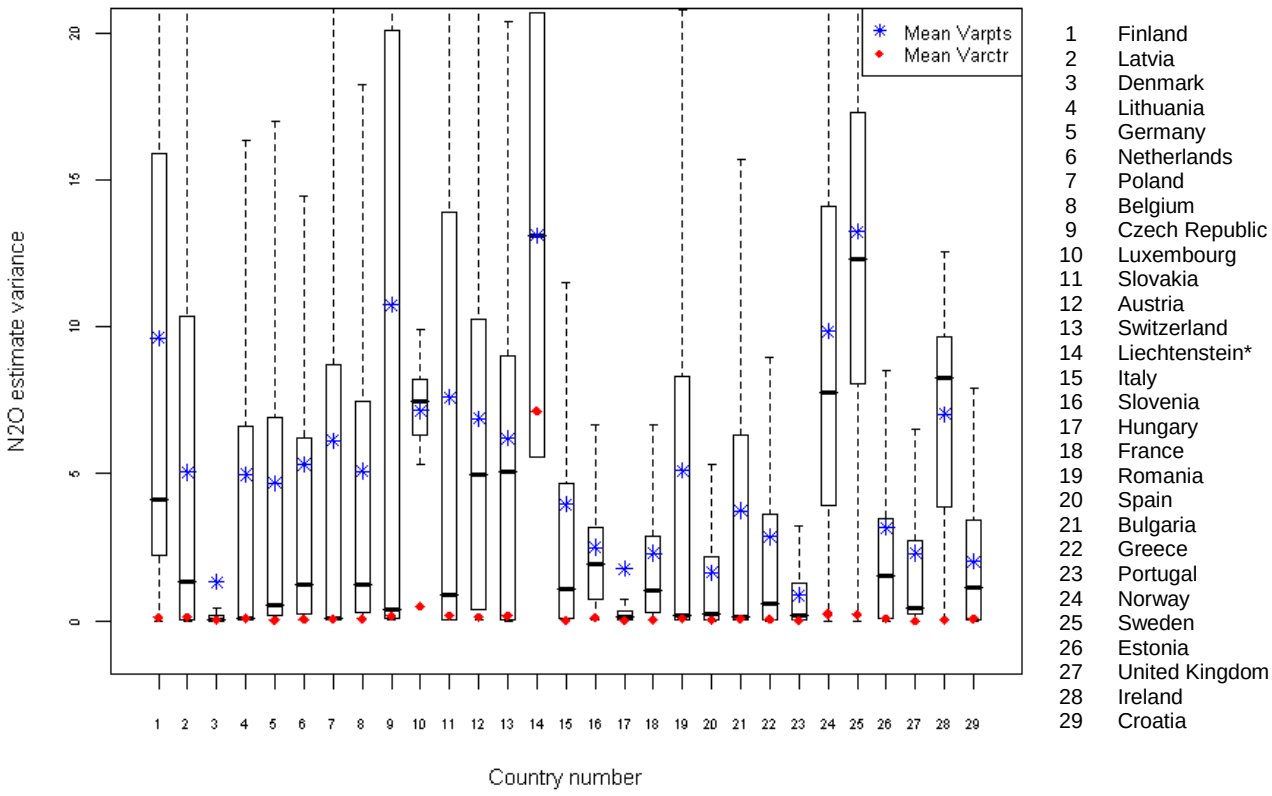


Figure 4-11: Comparison between output variance at point scale and at national scale

To quantify the decrease in variance after aggregating from point scale to national scale, the average of the point variances within each country was calculated and compared with the variance at national scale. The variance of N_2O reduced when up-scaling to national scale, European countries. Average percentage of reduction is more than 95%. The reduction is large when country area is large. For example, France with areas of 550.000 km^2, percentage of variance reduction is 99.73%. Liechtenstein* with areas of less than 200 km^2, the variance was only reduced by 45.63%. The result of variance reduction computation for all European countries can be completely found in Table 4-5.

Table 4-5: Percentage of uncertainty reduction for all European countries

Country	Areas (10000 km^2)	Variance at point scale (Mean Varpts)	Variance at national scale (Varctr)	% uncertainty reduction
Finland	33.31	9.627	0.117	98.78%
Latvia	6.43	5.047	0.110	97.83%
Denmark	4.25	1.311	0.011	99.17%
Lithuania	6.47	4.966	0.079	98.41%
Germany	35.59	4.687	0.013	99.72%
Netherlands	3.54	5.313	0.041	99.23%
Poland	31.05	6.131	0.041	99.33%
Belgium	3.05	5.090	0.055	98.93%
Czech Republic	7.85	10.750	0.160	98.51%
Luxembourg	0.26	7.142	0.497	93.03%
Slovakia	4.88	7.598	0.156	97.95%
Austria	8.37	6.883	0.112	98.38%
Switzerland	4.14	6.201	0.157	97.47%
Liechtenstein*	0.02	13.113	7.130	45.63%
Italy	29.96	3.956	0.017	99.58%
Slovenia	2.04	2.497	0.090	96.39%
Hungary	9.27	1.759	0.012	99.33%
France	54.66	2.295	0.006	99.73%
Romania	23.68	5.108	0.053	98.97%
Spain	50.57	1.642	0.007	99.60%
Bulgaria	11.08	3.742	0.063	98.32%
Greece	13.00	2.848	0.040	98.61%

Country	Areas (10000 km^2)	Variance at point scale (Mean Varpts)	Variance at national scale (Varctr)	% uncertainty reduction
Portugal	9.18	0.863	0.012	98.64%
Norway	31.89	9.852	0.228	97.69%
Sweden	44.33	13.239	0.205	98.45%
Estonia	4.57	3.161	0.072	97.72%
United Kingdom	24.33	2.294	0.002	99.92%
Ireland	6.94	7.028	0.015	99.78%
Croatia	5.58	2.016	0.060	97.04%

5 Discussion

5.1 Statistical models used for quantifying model input uncertainty: pH and OC

The results of using geostatistical modelling to predict pH and OC were presented in Section 4-1 showed the pattern of spatial variation of pH and OC value over Europe. The question raised here is whether the statistical models used are reasonable. To answer this question, we considered the following. The results of analyzing sampled data of pH and OC proved that the difference in pH and OC mean and their variation amongst cluster existed. It means that when estimating pH value at specific location over Europe, the cluster it belongs to should be taken into account. The mean value of pH and OC over all examined locations in Europe is not constant but it varies according to which clusters it is. Obviously, pH and OC have non-stationarity in their mean over space. Thus, an RK approach was used. The computational approach of RK is that the predictions are made separately for the mean and residuals and then added back together (Hengl et al. 2003). This is the reason of establishing the model as in equation 3-1. The non-constant mean is expressed as a function of clusters. The residuals expressed as the variation of the constant standard deviation of cluster at each location belonging to this cluster. The variograms were then fitted for the variation of the residuals.

The spherical model with nugget was chosen to fit the experimental variogram of the residual variation for both pH and OC value. This simple model is expected to be good enough to model the spatial correlation of pH and OC value over locations of 10x10 km spatial interval. However, fitted experimental variogram of pH sample data in arable and grassland can be improved to get better result.

Compared to other results of mapping pH and OC value and with soil texture map of Europe, we can say that the statistical models used for pH and OC are very reasonable. For example, a map of pH estimated from top soil in Southern UK created by the National Soil Resources Institute, Cranfield University (2002) shows a pattern of pH spatial variation that is very similar to the Kriging mean obtained by using our model in the same area (Figure 5-1). In the case of OC, for instance in Spain, the map of soil texture (Appendix 5) indicates mostly clay; the simulated map shows low OC values. Contrastingly, in the predominantly peat soils in Finland, high OC values are simulated. The pattern of OC spatial variation in our simulated map is also very comparable to the OC map presented in the map of organic carbon in top soils in Europe (Version 1.2 – 2003) (Appendix 5).

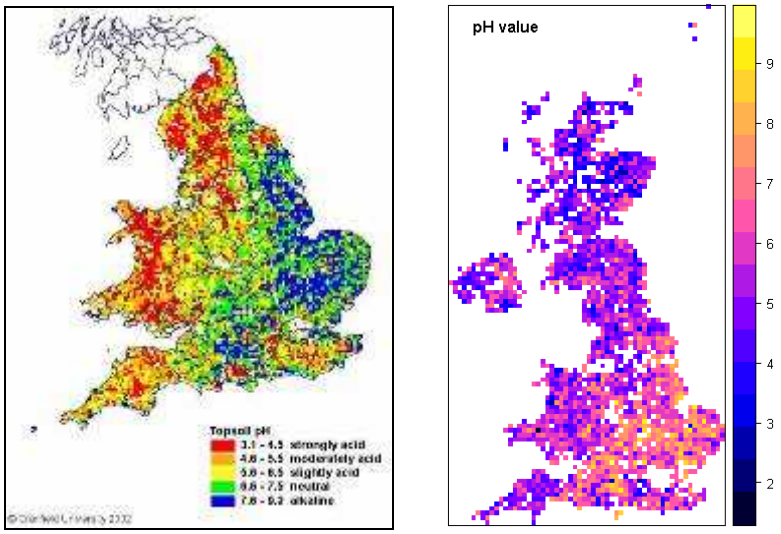


Figure 5-1: pH map of UK created by Cranfield University (left)and our single simulated map (right)

The uncertainty about pH and OC value at each location depends on which cluster it belongs to. Distribution of pH and OC in each cluster depends on sampled data available in this cluster. Because the singled simulated map of pH and OC is a combination of simulation maps of three land uses, the variation of pH and OC value at each location also depends on which land use this location belongs to. For OC, the three land uses have the same common clusters 14 and 21. Cluster 21 was characterized with high OC value and high variation. Accordingly the single simulated map of OC shows high value and also high uncertainty about OC at all allocations belong to cluster 21. Meanwhile, cluster 14 had lower OC value and lower variation than cluster 21; consequently, single simulated map of OC has lower value and lower uncertainty of OC in cluster 14. Maps of pH and OC clusters can be found in Appendix 1-1.

Results of quantifying uncertainty of pH and OC at all locations show high variation in OC value than in pH value. OC data had a very skewed distribution; therefore, a logarithmic transformation was first carried out for OC. The back-transformed value of OC at some locations was extremely high. These extreme high values resulted from using the antilog. Values greater than 100% were replaced by 100%, the maximum value OC can have. Perhaps, the variance of OC simulated value presented in section 4-1-2 computed by the variance of 800 simulated map may be better expressed by the variance computed from another complex approach such as the one presented in (Webster and Oliver 2007) or the one suggested by (Yamamoto 2007).

5.2 Statistical models used for quantifying model parameters and model structure uncertainty

a) N_2O emission factor uncertainty

According to (IPCC 2000), uncertainty of EFs are relatively small when the standard deviation divided by the mean is less than 0.3. Following this definition, the EF of crop residuals has an extremely high level of uncertainty amongst the four EFs, with the ratio ranges from 0.8 to 9.7 at all examined locations. Uncertainty of fertilizer EF is also high; the ratio ranges from 0.4 to 0.7. EF of manure and soil organic nitrogen are low because the ratios are all smaller than 0.3 (maximum 0.25 and 0.1, respectively).

In the case of quantifying EF uncertainty, there was no experimental data available for simulating EFs over space. Information about uncertainty of EF was completely obtained from expert and literature study. The autocorrelation in space could not be characterized as well and zero spatial correlation was assumed. The simulated maps of EFs consequently look like speckled noise. Thus, uncertainty of each EF was assumed to have similar distribution over Europe for each kind fertilizer application. However, the value of the EFs and their uncertainty proved not to be the same over space. N_2O EFs vary depending on environmental effects (e.g. soil type, weather), land use and management effects (e.g. type of manure and fertilizer, application rates, time of application) (IPCC 2000; Kuikman et al. 2006; Velthof 2008).

Is the result of quantify EFs' uncertainty using statistical model in our study rational to use for assessing uncertainty propagation of EFs to INTEGRATOR output? As said in section 2-1 these four EFs are the reference EFs. It means that the true EFs at each location in each European country will be a fraction of one of these four references, which already take all effects mentioned above into account. This is the way how INTEGRATOR estimates N_2O soil emission from agricultural area. Accordingly, uncertainty of the four reference EFs is first propagated to EFs' value at specific location and from this the uncertainty is then propagated to the outputs of INTEGRATOR. Therefore, the uncertainty quantifying under our statistical models is rather sufficient although their autocorrelation were not taken into account because of no data available. Many researches studied how to estimate average annual EFs for country and quantifying the uncertainty of EFs in the values themselves and over time but not focus on quantifying their uncertainty over space. We expected that there will be large difference in EFs' uncertainty propagation between the inputs uncertainty varied in space and constant in space; and when autocorrelation can be described in the case of EFs varied in space. The extent of propagation of each EF's uncertainty is also expected to be different.

b) Uncertain regression model using for natural area

The covariance matrix represented in section 4-3-1 expressed the uncertainty of regression coefficients and their dependency. The results of checking (Appendix 3-1) indicate that the randomly drawn results are reasonable. It means that the statistical model worked well. Their uncertainty is affected by the size of the sampled data used to fit the regression model. The small size of sampled data used for fitting caused the probability distribution of coefficients had high variance (Montgomery et al. 2001; Brus and Jansen 2004).

The regression residual is independent from uncertainty in the regression coefficients. Simulating the residual value over space was done while taking spatial correlation into account. Because value of adjusted R^2 is quite low (0.2016), the variance estimate of fitted regression model is high. Moreover, number of observations is small and clustering into some areas. This led to a large Kriging variance of the residual. Both caused high uncertainty about model structure at all examined locations.

5.3 INTEGRATOR output uncertainty and possibility of uncertainty reduction

The result of quantifying total output uncertainty shows that uncertainty about N_2O emission in natural areas is higher than in agricultural area. Figure 5-2 shows that uncertainty about N_2O emission at most of locations in natural land are high, with the coefficient of variance larger than 50% at almost all examined locations. Higher uncertainty in natural areas can be explained by higher sensitivity of the model and high uncertainty in the inputs, parameters and model structure. It is difficult to see the correspondence between high uncertainty in the inputs, parameters and model structure and high uncertainty in the outputs from the total uncertainty map. But it can be seen clearer in the contribution analysis results. In the model used for natural areas, pH and OC seem to be negatively correlated but their uncertainties were quantified without taking their correlation into account. Uncertainty of pH input might increase when taking its correlation with OC in account. Uncertainty in N_2O emissions from agricultural areas were caused by uncertainty propagation of pH and EFs; while uncertainty of N_2O estimate from natural areas were caused by uncertainty propagation of pH, OC and the regression model. The EFs are the most important source of uncertainty in N_2O emission from agricultural land. In natural area, uncertainty from regression model which describing the relationship between N_2O estimate and other factors had largest uncertainty contribution. In the model used for agricultural land, pH input only played a role of environmental factor effect on EFs. Thus, the contribution of pH in agricultural land is very low. The contribution of pH in natural areas is higher because there pH played a role as one of the inputs.

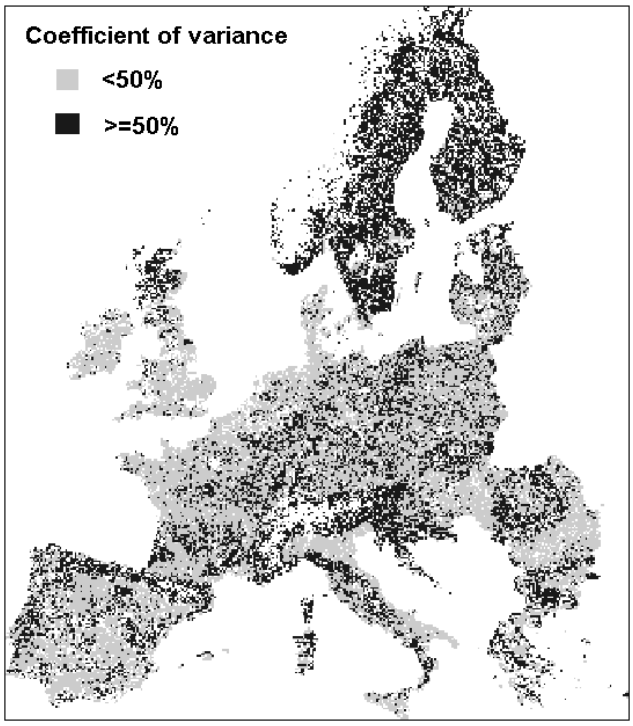


Figure 5-2: Coefficient of variance of N_2O estimate from INTEGRATOR model

The average uncertainty contribution of uncertain items at each country depends on the size of the natural and agricultural area in this country. This is the reason why the uncertainty contribution differed amongst countries (Appendix 4-3). For example, when comparing the Netherlands, Finland and Greece, Netherlands has a relatively large area of agricultural land (Figure 5-3). On the contrary, Finland has small area of agricultural land but a much larger area of natural land. Consequently, in the Netherlands, the average uncertainty contribution of EFs is highest. The highest uncertainty contribution in Finland is from the regression model which mainly comes from the natural area. In Greece, the contribution of the EFs (44.97%) and regression model (43.19%) is almost equal, which is owing to the balance between agricultural and natural areas. This is interesting to consider when the model is implemented for individual countries and uncertainty reduction has to be achieved. Average contribution at all locations over Europe is well balanced, OC: 12.30%, pH: 0.69%; EF: 46.37% and regression model: 40.64 % due to the approximately balance between agricultural and natural areas over Europe as a whole. Note that the sum of percentage of contribution was not absolutely but only approximately equal to 100%. All results were expressed in percentage of percentage. Due to a lot of missing points in Norway, the result for this country are not accurate.

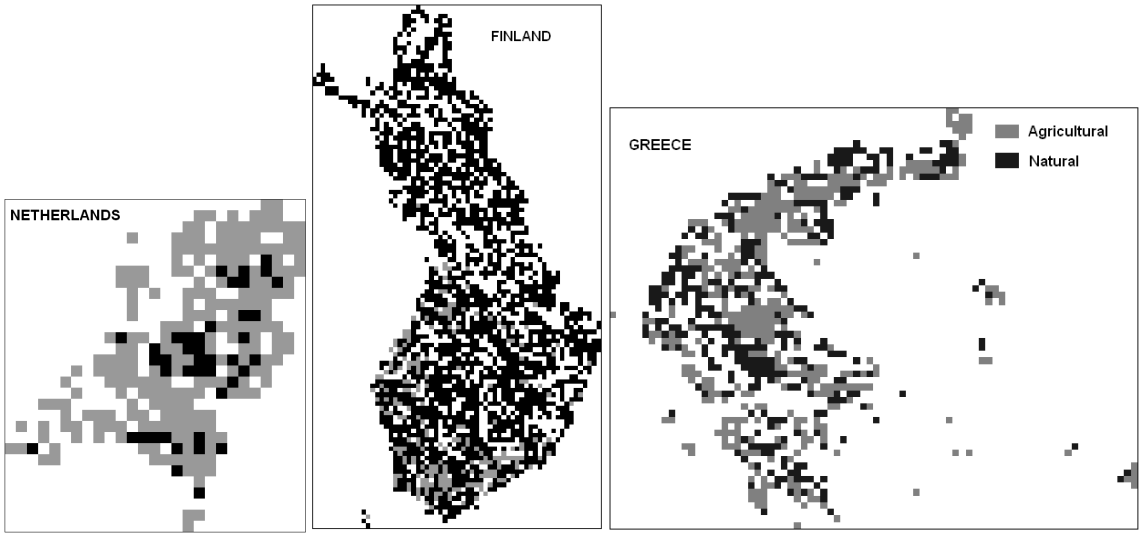


Figure 5-3: Land use/cover maps of Netherlands (left), Finland (middle) and Greece (right)

Uncertainty contributions of EFs and regression model are so-called grouped uncertainty contributions. EF uncertainty contribution consists of uncertainty contribution of four EFs. Differences in the extent of contribution of individual EFs can be expected. Similarly, uncertainty of regression model comprises of a combined effect of several regression coefficients (uncertainty in model parameters) and regression residual (uncertainty in model structure). The two contributions were proved to be different as well; and model structure has higher uncertainty contribution. Separately analyzing their individual uncertainty contribution will give better understanding.

Approach to reduce uncertainty about N_2O estimate from agricultural and natural area is by decreasing uncertainty of the most important sources in each area, EFs and regression model respectively. However, uncertainty of output in natural area is much higher. Average uncertainty over Europe can be reduced significantly by first reducing the uncertainty in natural areas.

5.4 Assessment of spatial aggregation effect on uncertainty of model output

The result of analyzing spatial aggregation effect in section 4-5 evidently shows that aggregation of outputs from point to national scale decreased the uncertainty in the outputs. This is because of average effect which cause the uncertainty in the outputs to level off. Thus, Varpts is the variance at point scale and can not stand for uncertainty about the outputs at country scale. If we talk about uncertainty at national scale, the Varctr value should be used. Variance reduced due to spatial aggregation does not mean that the result is better or worse. The validity of the model output should be evaluated at the scale for which it was built (Heuvelink 1998). When we talk about uncertainty of model outputs, we should mention the scale for which the model was implemented.

Question is whether spatial aggregation affects on uncertainty contributions as well. Although uncertainty interval of INTEGRATOR outputs changed after aggregated to different scale, the uncertainty contribution of each uncertain item is not seriously affected. We made this conclusion based on analyzing uncertainty contribution of all uncertain items in agricultural area and natural area at two scales: point scale and region scale. The results are presented in table 5-1.

Table 5-1: Analyzing uncertainty contribution at different spatial aggregation scale

Uncertainty contribution in agricultural area (%)			**Uncertainty contribution in natural area (%)**		
	Emission factor	**pH**	**pH**	**Organic Carbon**	**Regression model**
Point scale	99.98	0.02	1.29	21.83	76.88
Region scale	99.80	0.20	0.08	39.17	60.75

Uncertainty contribution at point scale was calculated as average of uncertainty contribution at all examined points. At scale of area, we followed the same steps of calculating Varctr (Figure 3-6) to calculate the variance of N_2O estimate from natural and agricultural areas; and then computed the percentage of uncertainty contribution. Even though averaging effect reduce the uncertainty of EFs and increased ten times

the uncertainty in pH; EFs still is the most important source of uncertainty. And in natural area, uncertainty in regression model is still the most important source.

6 Conclusions

In conclusion, the objective of this study is completely fulfilled. Uncertainty of INTEGRATOR inputs (soil pH and soil organic carbon), model parameters (emission factors and coefficients in relationship describing N_2O exchange rate from soil) and model structure uncertainty (random error in empirical model used for natural area) were quantified. Their uncertainty propagation through INTEGRATOR model was assessed. Analysing change in output uncertainty due to aggregating to different scale was carried out.

Estimate of N_2O emission from soil in Europe from INTEGRATOR shows different levels of uncertainty at different examined locations and between natural area and agricultural area due to error propagation from different inputs, parameters and model structure. The standard deviation of N_2O estimate varies from very low at 0.007 kg N/ha/yr to around 16.58 kg N/ha/yr over all examined locations at 10x10km interval cover whole Europe. There is more uncertainty in estimated emission of N_2O from natural area than from agricultural area. In natural areas, almost all examined locations had coefficient of variation larger than 50%; while in agricultural area, the coefficient of variation is mostly smaller than 50%.

Conclusion for each research question:

1. How to quantify uncertainty in INTEGRATOR inputs, parameters and model structure taking spatial correlation and cross-correlation into account?

   Uncertainty in INTEGRATOR inputs, parameters and model structure can be quantified by first defining their probability distribution based on sampled data, literature study or expert judgments. For continuous inputs soil pH and OC, emission factors and random error in empirical model, stochastic simulation can be used to draw random value from their probability distribution. Uncertainty about their values is obtained from Kriging prediction variance which already takes their spatial correlation into account, except emission factor uncertainty. For regression coefficients, their value can be randomly sampled from their joint probability distribution. Their uncertainty can be obtained from their covariance matrix taking their cross-correlation into account.

   In general, uncertainty in pH value is higher in agricultural area than in natural area; its standard deviation ranges from 0.3 to 1.3 at all examined locations. Contrary to pH, organic carbon shows higher uncertainty in natural areas, its standard deviation ranges from 0.3% to 35% over all examined locations. Regression coefficients have a different level of uncertainty according to each of the regressors. Random error in the empirical model (regression residual) had standard deviation around 0.45. Uncertainty in emission factors also shows different amongst them; highest uncertainty is about emission factor of crop residual and lowest uncertainty is in the emission factor from manure.

2. To what extent propagate uncertainty of model inputs, model parameters and model structure to model outputs and differ in the magnitude of uncertainty propagation?

   Monte Carlo approach was used to analyzing uncertainty propagation to INTEGRATOR output. Total uncertainty about N_2O estimates reflects the extent of uncertainty propagation. The magnitude of uncertainty propagation varies with specific model used for natural and agricultural area and with each uncertain item. Empirical model used for natural area is much sensitive to uncertain items. The most important source of uncertainty in estimating N_2O from natural area is uncertainty in the empirical model itself, including the uncertain coefficients and the residual. In agricultural areas, uncertainty in emission factors is the most important source. It means that the emission factors and the empirical model are the two most important items of which any error in these two items can cause big error in the predicted result of INTEGRATOR.

3. How spatial aggregation level affect output's uncertainty employed in the analysis?

   Obviously, spatial aggregation from point scale to national scale reduced a lot of output uncertainty due to the averaging out effect. Percentage of uncertainty reduction is proportional to extent of area aggregated to.

4. Uncertainty interval of which inputs, model parameters or model structure should be reduced to substantially reduce uncertainty of the predicted N_2O emission from agricultural and natural areas over Europe?

   To reduce error in the outputs, the error of the most contributory item should be first decreased. In natural areas, decrease in empirical model uncertainty will achieve large error reduction in predicted N_2O emission. In agricultural areas, emission factor uncertainty should be first reduced to gain large reduction in the outputs. However, because uncertainty in the output of natural areas are much higher than that in

agricultural area, first reduction in uncertainty of outputs from natural area will largely decrease the uncertainty in the outputs over Europe.

We recommend that further uncertainty contribution analysis should be carried out separately between regression coefficients and regression residual because their contribution to the output error have proved to be different and the methods to reduce their uncertainty are also different. The uncertainty in regression coefficients depends on the sample size used to fit the model; while the uncertainty of regression residual might mainly depend on chosen structure of the model and independent variables used in regression model. Separate uncertainty contribution analysis should be also carried out for each emission factor, taking spatial correlation and/or cross-correlation into account. The model used for agricultural area has two main source of uncertainty from emission factors and amount of nitrogen input into soil. This study only assessed uncertainty contribution of emission factors and the given result is of the highest uncertainty contribution. But uncertainty in amount of nitrogen input into soil can be even larger. Further research should also be carried out for this second main source of uncertainty. The empirical model used for estimating N_2O from natural areas should be revised. Numbers of MC runs can be desirably increased.

This study looked at uncertainty propagation from model inputs, model parameters and model structure to model output in the case of using INTEGRATOR to predict N_2O emission from terrestrial area in Europe. It showed that uncertainties about N2O emissions can be quite large, especially in the predicted results from natural land. It is therefore important that uncertainties are quantified. Important decisions can be made based on the presented results. Therefore, the results must therefore be of sufficient accuracy. This study has shown that uncertainty propagation analysis can be an important tool to help quantify the uncertainty about emission maps. Finally, from this study it is recommended that uncertainty should always be mentioned together with the scale at which the model is implemented.

7 Bibliography

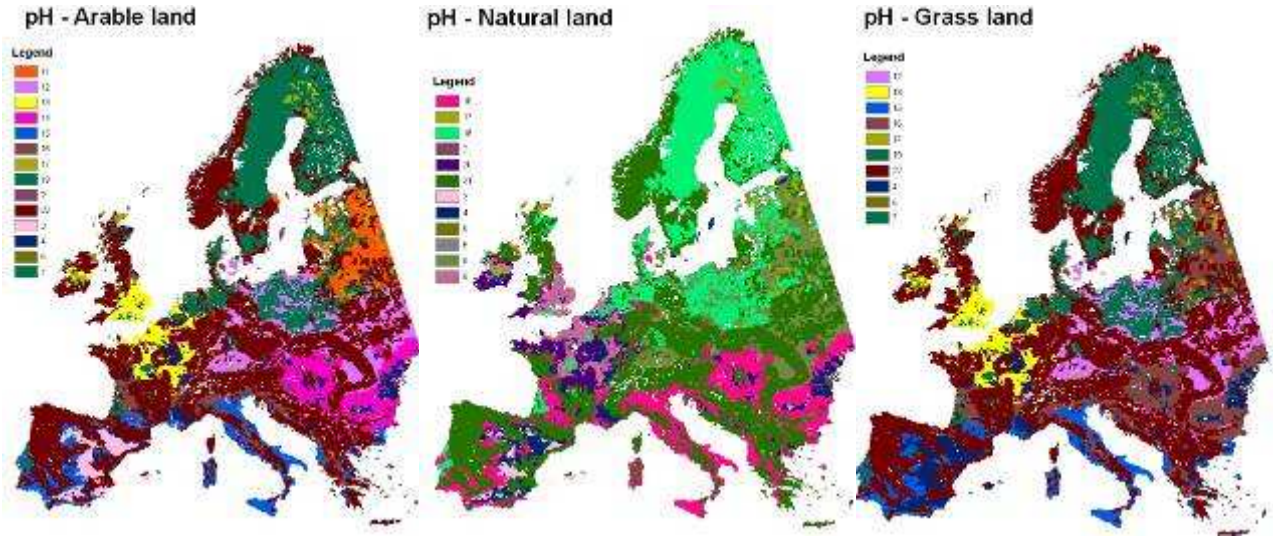
8 Appendices

8.1 Appendix 1: Data and extra results related to pH and OC

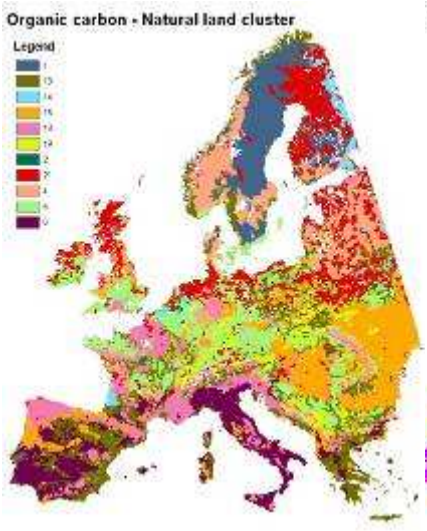
1.1 Soil cluster

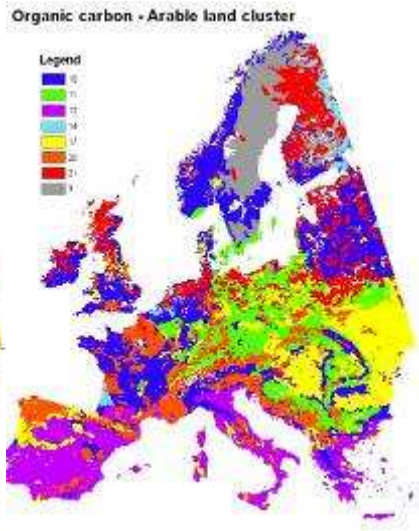
pH cluster			
Cluster	**pH**	**Texture**	**wetness**
1	calcareous	loam,clay	dry,wet
2	calcareous	clay, loam, unknown	dry,wet
3	calcareous	clay,loam,unknown	wet,dry
4	calcareous	clay,loam,unknown	dry,wet
5	eutric	peat	wet
6	eutric	sand	dry
7	eutric	sand	dry
8	eutric	clay,loam,volcanic,unknown	wet,dry
9	eutric	clay,loam,volcanic,unknown	dry,wet
10	eutric	clay,loam,volcanic,unknown	wet,dry
11	eutric	clay,loam,volcanic,unknown	wet,dry
12	eutric	clay,loam,volcanic,unknown	dry,wet
13	eutric	clay,loam,volcanic,unknown	dry,wet
14	eutric	clay,loam,volcanic,unknown	dry,wet
15	eutric	clay,loam,volcanic,unknown	dry,wet
16	eutric	clay,loam,volcanic,unknown	wet,dry
17	dystric	peat	wet
18	dystric	sand	dry,wet
19	dystric	sand	dry,wet
20	dystric	loam,volcanic,unknown	wet,dry

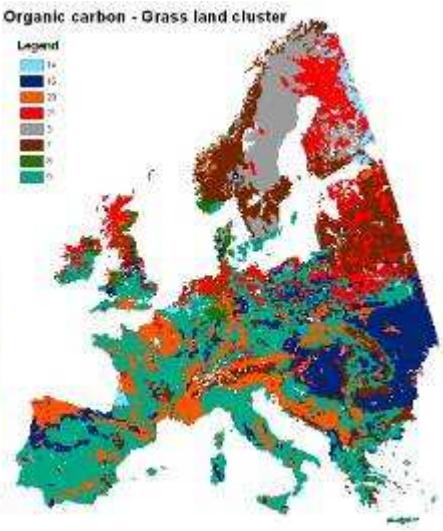
pH cluster			
Cluster	**pH**	**Texture**	**wetness**
21	dystric	loam,volcanic,unknown	dry,wet
22	dystric	loam,volcanic,unknown	wet,dry



OC cluster				
Cluster	**Organic matter**	**pH**	**Texture**	**wetness**
1	low	dystric,eutric	sand	dry,wet
2	low	dystric,eutric	sand	dry,wet
3	low	dystric,eutric	sand	dry,wet
4	low	dystric,eutric,calcareous	clay,loam,volcanic	wet,dry
5	low	dystric,calcareous,eutric	clay,loam,volcanic	dry,wet
6	low	eutric,calcareous,dystric	clay,loam,volcanic	dry,wet
7	low	eutric,dystric,calcareous	clay,loam,volcanic,unknown	wet,dry
8	low	eutric,dystric,calcareous	clay,loam,volcanic,unknown	dry,wet
9	low	eutric,dystric,calcareous	clay,loam,volcanic	dry,wet
10	low	eutric,dystric,calcareous	clay,loam,volcanic,unknown	dry,wet
11	low	eutric,dystric,calcareous	clay,loam,volcanic,unknown	dry,wet
12	low	eutric,dystric,calcareous	clay,loam,volcanic,unknown	dry,wet
13	low	eutric,dystric,calcareous	unknown	wet,dry
14	high	dystric	sand	dry
15	high	eutric,dystric,calcareous	clay,loam,volcanic	dry,wet
16	high	dystric,eutric,calcareous	clay,loam,volcanic	wet,dry
17	high	eutric	clay,loam,volcanic	dry,wet
18	high	dystric,calcareous	unknown	dry
19	high	dystric,calcareous	unknown	dry
20	high	dystric,calcareous	unknown	dry
21	high	eutric,dystric	peat	wet

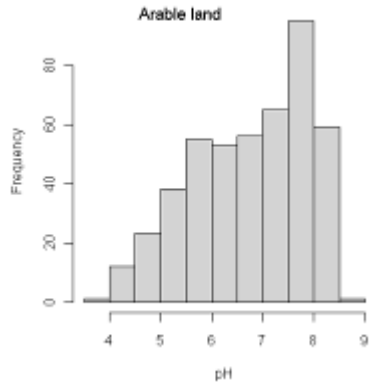


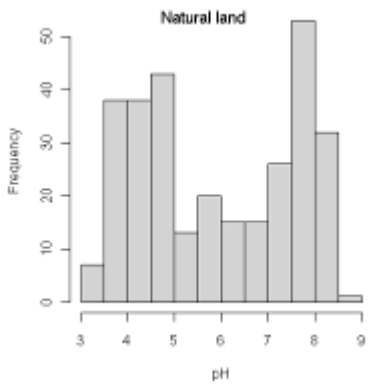


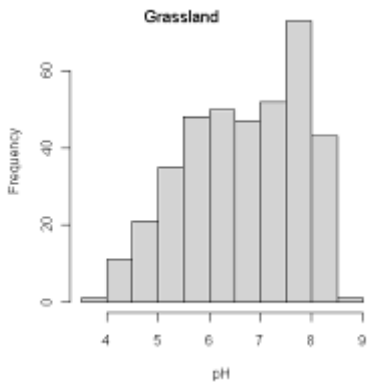


1.2 pH

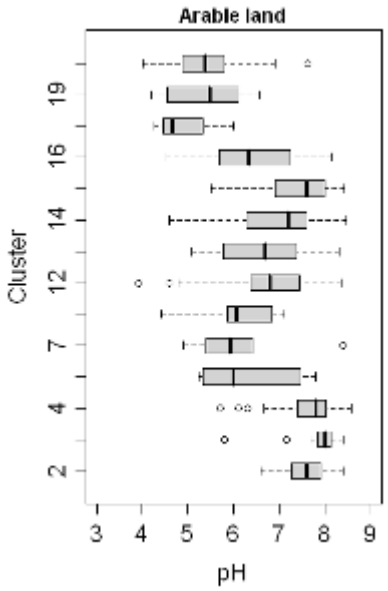
Histogram of pH

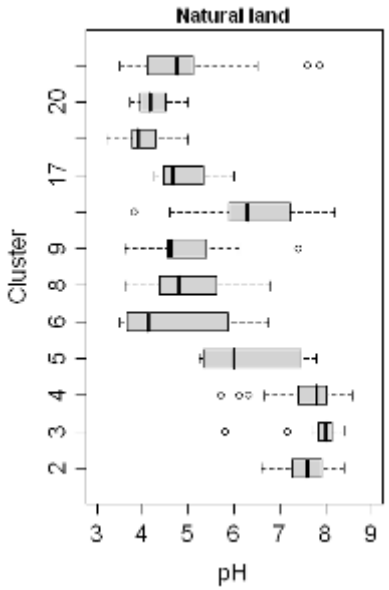


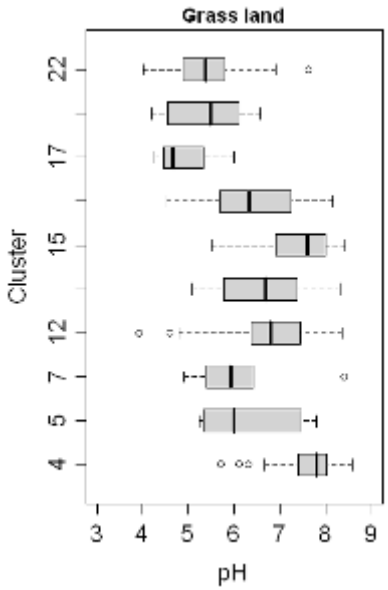




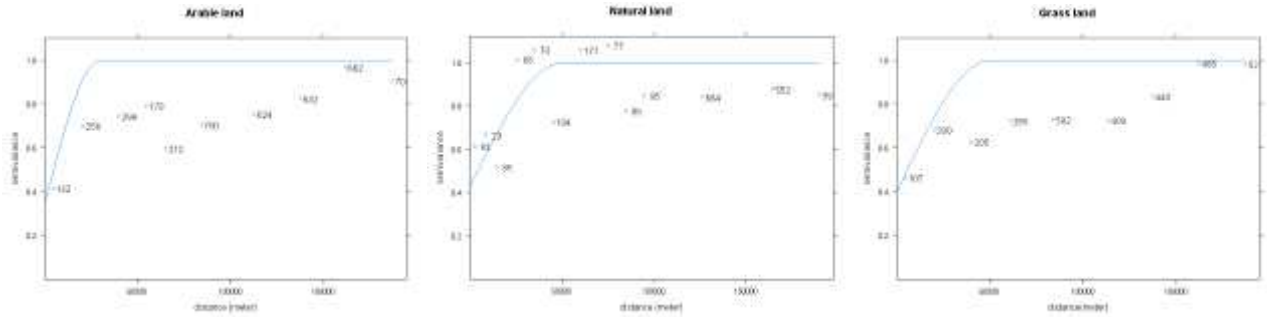
Box plot of pH





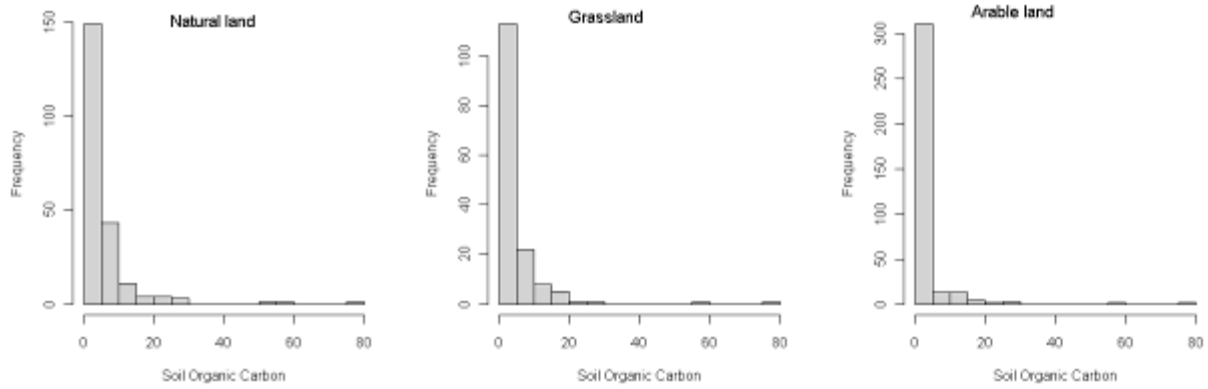


Fitted variogram of pH

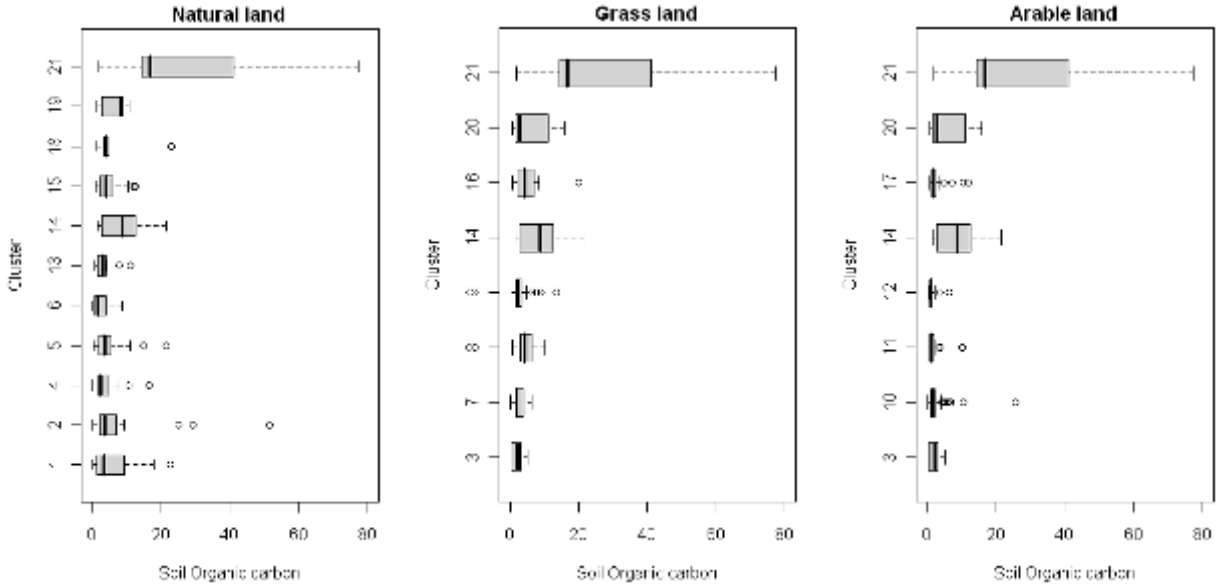


1.3 Organic carbon

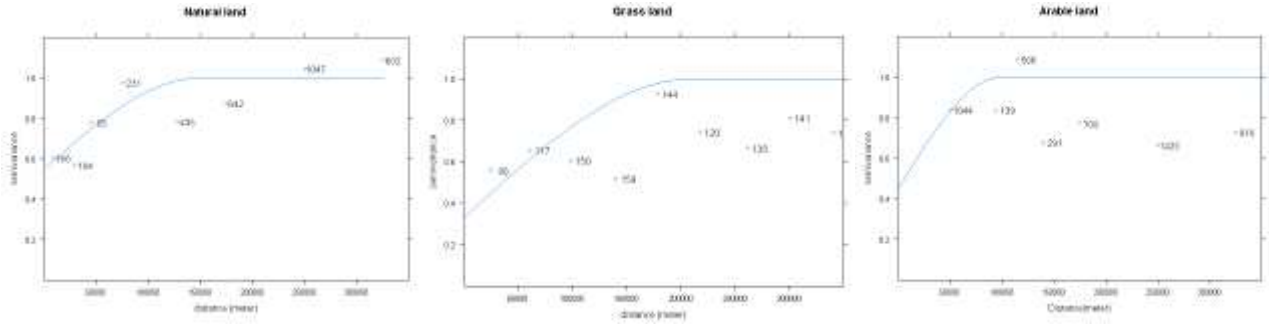
Histogram of OC



Box plot of OC



Fitted variogram of OC

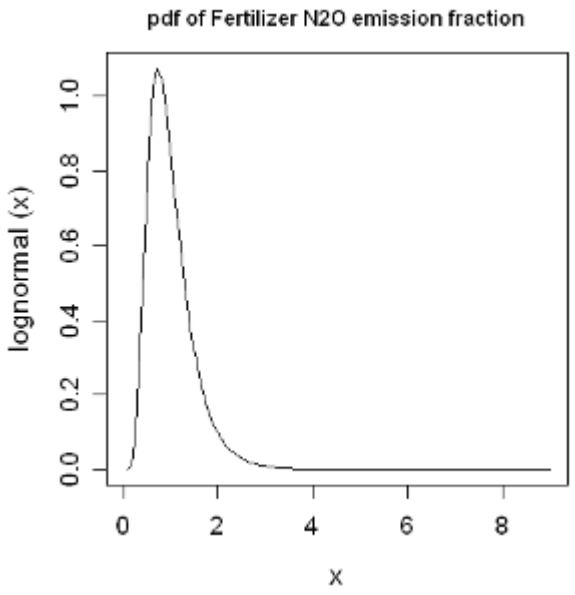


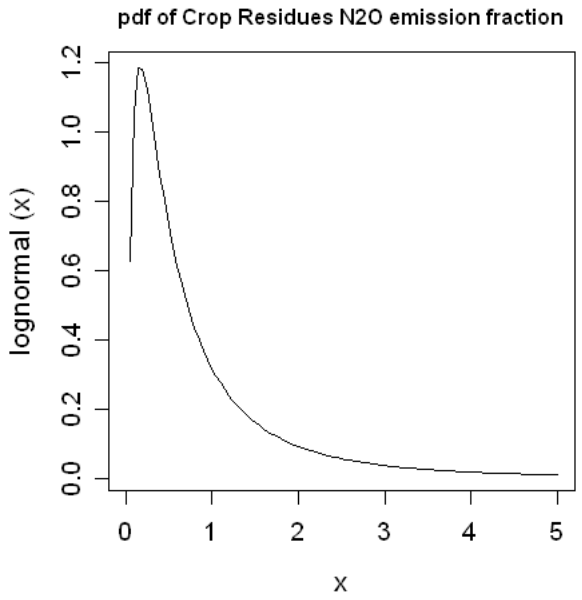
8.2 Appendix 2: Data and extra results related to Emission factors

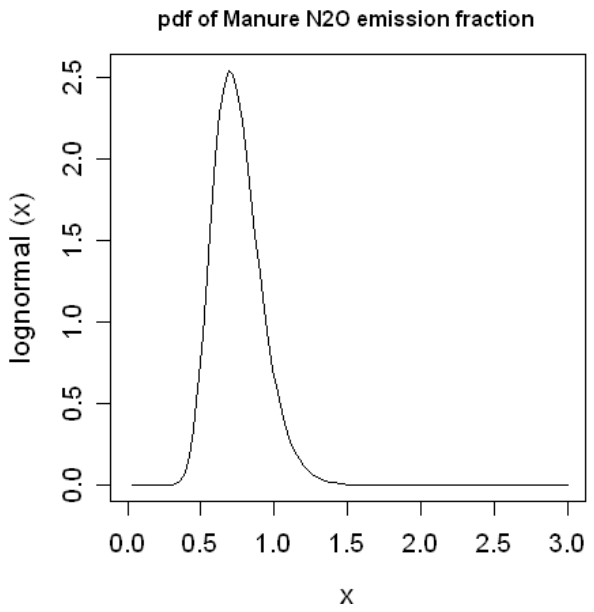
2.1 Data used to quantify EFs uncertainty

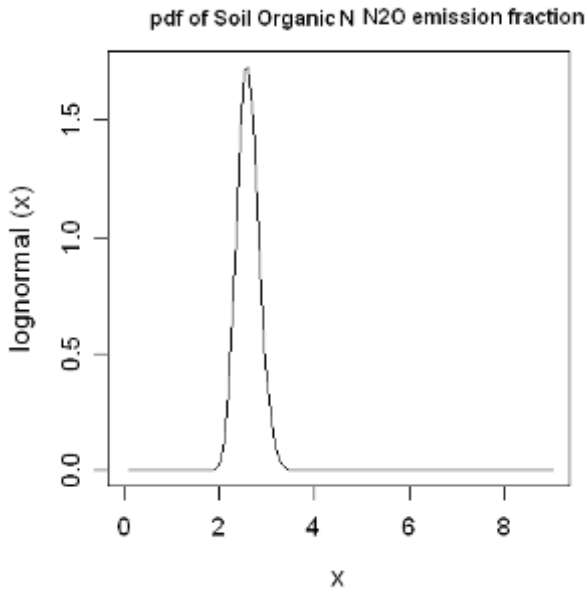
N source	Type	Ratio		Mean	5%	95%	Reference
Fertilizer	nitrate fertilizer			1.00%	0.50%	8.00%	(Clayton et al., 1997; Dobbie and Smith, 2003; Jones et al., 2005 & 2007; Velthof et al., 1997).
	ammonium fertilizer	0.50	× nitrate fertilizer				
	urea	1.50	× ammonium fertilizer				
Manure	pig slurry			0.75%	0.12%	0.97%	Chadwick et al. (2000).
	pig slurry	0.67	× incorporation				
	pig solid manure	0.33	× pig slurry				
	cattle soilid manure	0.33	× pig slurry				
	cattle slurry	0.67	× pig slurry				
	cattle slurry	0.67	× incorporation				
	poultry manure	0.33	× pig slurry				
	grazing	2.00	× nitrate fertilizer				
	other manure		average poultry, pig solid, cattle solid				
Soil organic N	Net mineralization			2.60%	1.80%	2.90%	Vinther et al. (2004)
Crop residues	arable crops			1.00%	0.13%	7%	Expert Judgement

2.2 Probability density distribution of four emission factors

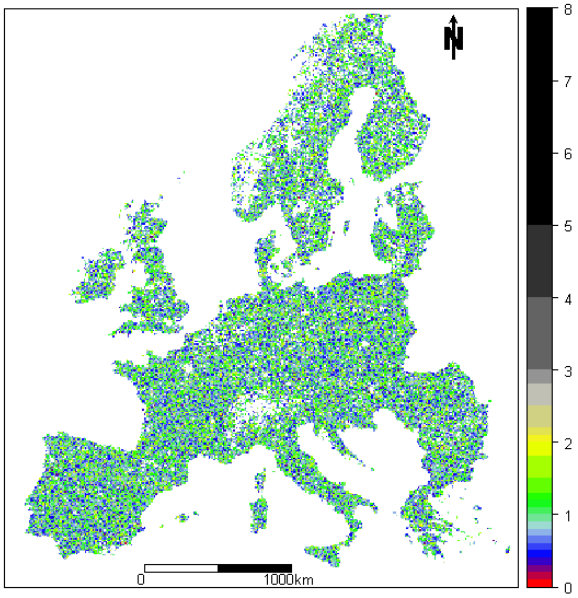




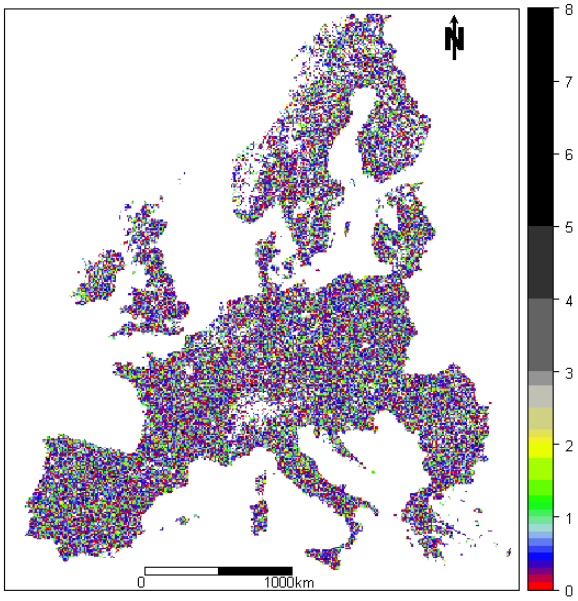




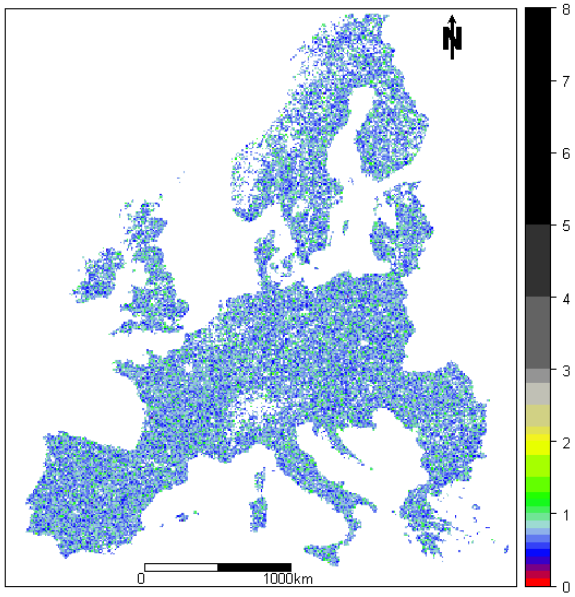
2.3 Simulated maps of EFs



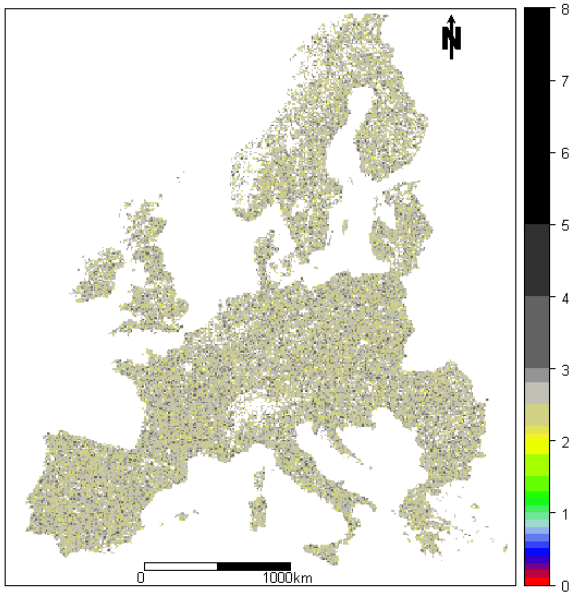
(a) Simulation map of Fertilizer emission factor



(b) Simulation map of Crop residuals emission factor



(c) Simulation map of manure emission factor

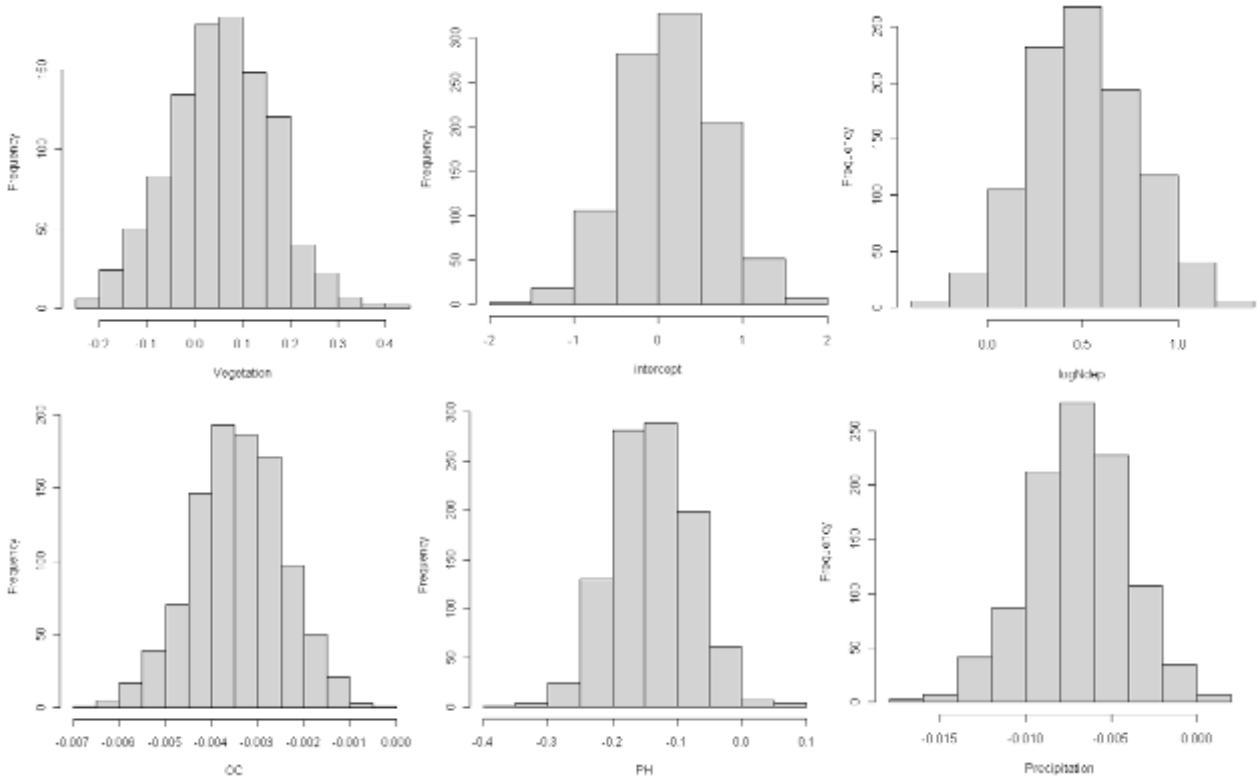


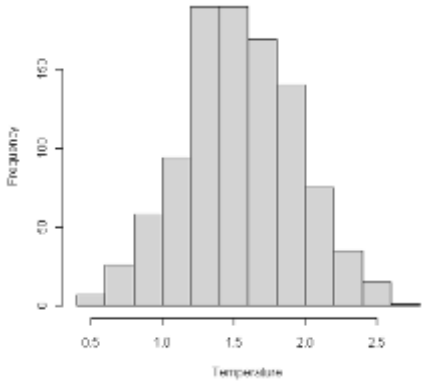
(d) Simulation map of organic N emission factor

8.3 Appendix 3: Data and extra results related to Regression model

3. 1 Uncertainty of Regression coefficients

Histogram of coefficients





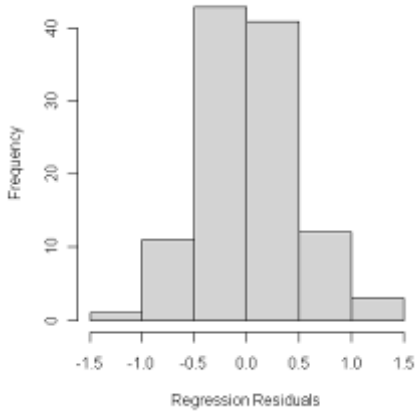
Comparison between variance of coefficients obtained from fitting regression model and from 1000 simulations

Coefficients	Variance from fitting	Variance from 1000 simulations
Intercept	0.3165881	***0.3329598***
Precipitation	0.0000088	***0.0000096***
FractionT<0	0.1574563	***0.1635866***
logNdep	0.0843314	***0.0035824***
pH	0.0037991	***0.0035824***
OC	0.0000010	***0.0000010***
Veg:dec	0.0113972	***0.0103550***

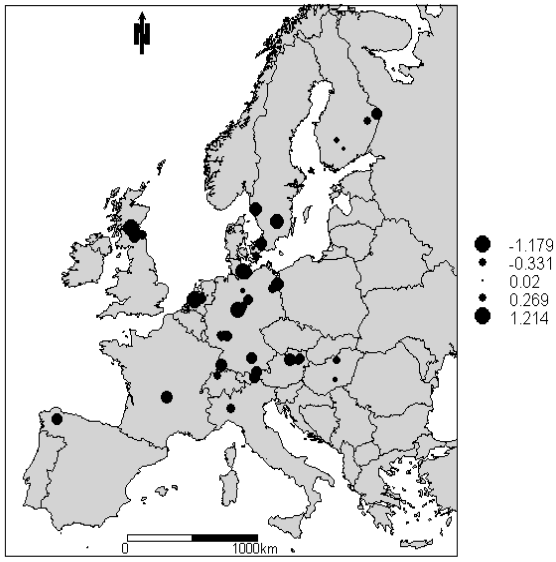
Variance – covariance matrix of regression coefficients

	Intercept	Precipitation	FractionT<0	logNdep	pH	OC	veg:dec
Intercept	0.3165881	***-0.0009268***	***-0.0712429***	***-0.1251004***	***-0.0121411***	***-0.0002802***	***-0.0116131***
Precipitation	-0.0007828	0.0000088	***-0.0000861***	***0.0001505***	***-0.0000009***	***0.0000008***	***0.0000784***
FractionT<0	-0.0690733	-0.0001434	0.1574563	***0.0448287***	***-0.0030788***	***-0.0000428***	***-0.0051885***
logNdep	-0.1184398	0.0001197	0.0444586	0.0843314	***-0.0039784***	***0.0000992***	***0.0024992***
pH	-0.0127540	-0.0000063	-0.0021900	-0.0041685	0.0037991	***0.0000102***	***-0.0002853***
OC	-0.0002627	0.0000007	-0.0000535	0.0000984	0.0000071	0.0000010	***0.0000127***
veg:dec	-0.0117172	0.0000790	-0.0072396	0.0025470	-0.0003727	0.0000233	0.0113972

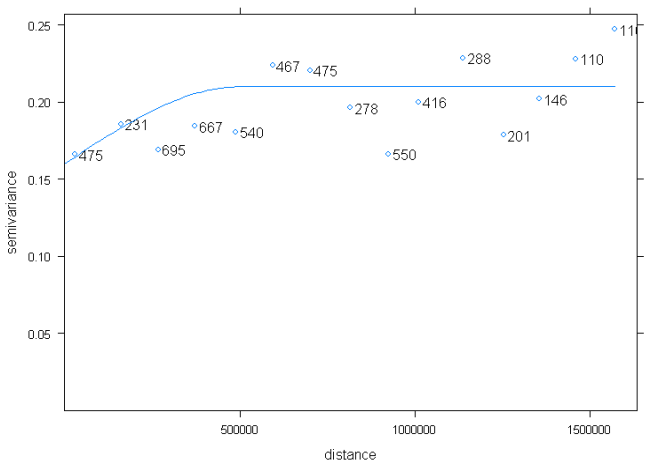
3.2 Regression residuals



Normal distribution of the regression residual



Llocation of observations used to fit regression model



Fitting the variogram for regression residuals

8.4 Appendix 4: Extensive uncertainty analysis results

4.1 Description of uncertain inputs and parameters in INTEGRATOR

Inputs/ Parameters	Units	Type	pdf	Mean		Range	
						5%	95%
Emission factors	%	Model parameter	Log-normal	Fertilizer	1.00	0.50	8.00
				Crop residues	1.00	0.13	7.00
				Soil organic N	2.60	1.80	2.90
				Manure	0.75	0.12	0.97
Regression coefficients	-	Model parameter	Multivariate normal distribution	β_0	0.133478		
				β_1	-0.006914		
				β_2	1.536506		
				β_3	0.497303		
				β_4	-0.137458		
				β_5	-0.003439		
				β6	0.055959		
Regression residual		Model parameter	Normal	0			
Soil pH	-	Model input		Varied in cluster			
Soil OC	%	Model input	Log-normal	Varied in cluster			

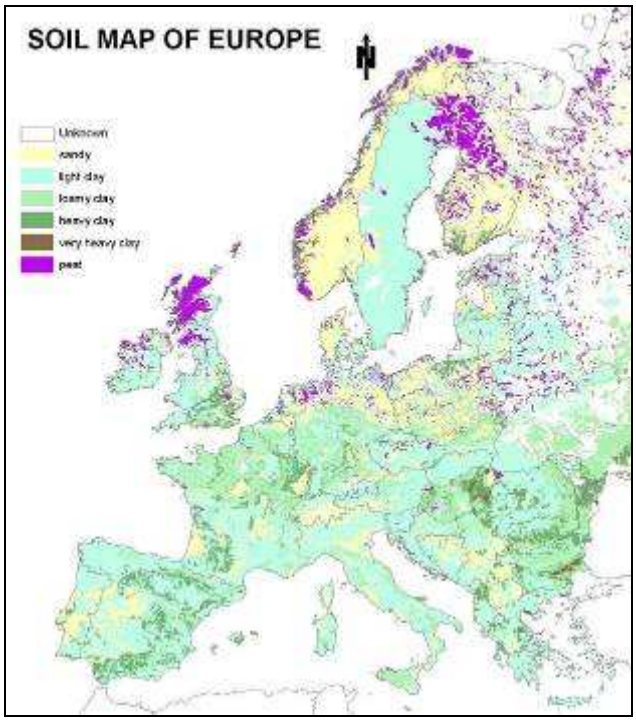
4.2 N_2O soil emission estimated by INTERGRATOR and uncertainty interval

Country	Average N2O soil emission (kg N/ha/yr)	Uncertainty interval (kg N/ha/yr)
Finland	1.429	± 0.342
Latvia	1.178	± 0.331
Denmark	18.571	± 0.105
Lithuania	1.407	± 0.281
Germany	3.006	± 0.115
Netherlands	17.390	± 0.203
Poland	3.378	± 0.202
Belgium	5.461	± 0.234
Czech Republic	2.568	± 0.400
Luxembourg	3.465	± 0.705
Slovakia	2.229	± 0.395
Austria	2.063	± 0.334
Switzerland	1.273	± 0.396
Liechtenstein	1.862	± 2.670
Italy	2.182	± 0.129
Slovenia	1.311	± 0.300
Hungary	9.048	± 0.109
France	2.610	± 0.078
Romania	1.776	± 0.230
Spain	1.141	± 0.081
Bulgaria	1.241	± 0.251
Greece	1.079	± 0.199
Portugal	1.881	± 0.108
Norway	1.702	± 0.477
Sweden	2.066	± 0.452
Estonia	0.873	± 0.268
United Kingdom	5.936	± 0.042
Ireland	5.676	± 0.123
Croatia	0.919	± 0.244

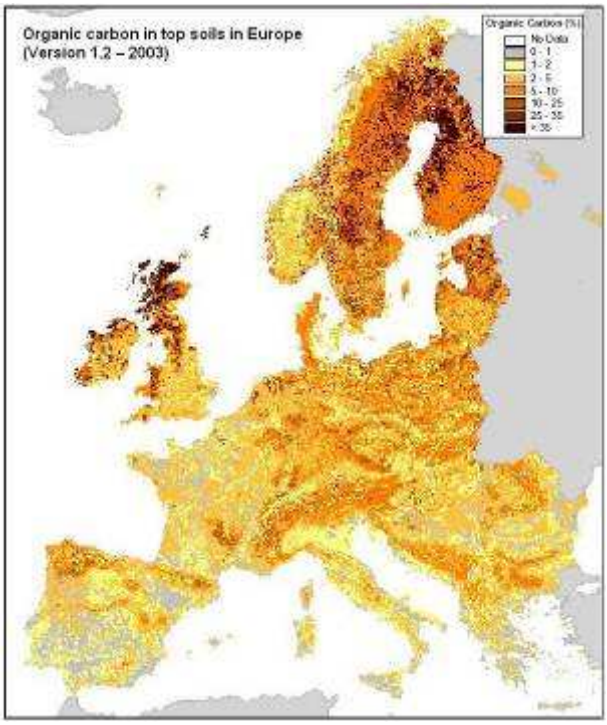
4.3 Percentage of uncertainty contribution in European country at point scale

Country	Percentage of uncertainty contribution			
	Organic Carbon	pH	Emission Factors	Regression Model
Finland	39.26%	0.65%	6.80%	53.30%
Latvia	21.24%	0.85%	37.42%	40.49%
Denmark	4.67%	0.37%	82.00%	12.96%
Lithuania	18.44%	0.86%	50.84%	29.85%
Germany	9.48%	0.46%	60.25%	29.80%
Netherlands	16.41%	0.33%	75.57%	7.69%
Poland	9.66%	0.37%	59.76%	30.21%
Belgium	12.59%	1.33%	59.86%	26.22%
Czech Republic	4.12%	0.57%	58.34%	36.97%
Luxembourg	3.86%	1.09%	36.31%	58.74%
Slovakia	3.84%	0.38%	48.95%	46.84%
Austria	7.11%	0.81%	33.12%	58.96%
Switzerland	8.79%	0.39%	27.57%	63.24%
Liechtenstein*	4.96%	0.82%	0.00%	94.22%
Italy	7.54%	1.12%	54.53%	36.81%
Slovenia	8.34%	1.13%	17.26%	73.26%
Hungary	4.38%	0.53%	71.44%	23.64%
France	8.48%	0.62%	60.45%	30.45%
Romania	2.07%	0.26%	60.80%	36.87%
Spain	4.67%	0.52%	61.86%	32.95%
Bulgaria	5.10%	0.86%	56.16%	37.88%
Greece	10.62%	1.22%	44.97%	43.19%
Portugal	2.38%	0.34%	53.81%	43.47%
Norway	14.74%	0.70%	7.48%	77.08%
Sweden	22.03%	1.15%	7.53%	69.29%
Estonia	12.27%	0.16%	29.28%	58.29%
United Kingdom	10.72%	0.53%	52.82%	35.94%
Ireland	0.64%	0.01%	91.99%	7.37%
Croatia	25.61%	2.80%	29.33%	42.26%

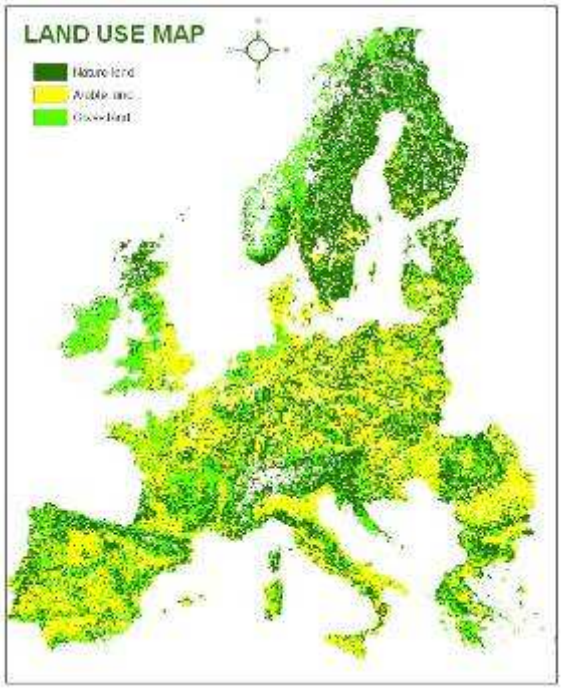
8.5 Appendix 5: Map of soil texture, land use cover and Organic carbon v1.2



Soil texture map of Europe



Map of Organic carbon



Map of three different land use cover

8.6 Appendix 6: Developed R code

Quantify pH uncertainty

```
#Part 1: read data
memory.limit(size=4095)
rm(list = ls())    # clean up working environment
graphics.off()    # terminate graphics devices
library(foreign)
library(sp)
library(rgdal)
library(gstat)
library(lattice)
library(maptools)
library(RColorBrewer)
setwd("H:\\phuong\\pH_natural")
clPH = read.table("clusterPH.txt",header=TRUE,na.string = "-9",sep="\t")
clPH = subset(clPH,(c2==1 | c3==1 | c4==1 | c5==1 | c6==1 | c8==1 | c9==1 | c10==1 |
c17==1 | c18==1 |c20==1 | c21==1))
PHdata = subset(clPH,!is.na(clPH$PH))
attach(PHdata)
clPH = NULL

PHdata$x = PHdata$x + 1 # move 1 meter to avoid exact overlap
coordinates(PHdata) = ~x + y
#------ make set of cluster grids ------
cluster = readGDAL("phnat_clus2.asc")
cluster$c2 = cluster$band1
cluster$c3 = readGDAL("phnat_clus3.asc")$band1
cluster$c4 = readGDAL("phnat_clus4.asc")$band1
cluster$c5 = readGDAL("phnat_clus5.asc")$band1
cluster$c6 = readGDAL("phnat_clus6.asc")$band1
cluster$c8 = readGDAL("phnat_clus8.asc")$band1
cluster$c9 = readGDAL("phnat_clus9.asc")$band1
cluster$c10 = readGDAL("phnat_clus10.asc")$band1
cluster$c17 = readGDAL("phnat_clus17.asc")$band1
cluster$c18 = readGDAL("phnat_clus18.asc")$band1
cluster$c20 = readGDAL("phnat_clus20.asc")$band1
cluster$c21 = readGDAL("phnat_clus21.asc")$band1
cluster$band1=NULL

eu_poly = readShapePoly("world_admin_esri_EUcountries_15m_etrs3035.shp") # read
simplified EU boundary
#Part 2: illustrate data original pH data
windows(width = 5,height = 5)
par(mar=c(4,4,1,1)+0.9,cex.lab=1,cex.axis=1)
hist(PH,main='',xlab='pH',col='lightgrey')
savePlot(filename="Image_results\\histpH",type="png")

PHdata$ph2 = 0*PHdata$PH
PHdata$ph3 = 0*PHdata$PH
PHdata$ph4 = 0*PHdata$PH
PHdata$ph5 = 0*PHdata$PH
PHdata$ph6 = 0*PHdata$PH
PHdata$ph8 = 0*PHdata$PH
PHdata$ph9 = 0*PHdata$PH
PHdata$ph10 = 0*PHdata$PH
PHdata$ph17 = 0*PHdata$PH
PHdata$ph18 = 0*PHdata$PH
PHdata$ph20 = 0*PHdata$PH
PHdata$ph21 = 0*PHdata$PH

for (i in 1:length(PHdata$PH))
{ if (clusternr[i]==2) PHdata$ph2[i]=PH[i] else PHdata$ph2[i]=NaN
+ if (clusternr[i]==3) PHdata$ph3[i]=PH[i] else PHdata$ph3[i]=NaN
+ if (clusternr[i]==4) PHdata$ph4[i]=PH[i] else PHdata$ph4[i]=NaN
+ if (clusternr[i]==5) PHdata$ph5[i]=PH[i] else PHdata$ph5[i]=NaN
+ if (clusternr[i]==6) PHdata$ph6[i]=PH[i] else PHdata$ph6[i]=NaN
+ if (clusternr[i]==8) PHdata$ph8[i]=PH[i] else PHdata$ph8[i]=NaN
+ if (clusternr[i]==9) PHdata$ph9[i]=PH[i] else PHdata$ph9[i]=NaN
+ if (clusternr[i]==10) PHdata$ph10[i]=PH[i] else PHdata$ph10[i]=NaN
+ if (clusternr[i]==17) PHdata$ph17[i]=PH[i] else PHdata$ph17[i]=NaN
+ if (clusternr[i]==18) PHdata$ph18[i]=PH[i] else PHdata$ph18[i]=NaN
+ if (clusternr[i]==20) PHdata$ph20[i]=PH[i] else PHdata$ph20[i]=NaN
+ if (clusternr[i]==21) PHdata$ph21[i]=PH[i] else PHdata$ph21[i]=NaN}

windows(width=10, height=5)
par(mar=c(4,4,1,1)+0.6,cex.lab=1.4,cex.axis=1.5)
boxp =
 boxplot(PHdata$ph2,PHdata$ph3,PHdata$ph4,PHdata$ph5,PHdata$ph6,PHdata$ph8,PHdata$ph9,
```

```
 PHdata$ph10,PHdata$ph17,PHdata$ph18,PHdata$ph20,PHdata$ph21,boxwex=0.5,ylim=c(3,9),
ylab='PH',names=c("2","3","4","5","6","8","9","10","17","18","20","21"),col="lightgray")
savePlot(filename="Image_results\\boxplotPH",type="png")

windows(width = 7, height = 7)
spplot(PHdata,zcol="PH",xlim=c(2500000,6000000),ylim=c(1500000,5100000),
 key.space=list(x=0.02,y=0.96,corner=c(0,1)),sp.layout=list("sp.polygons",
 eu_poly),col.regions=brewer.pal(4,"Oranges"))
savePlot(filename = "Image_results\\PHdata", type = "png")

windows(width = 8, height = 7)
bubble(PHdata,zcol="PH",xlim=c(2500000,6000000),ylim=c(1500000,5100000),
sp.layout=list("sp.polygons",eu_poly,fill="lightgrey"),fill=TRUE,maxsize=1.4,col="black"
)
savePlot(filename = "Image_results\\PHbubble", type = "png")

# Part 3: data transformation
PHdata$PHtransf = PHdata$PH
for (i in 1:length(PHdata$PHtransf))
{if (clusternr[i]==2) PHdata$PHtransf[i] =
  (PHdata$PH[i]- mean(PHdata$ph2,na.rm=TRUE))/sd(PHdata$ph2,na.rm=TRUE)
 if (clusternr[i]==3) PHdata$PHtransf[i] =
  (PHdata$PH[i]- mean(PHdata$ph3,na.rm=TRUE))/sd(PHdata$ph3,na.rm=TRUE)
 if (clusternr[i]==4) PHdata$PHtransf[i] =
  (PHdata$PH[i]- mean(PHdata$ph4,na.rm=TRUE))/sd(PHdata$ph4,na.rm=TRUE)
 if (clusternr[i]==5) PHdata$PHtransf[i] =
  (PHdata$PH[i]- mean(PHdata$ph5,na.rm=TRUE))/sd(PHdata$ph5,na.rm=TRUE)
 if (clusternr[i]==6) PHdata$PHtransf[i] =
  (PHdata$PH[i]- mean(PHdata$ph6,na.rm=TRUE))/sd(PHdata$ph6,na.rm=TRUE)
 if (clusternr[i]==8) PHdata$PHtransf[i] =
  (PHdata$PH[i]- mean(PHdata$ph8,na.rm=TRUE))/sd(PHdata$ph8,na.rm=TRUE)
 if (clusternr[i]==9) PHdata$PHtransf[i] =
  (PHdata$PH[i]- mean(PHdata$ph9,na.rm=TRUE))/sd(PHdata$ph9,na.rm=TRUE)
 if (clusternr[i]==10) PHdata$PHtransf[i] =
  (PHdata$PH[i]- mean(PHdata$ph10,na.rm=TRUE))/sd(PHdata$ph10,na.rm=TRUE)
 if (clusternr[i]==17) PHdata$PHtransf[i] =
  (PHdata$PH[i]- mean(PHdata$ph17,na.rm=TRUE))/sd(PHdata$ph17,na.rm=TRUE)
 if (clusternr[i]==18) PHdata$PHtransf[i] =
  (PHdata$PH[i]- mean(PHdata$ph18,na.rm=TRUE))/sd(PHdata$ph18,na.rm=TRUE)
 if (clusternr[i]==20) PHdata$PHtransf[i] =
  (PHdata$PH[i]- mean(PHdata$ph20,na.rm=TRUE))/sd(PHdata$ph20,na.rm=TRUE)
 if (clusternr[i]==21) PHdata$PHtransf[i] =
  (PHdata$PH[i]- mean(PHdata$ph21,na.rm=TRUE))/sd(PHdata$ph21,na.rm=TRUE)
}
windows(width = 5,height = 5)
par(mar=c(4,4,1,1)+0.6,cex.lab=1.4,cex.axis=1.5)
hist(PHdata$PHtransf,main='',xlab='pH transformed',col='lightgrey')
savePlot(filename="D:\\PH_natural\\Image_results\\histpHtransf",type="png")

#Part 4: Estimate sample variogram and variogram fitting
fPH = gstat(id=c("PHtransf"),formula=PHtransf~1,beta=0,data=PHdata,nmax=24)
fPH.vg =
variogram(fPH,boundaries=c(7000,10000,20000,30000,40000,50000,70000,80000,90000,100000,1
50000,180000,200000))
fPH.vgm = fit.variogram(fPH.vg,vgm(0.57,"Sph",50000,0.43), fit.sill=FALSE)
windows(width = 8, height = 6)
plot(fPH.vg,fPH.vgm,plot.numbers = TRUE,xlab="distance (meter)",main="Natural
land",xlim=c(0,84000))
savePlot(filename="Image_results\\vrgam_PHtransfn",type="png")
fPH = gstat(fPH,id=c("PHtransf"),model=fPH.vgm)

#Part 6: Simulation
N=100
PHsim = predict.gstat(fPH,cluster,nsim=N,BLUE=FALSE)
PHsim = as(PHsim,"SpatialGridDataFrame")

PHsim$mean = mean(PHdata$ph2,na.rm=TRUE)*cluster$c2
PHsim$mean = PHsim$mean + mean(PHdata$ph3,na.rm=TRUE)*cluster$c3
PHsim$mean = PHsim$mean + mean(PHdata$ph4,na.rm=TRUE)*cluster$c4
PHsim$mean = PHsim$mean + mean(PHdata$ph5,na.rm=TRUE)*cluster$c5
PHsim$mean = PHsim$mean + mean(PHdata$ph6,na.rm=TRUE)*cluster$c6
PHsim$mean = PHsim$mean + mean(PHdata$ph8,na.rm=TRUE)*cluster$c8
PHsim$mean = PHsim$mean + mean(PHdata$ph9,na.rm=TRUE)*cluster$c9
PHsim$mean = PHsim$mean + mean(PHdata$ph10,na.rm=TRUE)*cluster$c10
PHsim$mean = PHsim$mean + mean(PHdata$ph17,na.rm=TRUE)*cluster$c17
PHsim$mean = PHsim$mean + mean(PHdata$ph18,na.rm=TRUE)*cluster$c18
PHsim$mean = PHsim$mean + mean(PHdata$ph20,na.rm=TRUE)*cluster$c20
PHsim$mean = PHsim$mean + mean(PHdata$ph21,na.rm=TRUE)*cluster$c21

PHsim$sd = sd(PHdata$ph2,na.rm=TRUE)*cluster$c2
```

```
PHsim$sd = PHsim$sd + sd(PHdata$ph3,na.rm=TRUE)*cluster$c3
PHsim$sd = PHsim$sd + sd(PHdata$ph4,na.rm=TRUE)*cluster$c4
PHsim$sd = PHsim$sd + sd(PHdata$ph5,na.rm=TRUE)*cluster$c5
PHsim$sd = PHsim$sd + sd(PHdata$ph6,na.rm=TRUE)*cluster$c6
PHsim$sd = PHsim$sd + sd(PHdata$ph8,na.rm=TRUE)*cluster$c8
PHsim$sd = PHsim$sd + sd(PHdata$ph9,na.rm=TRUE)*cluster$c9
PHsim$sd = PHsim$sd + sd(PHdata$ph10,na.rm=TRUE)*cluster$c10
PHsim$sd = PHsim$sd + sd(PHdata$ph17,na.rm=TRUE)*cluster$c17
PHsim$sd = PHsim$sd + sd(PHdata$ph18,na.rm=TRUE)*cluster$c18
PHsim$sd = PHsim$sd + sd(PHdata$ph20,na.rm=TRUE)*cluster$c20
PHsim$sd = PHsim$sd + sd(PHdata$ph21,na.rm=TRUE)*cluster$c21
cluster=NULL
PHdata=NULL
for (i in 1:N)
{
sim=paste("sim",as.character(i),sep="")
PHsim[[sim]] = PHsim$sd*PHsim[[sim]]+ PHsim$mean
}
PHsim$sd=NULL
PHsim$mean=NULL
# Create .txt outputs
PHsimulation=as.data.frame(PHsim)
ID=c(1:length(PHsimulation$sim1))
PHsimulation$ID=ID
for (i in 1:N)
{
colna=c("x","y","ID",paste("sim",as.character(i),sep=""))
PHsimu= subset(PHsimulation,select=colna)
write.table(PHsimu, file= paste("For_UA_input\\phnatsim",as.character(i),".txt",sep=""),
col.names=c("x","y","ID",paste("sim",as.character(i),sep="")),row.names=FALSE,
quote=FALSE, sep=" ", na="-9999")
}
```

Quantify OC uncertainty

```
#Part 1: Read data
memory.limit(size=4095)
rm(list = ls())    # clean up working environment
graphics.off()    # terminate graphics devices
library(foreign)
library(sp)
library(rgdal)
library(gstat)
library(lattice)
library(maptools)
library(RColorBrewer)
setwd("H:\\phuong\\OC_natural") #set working directory
clOC = read.table("clusterOC.txt",header=TRUE, na.string="-9") #read .txt data
clOC = subset(clOC,(c1==1 | c2==1 | c4==1 | c5==1 | c6==1 | c13==1 | c14==1 | c15==1 |
c18==1 | c19==1 |c21==1))
OCdata = subset(clOC,!is.na(clOC$OC))
attach(OCdata)
clOC = NULL
OCdata$x = OCdata$x + 1 # move 1 meter to avoid exact overlap
coordinates(OCdata) = ~x + y #create SpatialPointsDataFrame
#------ make set of cluster grids ------
cluster = readGDAL("ocnat_clus1.asc")
cluster$c1 = cluster$band1
cluster$c2 = readGDAL("ocnat_clus2.asc")$band1
cluster$c4 = readGDAL("ocnat_clus4.asc")$band1
cluster$c5 = readGDAL("ocnat_clus5.asc")$band1
cluster$c6 = readGDAL("ocnat_clus6.asc")$band1
cluster$c13 = readGDAL("ocnat_clus13.asc")$band1
cluster$c14 = readGDAL("ocnat_clus14.asc")$band1
cluster$c15 = readGDAL("ocnat_clus15.asc")$band1
cluster$c18 = readGDAL("ocnat_clus18.asc")$band1
cluster$c19 = readGDAL("ocnat_clus19.asc")$band1
cluster$c21 = readGDAL("ocnat_clus21.asc")$band1
cluster$band1=NULL
spplot(cluster,col.regions=grey(c(0.8,0)),as.table=TRUE)
savePlot(filename = "Image_results\\OCnat_clusters", type = "png")
eu_poly = readShapePoly("world_admin_esri_EUcountries_15m_etrs3035.shp") # read
simplified EU boundary
#-------------------------------------------------
#Part 2: illustrate OC original data
#OC histogram
windows(width = 5,height = 5)
par(mar=c(4,4,1,1)+0.9,cex.lab=1,cex.axis=1)
hist(OC,freq=TRUE,breaks=c(0,5,10,15,20,25,30,35,40,45,50,55,60,65,70,75,80),main='',xla
b='Soil Organic Carbon',col='lightgrey')
```

```
savePlot(filename="Image_results\\histOCnat",type="png")
# making box plot
OCdata$OC1 = 0*OCdata$OC
OCdata$OC2 = 0*OCdata$OC
OCdata$OC4 = 0*OCdata$OC
OCdata$OC5 = 0*OCdata$OC
OCdata$OC6 = 0*OCdata$OC
OCdata$OC13 = 0*OCdata$OC
OCdata$OC14 = 0*OCdata$OC
OCdata$OC15 = 0*OCdata$OC
OCdata$OC18 = 0*OCdata$OC
OCdata$OC19 = 0*OCdata$OC
OCdata$OC21 = 0*OCdata$OC
for (i in 1:length(OCdata$OC))
+ {if (clusternr[i]==1)OCdata$OC1[i]=OC[i] else OCdata$OC1[i]=NaN
+ if (clusternr[i]==2) OCdata$OC2[i]=OC[i] else OCdata$OC2[i]=NaN
+ if (clusternr[i]==4) OCdata$OC4[i]=OC[i] else OCdata$OC4[i]=NaN
+ if (clusternr[i]==5) OCdata$OC5[i]=OC[i] else OCdata$OC5[i]=NaN
+ if (clusternr[i]==6) OCdata$OC6[i]=OC[i] else OCdata$OC6[i]=NaN
+ if (clusternr[i]==13) OCdata$OC13[i]=OC[i] else OCdata$OC13[i]=NaN
+ if (clusternr[i]==14) OCdata$OC14[i]=OC[i] else OCdata$OC14[i]=NaN
+ if (clusternr[i]==15) OCdata$OC15[i]=OC[i] else OCdata$OC15[i]=NaN
+ if (clusternr[i]==18) OCdata$OC18[i]=OC[i] else OCdata$OC18[i]=NaN
+ if (clusternr[i]==19) OCdata$OC19[i]=OC[i] else OCdata$OC19[i]=NaN
+ if (clusternr[i]==21) OCdata$OC21[i]=OC[i] else OCdata$OC21[i]=NaN}

windows(width=10, height=5)
par(mar=c(4,4,1,1)+0.6,cex.lab=1,cex.axis=1)
boxp =
boxplot(OCdata$OC1,OCdata$OC2,OCdata$OC4,OCdata$OC5,OCdata$OC6,OCdata$OC13,OCdata$OC14,
 OCdata$OC15,OCdata$OC18,OCdata$OC19,OCdata$OC21,boxwex=0.5,ylim=c(0,80),
 ylab='Soil Organic carbon',
 names=c("1","2","4","5","6","13","14","15","18","19","21"),col="lightgrey")
savePlot(filename="Image_results\\boxplotOCnat",type="png")
# map of sample location
windows(width = 7, height = 7)
spplot(OCdata,zcol="OC",xlim=c(2500000,6000000),ylim=c(1500000,5100000),
 key.space=list(x=0.02,y=0.96,corner=c(0,1)),sp.layout=list("sp.polygons",
 eu_poly),col.regions=brewer.pal(3,"Dark2"),main="Soil Organic Carbon" )
savePlot(filename = "Image_results\\OCnat_sample_point", type = "png")
# bubble map of sample
windows(width = 8, height = 7)
bubble(OCdata,zcol="OC",xlim=c(2500000,6000000),ylim=c(1500000,5100000),
 sp.layout=list("sp.polygons",eu_poly,fill="lightgrey"),fill=TRUE,maxsize=2, main="Soil
Organic Carbon",col="black")
savePlot(filename = "Image_results\\OCbubble", type = "png")
#-----------------------------------------------------------
#Part 3: OC log-natural
OCdata$OClog=log(OC)

OCdata$OClog1 = log(OCdata$OC1)
OCdata$OClog2 = log(OCdata$OC2)
OCdata$OClog4 = log(OCdata$OC4)
OCdata$OClog5 = log(OCdata$OC5)
OCdata$OClog6 = log(OCdata$OC6)
OCdata$OClog13 = log(OCdata$OC13)
OCdata$OClog14 = log(OCdata$OC14)
OCdata$OClog15 = log(OCdata$OC15)
OCdata$OClog18 = log(OCdata$OC18)
OCdata$OClog19 = log(OCdata$OC19)
OCdata$OClog21 = log(OCdata$OC21)
#-----------------------------------------------------------
#Part 4: OC transformation (taking into account the variantion in clusters)
OCdata$OCtransf = OCdata$OClog
for (i in 1:length(OCdata$OCtransf))
{if (clusternr[i]==1) OCdata$OCtransf[i] =
  (OCdata$OClog[i]- mean(OCdata$OClog1,na.rm=TRUE))/sd(OCdata$OClog1,na.rm=TRUE)
 if (clusternr[i]==2) OCdata$OCtransf[i] =
  (OCdata$OClog[i]- mean(OCdata$OClog2,na.rm=TRUE))/sd(OCdata$OClog2,na.rm=TRUE)
 if (clusternr[i]==4) OCdata$OCtransf[i] =
  (OCdata$OClog[i]- mean(OCdata$OClog4,na.rm=TRUE))/sd(OCdata$OClog4,na.rm=TRUE)
 if (clusternr[i]==5) OCdata$OCtransf[i] =
  (OCdata$OClog[i]- mean(OCdata$OClog5,na.rm=TRUE))/sd(OCdata$OClog5,na.rm=TRUE)
 if (clusternr[i]==6) OCdata$OCtransf[i] =
  (OCdata$OClog[i]- mean(OCdata$OClog6,na.rm=TRUE))/sd(OCdata$OClog6,na.rm=TRUE)
 if (clusternr[i]==13) OCdata$OCtransf[i] =
  (OCdata$OClog[i]- mean(OCdata$OClog13,na.rm=TRUE))/sd(OCdata$OClog13,na.rm=TRUE)
 if (clusternr[i]==14) OCdata$OCtransf[i] =
  (OCdata$OClog[i]- mean(OCdata$OClog14,na.rm=TRUE))/sd(OCdata$OClog14,na.rm=TRUE)
 if (clusternr[i]==15) OCdata$OCtransf[i] =
```

```
  (OCdata$OClog[i]- mean(OCdata$OClog15,na.rm=TRUE))/sd(OCdata$OClog15,na.rm=TRUE)
 if (clusternr[i]==18) OCdata$OCtransf[i] =
  (OCdata$OClog[i]- mean(OCdata$OClog18,na.rm=TRUE))/sd(OCdata$OClog18,na.rm=TRUE)
 if (clusternr[i]==19) OCdata$OCtransf[i] =
  (OCdata$OClog[i]- mean(OCdata$OClog19,na.rm=TRUE))/sd(OCdata$OClog19,na.rm=TRUE)
 if (clusternr[i]==21) OCdata$OCtransf[i] =
  (OCdata$OClog[i]- mean(OCdata$OClog21,na.rm=TRUE))/sd(OCdata$OClog21,na.rm=TRUE)
}
OCdata$OC1=NULL
OCdata$OC2=NULL
OCdata$OC4=NULL
OCdata$OC5=NULL
OCdata$OC6=NULL
OCdata$OC13=NULL
OCdata$OC14=NULL
OCdata$OC15=NULL
OCdata$OC18=NULL
OCdata$OC19=NULL
OCdata$OC21=NULL
OCdata$OC=NULL
# histogram of ε
windows(width = 5,height = 5)
par(mar=c(4,4,1,1)+0.6,cex.lab=1,cex.axis=1)
hist(OCdata$OCtransf,main='',xlab='OC transformed',col='lightgrey', freq=FALSE)
savePlot(filename="Image_results\\histOC_nat",type="png")
# bubble map of ε
windows(width = 8, height = 7)
bubble(OCdata,zcol="OCtransf",xlim=c(2500000,6000000),ylim=c(1500000,5100000),
 sp.layout=list("sp.polygons",eu_poly),fill=TRUE,maxsize=2, main="Soil OC
 transformed",col="black")
savePlot(filename = "Image_results\\OCnat_2transf_bubble", type = "png")
#--------------------------------
#Part 5: Fit variogram of ε
fOC = gstat(id=c("OCtransf"),formula=OCtransf~1,beta=0,data=OCdata,nmax=24)
fOC.vg =
variogram(fOC,boundaries=c(20000,40000,50000,100000,150000,200000,300000,350000))
fOC.vgm = fit.variogram(fOC.vg,vgm(0.45,"Sph",300000,0.55),fit.sills=FALSE)
windows(width = 8, height = 6)
plot(fOC.vg,fOC.vgm,plot.numbers = TRUE,ylim=c(0,1.2),xlim=c(0,350000),xlab="distance
(meter)",main="Natural land")
savePlot(filename="Image_results\\vrgam_OCnat2transfn",type="png")
fOC = gstat(fOC,id=c("OCtransf"),model=fOC.vgm)

# Part 6: Simulation
N=100
OCsim = predict.gstat(fOC,cluster,nsim=N,BLUE=FALSE)
OCsim = as(OCsim,"SpatialGridDataFrame")
#compute OC from ε
OCsim$meanlog<-mean(OCdata$OClog1,na.rm=TRUE)*cluster$c1
OCsim$meanlog=OCsim$meanlog+mean(OCdata$OClog2,na.rm=TRUE)*cluster$c2
OCsim$meanlog=OCsim$meanlog+mean(OCdata$OClog4,na.rm=TRUE)*cluster$c4
OCsim$meanlog=OCsim$meanlog+mean(OCdata$OClog5,na.rm=TRUE)*cluster$c5
OCsim$meanlog=OCsim$meanlog+mean(OCdata$OClog6,na.rm=TRUE)*cluster$c6
OCsim$meanlog=OCsim$meanlog+mean(OCdata$OClog13,na.rm=TRUE)*cluster$c13
OCsim$meanlog=OCsim$meanlog+mean(OCdata$OClog14,na.rm=TRUE)*cluster$c14
OCsim$meanlog=OCsim$meanlog+mean(OCdata$OClog15,na.rm=TRUE)*cluster$c15
OCsim$meanlog=OCsim$meanlog+mean(OCdata$OClog18,na.rm=TRUE)*cluster$c18
OCsim$meanlog=OCsim$meanlog+mean(OCdata$OClog19,na.rm=TRUE)*cluster$c19
OCsim$meanlog=OCsim$meanlog+mean(OCdata$OClog21,na.rm=TRUE)*cluster$c21

OCsim$sdlog=sd(OCdata$OClog1,na.rm=TRUE)*cluster$c1
OCsim$sdlog=OCsim$sdlog+sd(OCdata$OClog2,na.rm=TRUE)*cluster$c2
OCsim$sdlog=OCsim$sdlog+sd(OCdata$OClog4,na.rm=TRUE)*cluster$c4
OCsim$sdlog=OCsim$sdlog+sd(OCdata$OClog5,na.rm=TRUE)*cluster$c5
OCsim$sdlog=OCsim$sdlog+sd(OCdata$OClog6,na.rm=TRUE)*cluster$c6
OCsim$sdlog=OCsim$sdlog+sd(OCdata$OClog13,na.rm=TRUE)*cluster$c13
OCsim$sdlog=OCsim$sdlog+sd(OCdata$OClog14,na.rm=TRUE)*cluster$c14
OCsim$sdlog=OCsim$sdlog+sd(OCdata$OClog15,na.rm=TRUE)*cluster$c15
OCsim$sdlog=OCsim$sdlog+sd(OCdata$OClog18,na.rm=TRUE)*cluster$c18
OCsim$sdlog=OCsim$sdlog+sd(OCdata$OClog19,na.rm=TRUE)*cluster$c19
OCsim$sdlog=OCsim$sdlog+sd(OCdata$OClog21,na.rm=TRUE)*cluster$c21
cluster=NULL
OCdata=NULL
for (i in 1:N)
{
sim=paste("sim",as.character(i),sep="")
OCsim[[sim]] = exp(OCsim$sdlog*OCsim[[sim]]+ OCsim$meanlog)
}
OCsim$meanlog=NULL
```

```
OCsim$sdlog=NULL
# create .txt outputs
OCsimulation=as.data.frame(OCsim)
for (i in 1:N)
{
sim=paste("sim",as.character(i),sep="")
for (j in 1:length(OCsimulation$sim1))
OCsimulation[[sim]]=ifelse (OCsimulation[[sim]]>100,100, OCsimulation[[sim]])
}
ID=c(1:length(OCsimulation$sim1))
OCsimulation$ID=ID
for (i in 1:N)
{
colna=c("x","y","ID",paste("sim",as.character(i),sep=""))
OCsimu= subset(OCsimulation,select=colna)
write.table(OCsimu, file=
paste("For_UA_input\\ocnatsim",as.character(i),".txt",sep=""),col.names=c("x","y","ID",p
aste("sim",as.character(i),sep="")),row.names=FALSE, quote=FALSE, sep=" ", na="-9999")
}
```

Quantify uncertainty of Emission Factor

```
#load packages
library(stats)
library(sp)
library(rgdal)
library(gstat)
rm(list = ls())
memory.limit(size=4095)
#---------------------------------------------------------------------------
#Function to estimate emission factor distribution parameters, mu and sigma
distparest<-function(mY,p5Y,p95Y)
{
p5X=log(p5Y)
p95X=log(p95Y)
mX=log(mY)
G<-function(sigma) (pnorm(p95X, mean=mX-(sigma^2)/2, sd=sigma)-pnorm(p5X, mean=mX-
(sigma^2)/2, sd=sigma))
x=0
Gre=1
while (Gre>0.9)
{
x=x+0.0000001
Gre=G(x)
}
sigma=x
pnorm(p95X, mean=mX-(sigma^2)/2, sd=sigma)
pnorm(p5X, mean=mX-(sigma^2)/2, sd=sigma)
return(sigma)
}
#---------------------------------------------------------------------------
#Function to simulate over Europe
EFsim <-function(mu,sigma,N)
{
europe = readGDAL("D:\\Emission_fraction\\Europe_gr.asc")
europe$value=europe$band1
europe$band1=NULL
gridcord=read.table("D:\\Emission_fraction\\grid_cord.txt", sep="\t", header=TRUE)
vx= gridcord$x
X=sample(vx,1)
for (i in 1:length(gridcord$x))
if (gridcord$x[i]==X) Y= gridcord$y[i]
EFfert=data.frame(x=X,y=Y,emfa=rnorm(1,mean=mu,sd=sigma))
coordinates(EFfert)=~x+y
sillEF=sigma^2
EF.vgm=vgm(sillEF,"Nug",0)
fEF = gstat(id=c("emfa"),formula=emfa~1,data=EFfert,beta=mu,nmax=24,model=EF.vgm)
EFsim = predict.gstat(fEF,europe,nsim=N,BLUE=FALSE)
return (EFsim)
}
#---------------------------------------------------------------------------
#Fucntion to back transform
backtrnsf <- function(N,EFsim)
{
for (i in 1:N)
{
sim=paste("sim",as.character(i),sep="")
EFsim[[sim]] = exp(EFsim[[sim]])
}
return(EFsim)
```

```
}
#-------------------------------------------------------------------------
#Function to create multi simulation output
simout <- function(N,EFsimulation,dir)
{
for (i in 1:N)
{
colna=c("x","y","ID",paste("sim",as.character(i),sep=""))
EFsim= subset(EFsimulation,select=colna)
write.table(EFsim, file= paste(dir,as.character(i),".txt",sep=""), row.names=FALSE,
quote=FALSE, sep=" ", na="-9999")
}
}
#-------------------------------------------------------------------------
#Function to create EF constant map over Europe
EFmap <-function(mu,sigma,N)
{
EFmap = readGDAL("D:\\Emission_fraction\\Europe_gr.asc")
for (i in 1:N)
{
EFvalue=rlnorm(1,mu,sigma)
value=paste("EFvalue",as.character(i),sep="")
EFmap[[value]]=EFmap$band1*EFvalue
}
EFmap$band1=NULL
return (EFmap)
}
#-------------------------------------------------------------------------
#Function to create multi constant output
conout <- function(N,EFcon,dir)
{
for (i in 1:N)
{
colna=c("x","y","ID",paste("EFvalue",as.character(i),sep=""))
EFconmap= subset(EFcon,select=colna)
write.table(EFconmap, file= paste(dir,as.character(i),".txt",sep=""), row.names=FALSE,
quote=FALSE, sep=" ", na="-9999")
}
}
#set number of simulation
nsim=100
#-------------------------------------------------------------------------
#Emission factor of fertilizer mean=1, p5=0.5, p95=8
p5Yfert=0.5
p95Yfert=8
mYfert=1
#estimate parameter distribution
sigmafert= distparest(mYfert,p5Yfert,p95Yfert)
mufert=log(mYfert)-((sigmafert^2)/2)
windows(width=5, height=5)
par(mar=c(4,4,2,1)+0.9,cex.lab=1.3,cex.axis=1.2, cex.main=1)
fG<-function(x)(pnorm(log(p95Yfert), mean=log(mYfert)-(x^2)/2, sd=x)-pnorm(log(p5Yfert),
mean=log(mYfert)-(x^2)/2, sd=x))
curve(fG,xlim=c(0,1),ylab="G(sigma)",xlab="sigma", main="Fertilizer - N2O Emission
Fraction")
abline(a=NULL, b=NULL, v=NULL, h=0.9)
abline(a=NULL, b=NULL, v=sigmafert, h=NULL)
windows(width=5, height=5)
par(mar=c(4,4,2,1)+0.9,cex.lab=1.3,cex.axis=1.2, cex.main=1)
lognormal=function(x)(1/(sqrt(2*pi)*sigmafert*x)*exp(-((log(x)-
mufert)^2/(2*sigmafert^2))))
curve(lognormal, xlim=c(0,9), main="pdf of Fertilizer N2O emission fraction")
#simulation
EFfertsim=EFsim(mufert,sigmafert,nsim)
#back transformation
EFfertsim <- backtrnsf(nsim,EFfertsim)
#create multi output
EFfertsimulation=as.data.frame(EFfertsim)
ID=c(1:length(EFfertsimulation$sim1))
EFfertsimulation$ID=ID
dir="D:\\Emission_fraction\\EF_fert\\output\\vary\\femis_fert"
simout (nsim,EFfertsimulation,dir)
#EF fertilizer constant map
EFfertmap = EFmap(mufert,sigmafert,nsim)
EFfertcon=as.data.frame(EFfertmap)
ID=c(1:length(EFfertcon$EFvalue1))
EFfertcon$ID=ID
dir="D:\\Emission_fraction\\EF_fert\\output\\constant\\femisfert_con"
conout (nsim,EFfertcon,dir)
#-------------------------------------------------------------------------
```

```
#Emission factor of manure mean=0.75, p5=0.12, p95=0.97
p5Ymanu=0.12
p95Ymanu=0.97
mYmanu=0.75
#estimate parameter distribution
sigmanu= distparest(mYmanu,p5Ymanu,p95Ymanu)
mumanu=log(mYmanu)-((sigmanu^2)/2)
windows(width=5, height=5)
par(mar=c(4,4,2,1)+0.9,cex.lab=1.3,cex.axis=1.2, cex.main=1)
fG<-function(x)(pnorm(log(p95Ymanu), mean=log(mYmanu)-(x^2)/2, sd=x)-pnorm(log(p5Ymanu),
mean=log(mYmanu)-(x^2)/2, sd=x))
curve(fG,xlim=c(0,1),ylab="G(sigma)",xlab="sigma", main="Manure - N2O Emission
Fraction")
abline(a=NULL, b=NULL, v=NULL, h=0.9)
abline(a=NULL, b=NULL, v=sigmanu, h=NULL)
windows(width=5, height=5)
par(mar=c(4,4,2,1)+0.9,cex.lab=1.3,cex.axis=1.2, cex.main=1)
lognormal=function(x)(1/(sqrt(2*pi)*sigmanu*x)*exp(-((log(x)-mumanu)^2/(2*sigmanu^2))))
curve(lognormal, xlim=c(0,3), main="pdf of Manure N2O emission fraction")
#simulation
EFmanusim=EFsim(mumanu,sigmanu,nsim)
#back transformation
EFmanusim <- backtrnsf(nsim,EFmanusim)
#create multi output
EFmanusimulation=as.data.frame(EFmanusim)
ID=c(1:length(EFmanusimulation$sim1))
EFmanusimulation$ID=ID
dir="D:\\Emission_fraction\\EF_manure\\output\\vary\\femis_man"
simout (nsim,EFmanusimulation,dir)
#EF manure constant map
EFmanumap = EFmap(mumanu,sigmanu,nsim)
EFmanucon=as.data.frame(EFmanumap)
ID=c(1:length(EFmanucon$EFvalue1))
EFmanucon$ID=ID
dir="D:\\Emission_fraction\\EF_manure\\output\\constant\\femisman_con"
conout (nsim,EFmanucon,dir)
#-------------------------------------------------------------------------
#Emission factor of soil organic N mean=2.6, p5=1.8, p95=2.9
p5YorgN=1.8
p95YorgN=2.9
mYorgN=2.6
#estimate parameter distribution
sigmaorgN= distparest(mYorgN,p5YorgN,p95YorgN)
muorgN=log(mYorgN)-((sigmaorgN^2)/2)
windows(width=5, height=5)
par(mar=c(4,4,2,1)+0.9,cex.lab=1.3,cex.axis=1.2, cex.main=1)
fG<-function(x)(pnorm(log(p95YorgN), mean=log(mYorgN)-(x^2)/2, sd=x)-pnorm(log(p5YorgN),
mean=log(mYorgN)-(x^2)/2, sd=x))
curve(fG, xlim=c(0,0.6),ylab="G(sigma)",xlab="sigma", main="Soil Organic N - N2O
Emission Fraction")
abline(a=NULL, b=NULL, v=NULL, h=0.9)
abline(a=NULL, b=NULL, v=sigmaorgN, h=NULL)
windows(width=5, height=5)
par(mar=c(4,4,2,1)+0.9,cex.lab=1.3,cex.axis=1.2, cex.main=1)
lognormal=function(x)(1/(sqrt(2*pi)*sigmaorgN*x)*exp(-((log(x)-
muorgN)^2/(2*sigmaorgN^2))))
curve(lognormal, xlim=c(0,9), main="pdf of Soil Organic Carbon N2O emission fraction")
#simulation
EForgNsim=EFsim(muorgN,sigmaorgN,nsim)
#back transformation
EForgNsim <- backtrnsf(nsim,EForgNsim)
#create multi output
EForgNsimulation=as.data.frame(EForgNsim)
ID=c(1:length(EForgNsimulation$sim1))
EForgNsimulation$ID=ID
dir="D:\\Emission_fraction\\EF_orgN\\output\\vary\\femis_orgN"
simout (nsim,EForgNsimulation,dir)
#EF soil N constant map
EForgNmap = EFmap(muorgN,sigmaorgN,nsim)
EForgNcon=as.data.frame(EForgNcon)
ID=c(1:length(EForgNcon$EFvalue1))
EForgNcon$ID=ID
dir="D:\\Emission_fraction\\EF_orgN\\output\\constant\\femisorgN_con"
conout (nsim,EForgNcon,dir)
#-------------------------------------------------------------------------
#Emission factor of crop residues mean=1, p5=0.13, p95=7
p5Ycrop=0.13
p95Ycrop=7
mYcrop=1
#estimate parameter distribution
```

```
sigmacrop= distparest(mYcrop,p5Ycrop,p95Ycrop)
mucrop=log(mYcrop)-((sigmacrop^2)/2)
windows(width=5, height=5)
par(mar=c(4,4,2,1)+0.9,cex.lab=1.3,cex.axis=1.2, cex.main=1)
fG<-function(x)(pnorm(log(p95Ycrop), mean=log(mYcrop)-(x^2)/2, sd=x)-pnorm(log(p5Ycrop),
mean=log(mYcrop)-(x^2)/2, sd=x))
curve(fG,xlim=c(0,2),ylab="G(sigma)",xlab="sigma",main="Crop Residues - N2O Emission
Fraction")
abline(a=NULL, b=NULL, v=NULL, h=0.9)
abline(a=NULL, b=NULL, v=sigmacrop, h=NULL)
windows(width=5, height=5)
par(mar=c(4,4,2,1)+0.9,cex.lab=1.3,cex.axis=1.2, cex.main=1)
lognormal=function(x)(1/(sqrt(2*pi)*sigmacrop*x)*exp(-((log(x)-
mucrop)^2/(2*sigmacrop^2))))
curve(lognormal, xlim=c(0,5), main="pdf of Crop Residues N2O emission fraction")
#simulation
EFcropsim=EFsim(mucrop,sigmacrop,nsim)
#back transformation
EFcropsim <- backtrnsf(nsim,EFcropsim)
#create multi output
EFcropsimulation=as.data.frame(EFcropsim)
ID=c(1:length(EFcropsimulation$sim1))
EFcropsimulation$ID=ID
dir="D:\\Emission_fraction\\EF_cropRe\\output\\vary\\femis_cr"
simout (nsim,EFcropsimulation,dir)
#EF crop residual constant map
EFcropmap = EFmap(mucrop,sigmacrop,nsim)
EFcropcon=as.data.frame(EFcropmap)
ID=c(1:length(EFcropcon$EFvalue1))
EFcropcon$ID=ID
dir="D:\\Emission_fraction\\EF_cropRe\\output\\constant\\femiscrop_con"
conout (nsim,EFcropcon,dir)
```

Quantify Regression Residual Uncertainty

```
#load packages
rm(list = ls()) # clean up working environment
graphics.off() # terminate graphics devices
library(foreign)
library(RColorBrewer)
library(sp)
library(gstat)
library(rgdal)
library(maptools)
library(lattice)
setwd("F:\\Thesis\\Regre_model")
#read data
eurmap = readGDAL("Europe_gr.asc")
coefdata= read.table("regr_nature_1yr_cord.txt", header=TRUE, na.strings = "-
9999",sep="\t")
coefdata=subset(coefdata,select=c(Prec,Frost,pH20,OC20,veg,Ndep,N2O,X,Y))
coefdata$logN2O = log10(coefdata$N2O)
coefdata$logNdep = log10(coefdata$Ndep)
coefdata$fveg=factor(coefdata$veg)
coefdata=na.omit(coefdata)
attach(coefdata)
#fitting regression model
lmN2O = lm(logN2O ~ Prec+Frost+logNdep+pH20+OC20+fveg,data=coefdata )
rres=residuals(lmN2O)
regr_res=as.data.frame(rres)
regr_res$X=coefdata$X
regr_res$Y=coefdata$Y
coordinates(regr_res)=~X+Y
#fitting residual Variogram
fRR = gstat(id="regr_res$rres",formula=regr_res$rres~1,beta=0,data=regr_res,nmax=25)
fRR.vg = variogram(fRR)
fRR.vgm = fit.variogram(fRR.vg,vgm(0.05,"Sph",500000,0.16),fit.sills=TRUE)
fRR = gstat(fRR,id="regr_res$rres",model=fRR.vgm)
# Simulation
N=100
RRsim = predict.gstat(fRR,eurmap,nsim=N,BLUE=FALSE)
RRsim=as.data.frame(RRsim)
ID=c(1:length(RRsim$sim1))
RRsim$ID=ID
#creating output
for (i in 1:N)
{
colna=c("x","y","ID",paste("sim",as.character(i),sep=""))
RRsimu= subset(RRsim,select=colna)
```

```
write.table(RRsimu, file=
paste("For_UA_input\\Res\\RegErr",as.character(i),".txt",sep=""), row.names=FALSE,
quote=FALSE, sep=" ", na="-9999")
}
```

Quantify Regression Coefficients uncertainty

```
#load packages
rm(list = ls()) # clean up working environment
graphics.off() # terminate graphics devices
setwd("F:\\Regre_model")
#read data
coefdata= read.table("regr_nature_1yr_cord.txt", header=TRUE, na.strings = "-
9999",sep="\t")
coefdata=subset(coefdata,select=c(Prec,Frost,pH20,OC20,veg,Ndep,N2O,X,Y))
coefdata$logN2O = log10(coefdata$N2O)
coefdata$logNdep = log10(coefdata$Ndep)
coefdata$fveg=factor(coefdata$veg)
coefdata=na.omit(coefdata)
attach(coefdata)
#fitting regression model
lmN2O = lm(logN2O ~ Prec+Frost+logNdep+pH20+OC20+fveg,data=coefdata )
# sampling from mutivariate distribution
covmax=vcov(lmN2O)
covmax=matrix(covmax[1:7,1:7],nrow = 7,ncol = 7)
Low_trimax=t(matrix(chol(covmax),nrow = 7, ncol = 7))
mu=matrix(coef(lmN2O),nrow = 1, ncol = 9)
mu=matrix(mu[,1:7],nrow = 7, ncol=1 )
N=1000
coefs=matrix(nrow = 7, ncol = 1)
for (i in 1:N)
{
Z=t(matrix(rnorm(7,0,1),nrow = 1, ncol = 7))
X=mu+Low_trimax%*%Z
coefs=cbind(coefs,X)
}
coefs=subset(coefs,select=2:(N+1))
regconst=coefs[1,]
regprec=coefs[2,]
regtemp=coefs[3,]
reglogNdep=coefs[4,]
regph=coefs[5,]
regorg=coefs[6,]
regveg=coefs[7,]
#write the outputs
write.table(regconst,"For_UA_input\\RCs_1000\\regconst.txt",row.names=FALSE,
col.names=FALSE, quote=FALSE, sep=" ", na="-9999")
write.table(regprec,"For_UA_input\\RCs_1000\\regprec.txt",row.names=FALSE,
col.names=FALSE, quote=FALSE, sep=" ", na="-9999")
write.table(regtemp,"For_UA_input\\RCs_1000\\regtemp.txt",row.names=FALSE,
col.names=FALSE, quote=FALSE, sep=" ", na="-9999")
write.table(reglogNdep,"For_UA_input\\RCs_1000\\reglogNdep.txt",row.names=FALSE,
col.names=FALSE, quote=FALSE, sep=" ", na="-9999")
write.table(regph,"For_UA_input\\RCs_1000\\regph.txt",row.names=FALSE, col.names=FALSE,
quote=FALSE, sep=" ", na="-9999")
write.table(regorg,"For_UA_input\\RCs_1000\\regorg.txt",row.names=FALSE,
col.names=FALSE, quote=FALSE, sep=" ", na="-9999")
write.table(regveg,"For_UA_input\\RCs_1000\\regveg.txt",row.names=FALSE,
col.names=FALSE, quote=FALSE, sep=" ", na="-9999")
```

Script for Monte Carlo run in batch mode

```
setwd("F:\\INTEGRATORModel") # set working directory
dyn.load("integratorUA.dll")
is.loaded("setmodeldirectory")
theDir= "F:\\INTEGRATORModel\\Integrator"
ier=0
.Fortran("setmodeldirectory",as.integer(ier),theDir)
is.loaded("initializedll")
ier=0
#initialize
scenario = "A1"
.Fortran("initializedll",as.integer(ier),scenario)
regconstfile=read.table("InputdataUA\\RCs\\regconst.txt",col.names="Value")
regprecfile = read.table("InputdataUA\\RCs\\regprec.txt",col.names="Value")
regtempfile = read.table("InputdataUA\\RCs\\regtemp.txt",col.names="Value")
regphfile = read.table("InputdataUA\\RCs\\regph.txt",col.names="Value")
regorgcfile = read.table("InputdataUA\\RCs\\regorg.txt",col.names="Value")
regndepfile = read.table("InputdataUA\\RCs\\reglogNdep.txt",col.names="Value")
```

```
regvegfile = read.table("InputdataUA\\RCs\\regveg.txt",col.names="Value")
N=100
for (i in 1:N)
{
#regression coefficients
#
RegConst= regconstfile$Value[i]
RegPrec = regprecfile$Value[i]
RegTemp = regtempfile$Value[i]
RegpH   = regphfile$Value[i]
RegOrgC = regorgcfile$Value[i]
RegNdep = regndepfile$Value[i]
RegVeg  = regvegfile$Value[i]

is.loaded("setcurrenttimestep")
#
#reset timestep to year 2000
#
.Fortran("setcurrenttimestep",as.integer(ier),as.integer(2000))
#
# provide pH and Oc data (files):
#
filpH =
       paste("F:\\INTEGRATORModel\\InputdataUA\\ph\\phlusim",as.character(i),".txt",sep="
       ")
is.loaded("providephfile")
.Fortran("providephfile",as.integer(ier),filpH)
filoc =
       paste("F:\\INTEGRATORModel\\InputdataUA\\oc\\oclusim",as.character(i),".txt",sep="
       ")
is.loaded("provideocfile")
.Fortran("provideocfile",as.integer(ier),filoc)
#
# provide files with emission fractions:
#
filfemfert =
       paste("F:\\INTEGRATORModel\\InputdataUA\\femis_fert\\vary\\femis_fert",as.characte
       r(i),".txt",sep="")
is.loaded("providefemfertfile")
.Fortran("providefemfertfile",as.integer(ier),filfemfert)
filfemman =
       paste("F:\\INTEGRATORModel\\InputdataUA\\femis_man\\vary\\femis_man",as.character(
       i),".txt",sep="")
is.loaded("providefemmanfile")
.Fortran("providefemmanfile",as.integer(ier),filfemman)
filfemorgn =
       paste("F:\\INTEGRATORModel\\InputdataUA\\femis_orgN\\vary\\femis_orgN",as.characte
       r(i),".txt",sep="")
is.loaded("providefemorgnfile")
.Fortran("providefemorgnfile",as.integer(ier),filfemorgn)
filfemcr =
       paste("F:\\INTEGRATORModel\\InputdataUA\\femis_cr\\vary\\femis_cr",as.character(i)
       ,".txt",sep="")
is.loaded("providefemcrfile")
.Fortran("providefemcrfile",as.integer(ier),filfemcr)
#
# regression coefficients for N2O from natural areas:
#
is.loaded("provideregpar")
.Fortran("provideregpar",as.integer(ier),as.double(RegConst),as.double(RegPrec),as.doubl
       e(RegTemp),as.double(RegpH),as.double(RegOrgC),as.double(RegNdep),as.double(RegVeg
       ))
#
# regression error term (spatially explicit)
#
filregerr =
       paste("F:\\INTEGRATORModel\\InputdataUA\\Res\\RegErr",as.character(i),".txt",sep="
       ")
is.loaded("provideregerrfile")
.Fortran("provideregerrfile",as.integer(ier),filregerr)
#
#run the model:
#
is.loaded("run")
.Fortran("run",as.integer(ier))
#
# get the results and write them to an output file:
#
is.loaded("getvalues")
```

```
resultfile =
       paste("F:\\INTEGRATORModel\\UAOutput\\scenA1\\MIX\\out_mix",as.character(i),".out"
       ,sep="")
.Fortran("getvalues",as.integer(ier),resultfile)
}
```

Analyze spatial aggregation effects on outputs unceratinty

```
#load package
library(foreign)
library(sp)
library(rgdal)
library(gstat)
library(lattice)
library(maptools)
library(RColorBrewer)
memory.limit(size=4095)
rm(list = ls())
graphics.off()
#set working directory
setwd("H:\\phuong\\upscale")
#function to calculate row variance
rowVars <- function(x, na.rm=FALSE, dims=1, unbiased=TRUE, SumSquares=FALSE,
                    twopass=FALSE) {
  if (SumSquares) return(rowSums(x^2, na.rm, dims))
  N <- rowSums(!is.na(x), FALSE, dims)
  Nm1 <- if (unbiased) N-1 else N
  if (twopass) {x <- if (dims==0) x - mean(x, na.rm=na.rm) else
                     sweep(x, 1:dims, rowMeans(x,na.rm,dims))}
  (rowSums(x^2, na.rm, dims) - rowSums(x, na.rm, dims)^2/N) / Nm1
}
eu.poly = readShapePoly("world_admin_esri_EUcountries_15m_etrs3035.shp")
ctryname=c(as.character(eu.poly$CNTRY_NAME[1:19]),as.character(eu.poly$CNTRY_NAME[22:23]
       ),as.character(eu.poly$CNTRY_NAME[26:27]),as.character(eu.poly$CNTRY_NAME[29:32]),
       as.character(eu.poly$CNTRY_NAME[34]),as.character(eu.poly$CNTRY_NAME[37]))

#compute variance for each country: varctr
N2OMIXsim= read.table("N2O800mix.txt",header=TRUE)
gridded(N2OMIXsim)<-~x1+y1
N2OMIXsim$country_code=overlay(N2OMIXsim,eu.poly)
N2OMIXaverage = aggregate(N2OMIXsim, na.rm = TRUE, by=list(N2OMIXsim$country_code), FUN
       = mean)
N2OMIXaverage$var=rowVars(subset(N2OMIXaverage,select=c(sim1:sim800)))
N2OMIXaverage$ctrname=ctry

#compute varieance for each location in country:varpts
varpoint=read.table("N2Ovar_st800mix.txt",header=TRUE)
gridded(varpoint)<-~x+y
varpoint$country_code = overlay(varpoint, eu.poly)
varp_average = aggregate(varpoint, na.rm = TRUE, by=list(varpoint$country_code), FUN =
       mean)
varp_average$ctryname=ctryname

#visualizing results
mean_varpts<-tapply(varpoint$var,varpoint$country_code,mean)
windows(width=9, height=7)
par(mar=c(4,4,1,1)+0.9,cex.lab=0.9,cex.axis=0.6)
bplot<-boxplot(varpoint$var~varpoint$country_code,ylim=c(
       1,20),na.rm=TRUE,col="white",boxwex = 0.5,names=c(1:29),outline=FALSE,
       xlab="Country number",ylab="N2O estimate variance")
xi <- seq(bplot$n)
points(xi, mean_varpts, col = "black", pch = 8)
points(xi, N2OMIXaverage$var, col = "black", pch = 18)
legend(x="topright", legend=c("Mean Varpts","Mean Varctr"), col = c("black","black"),
       pch = c(8, 18),cex=0.8)
savePlot("plotscaleeffect",type="png")
```