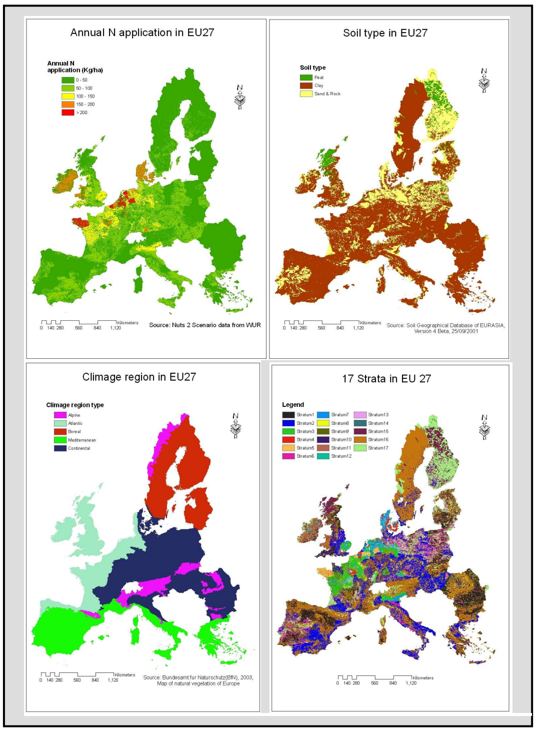
Centre for Geo-Information

Thesis Report GIRS-2009-08

Accuracy of Mean Annual N2O Emission Estimates from Agricultural and Natural Area in the EU27 using Stratified Random Sampling

Haolu Shang

April, 2009



WAGENINGEN UNIVERSITY
WAGENINGEN UR

Accuracy of Mean Annual N_2O Emission Estimates from Agricultural and Natural Area in the EU 27 using Stratified Random Sampling

Haolu Shang

Registration number 85 03 11 759 050

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: 25 April, 2009
Wageningen, The Netherlands

Thesis code number: GRS-80436
Thesis Report: GIRS-2009-08
Wageningen University and Research Centre
Laboratory of Geo-Information Science and Remote Sensing

Acknowledgements

During 8 moths, I have worked with my supervisors Gerard Heuvelink and Sytze de Bruin on the accuracy of mean annual N2O emission from agriculture and natural area in the EU 27 using stratified random sampling. This topic is a little bit difficult for me just because of my background is not so ecological. Thanks to the help of Gerard Vetlhof and Hans Kros, I can handle this problem. I am glad that my supervisors always give me the pressure to finish every thing by schedule. It is a human nature that people will loose their patient when facing the troubles and difficulty. Maybe the pressure is not a gift for me, but I use it to raise my motivation. To be honest, I also want to schedule my life well and use every minutes of the time. It seems a long way before I can handle it. However, the start point is right now.

At the beginning, the feeling that I was alone when I was doing my thesis is so strong that I did not want to ask any help from my classmates or my supervisors. Some people said that the Dutch people were as cold as the weather. I have to correct them. Dutch people are the hot point during the cold winter. Our thesis room is always warm as a family.

This is an amazing travel when I close the end of my working. My memory and schedule seems still at the stage of thesis. I learned so much from peoples around me, from literatures and from the complex and terrible N_2O emission topic. If you asked me what I would choose when I have a chance to start again, I will say, the same topic, the same supervisors, but working much faster.

Thanks to all the experts who help me to finish the interviewing. I am grateful to have the chance to discuss with them. Finally, a long travel has to be finished. I am waiting for another start.

Haolu Shang

Wageningen, 2009

Summary

Nitrous-oxide is one of the greenhouse gases, which attracts a lot attention in 21st century. There is much research on the N_2O emission for agricultural and natural area at the field scale. Based on the research at the field scale, some model-based methods, such as IPCC methods, are developed to estimate the annual N_2O emission at the continental scale. However, the uncertainty of these model-based methods is large. In this research, we evaluate the accuracy of estimation of mean annual N_2O emission from agricultural and natural area using stratified random sampling method. The study area is 27 countries in European Union (EU27). For applying stratified random sampling methods to EU27, the spatial factors controlling annual N_2O emission are derived from literature studying. After obtaining corresponding spatial data, they are processed to combine together. The compositions of combination are hierarchically clustered to create the strata, which are designed for the annual N_2O emission. Due to lack of field measurement data, we make a questionnaire on the mean annual N_2O emission for each stratum to consult experts. From the answers of experts, we calculate the within-stratum variance. Given a certain sample size, the total variance could be calculated from the within-stratum variance and the characteristic of each stratum. The main results of this research are: (1) there are four spatial factors controlling N_2O emission at the continental scale, which are land use type, annual N input, soil type and climate region; (2) We designed 17 strata over EU27 according to the combination of the four corresponding spatial data by agglomerative hierarchical clustering; (3) The optimistic scenario and pessimistic scenario of within-stratum variance are calculated from the answers of experts; (4) Given the total sample size of 200, the total variance calculated from optimistic scenario and pessimistic scenario is 0.98 kg N/ha and 3.41 kg N/ha, of which the range is much smaller than model-based methods.

This report has five parts. The first chapter gives an overview of this research and raises the objective and research questions. Chapter 2 describes the methodology used in this research, which has 5 sections. Each section tells the story of the methodology, which is used to answer the research question one by one. Chapter 3 shows the result of each methodology, except the statistical theory described in section 2.1. In the chapter 4, we discuss the methodologies and corresponding results we have. Finally, chapter 5 concludes the complete research process and gives the answers of the objective and the research questions.

1. Introduction

1.1. Background and problem definition

With the implementation of international agreements such as the Kyoto protocol and the general public awareness of climate change, national authorities and international bodies try to assess the national and continental emission of greenhouse gases. In this research, the mean annual Nitrous-oxide emission from agricultural and natural area of the 27 member states of the European Union (EU27) is studied.

Nitrous-oxide is one of the greenhouse gases and plays an important role in the global warming. Because of the complex biogeochemical processes and the structural simplifications involved in modelling, the estimation of N_2O emission by model-based methods, such as the IPCC methods, and MITERRA-EUROPE (Velthof et al. 2007), has large associated uncertainty. In this research, stratified random sampling is applied to estimate the mean annual N_2O emission from agricultural and natural land using a design-based approach. The purpose of this research is to evaluate the accuracy of this method.

N_2O is a product or inter-media of microbial transformations (nitrification and de-nitrification) of nitrogenous compounds. The dominant sink of nitrous-oxide is the reaction with atomic oxygen in the stratosphere to produce NO (Wrage et al. 2001). This process potentially increases the UV radiation at the earth surface and causes the destruction of stratospheric ozone (Mooney et al., 1987; Prather et al., 1995). Tropical forest soil appears to be the major natural source of nitrous oxide. Fertilized agricultural ecosystems also emit more nitrous oxide than do most natural ecosystems (Mooney et al., 1987). The increase of N_2O concentration in atmosphere is attributed to increasing anthropogenic emissions, through increasing production and use of N fertilizers, tropical land conversion from forest to agriculture, increased biomass burning, etc. (Prather et al., 1995).

Because of the complexity of the biogeochemical process of N_2O production, the structural uncertainty of modelling based methods is difficult to preclude. Models designed for a specific purpose make simplifying assumptions to represent ecosystem processes. Structural uncertainty in these models is relevant given that there are alternative hypotheses regarding which processes are the key to influence terrestrial N_2O emission, or which functional forms are correct for a given process. The processes of greenhouse gases production will be influenced by the multiple environmental resources and stresses (Schimel et al., 1997). It is difficult to reduce the structural uncertainty in the model and even to evaluate it. To take environmental factors into account, model-based estimation needs a lot of parameters, for example, temperature, soil carbon content, soil moisture and precipitation. This increases the work of collecting and processing data. Some information, such as soil temperature, may be very difficult to be measured at the national and even continental scale.

The model-based approach has large uncertainty. A stratified random sampling method is proposed in this research to measure mean annual N_2O emission from samples.

There are several basic design types in design-based sampling, such as simple random sampling, stratified random sampling and clustered random sampling. To increase the accuracy and reduce the sample size, we use stratified random sampling in this research. Stratified random sampling is an appropriate method to estimate population characteristic from a sample, especially when the population has a spatial relation with some known spatially distributed property (Olea, 1984;Matern, 1986). The basic idea of this method is that we stratify the whole study area according to control factors of N_2O emission, so that the variability of N_2O emission between strata is larger than that within each stratum. From an ecosystem perspective, the circumstances within stratum are more consistent and similar than between strata. The sample points of each stratum are randomly chosen within each stratum.

The N_2O emission is measured by chambers continuously at each sample location through the whole year. Given the total sample size, the variance of estimation depends on the strategy of sample method, also called sampling design. To get the optimal variance, the three attributes of the sampling design -- definition of strata, the total sample size, and the allocation of the sample size to each stratum should be decided.

This research is based on the hypothesis that the mean annual terrestrial greenhouse gases emission has spatial variability. This spatial variability is represented in the definition of the strata and controlled by spatial factors which influence the annual N_2O emission from terrestrial ecosystems. So finding these control factors of N_2O emission is one of the major research questions.

Nowadays, much research on N_2O emission is done for different ecosystems, such as, pasture, cropland, and forest. There are some important spatial factors controlling the N_2O emission at the field-scale. However, the spatial factors at the field-scale can not directly be used as the spatial factors at the continental scale, because of the large spatial variability of these factors. In this research, annual N_2O emission is measured, while the spatial factors at the field scale are influencing the daily or monthly N_2O emission. The temporal variability of these spatial factors in one year is also very large. All the corresponding spatial data of the spatial factors need to be accessible from internet. The difficulty of this research is how to determine the spatial factors, which are control the annual N_2O emission at the continental scale.

After obtaining all spatial data, we use them to create the strata designed to estimate the annual N_2O emission from agricultural and natural area in EU 27. The combination of all spatial data is processed to create the strata. Dissimilarity table of the result of combination is created according to the dissimilarity tables of spatial factors. We use software to cluster the result of combination to obtain the strata needed in this research.

Due to the lack of field measurement data over Europe, we consult experts to guess the annual N_2O emission for each stratum we designed. A questionnaire is designed for this research. The answers from experts are processed and analyzed to obtain the data we need in this research.

According to the data analyzed from answers of experts and the statistical theory of stratified random sampling, we calculate the total variance at a certain total sample size. The total variance is evaluated with the increase total sample size. The trend of total variance is also evaluated in this research.

1.2. Objective and research question

The aim of this research is to asses the accuracy of design-based estimation of mean annual nitrous-oxide emission by natural and agricultural land in EU 27 by applying stratified random sampling.

The European Union currently has 27 countries and most of the datasets at European Union scale coves 27 countries, therefore we set the study area as EU 27.

This research is expected to test the accuracy of estimator with certain classes of sample size. Due to the lack of measurement data over all the 27 countries, we use the expert's knowledge to get the variance of mean annual N_2O emission within each stratum. Accordingly, the effectiveness of sampling can be evaluated even without performing the measurements.

To implement the stratified random sampling method to estimation of the mean annual nitrous-oxide emission in EU 27, there are five questions that this research must answer:

1. What are the theory and the practice of stratified random sampling to estimate nitrous-oxide gas emission in EU 27?
2. What are important spatial factors controlling annual nitrous-oxide gas emissions from natural and agricultural land?
3. How to process the spatial data to create strata, which are designed for N_2O emission?
4. What are the annual within-stratum variances for nitrous-oxide by consulting expert?
5. What is the total variance of the estimation given a certain total sample size?

2. Methods

2.1. Theory of stratified random sampling

The stratified random sampling method is described in many statistics books (Barnett,1991; Hansen et al., 1993; Thompson, 2002; Levy and Lemeshow, 2008;). The total sample size n is a pre-determined size, as the sum of the stratum sample size. The stratum sample sizes will be denoted $n_1, n_2, \ldots, n_k$ ($\sum_{i=1,2\ldots,k} n_i = n$). The simple random sample from the *i*th stratum has members $z_{i1}, z_{i2}, \ldots, z_{ij}$ ($j = 1, 2, \ldots, n_i$). For every stratum, the sample points are selected using the same rule as the simple random sample. This means that in each stratum, sample locations are selected independently with equal probability. The measured annual N2O emission from these sample points in each stratum are denoted as $y_{i1}, y_{i2}, \ldots, y_{ij}$ ($j = 1, 2, \ldots, n_i$)

We denote the sample mean of *i*th stratum as $\bar{y}_i$, which can be calculated as:

$$\bar{y}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij} \qquad (2.1)$$

And standard deviation of the sample mean $\bar{y}_i$ is denoted as s_i, which can be calculated as:

$$s_i^2 = \frac{1}{(n_i - 1)}\sum_{j=1}^{n_i}\left(y_{ij} - \bar{y}_i\right)^2 \qquad (2.2)$$

If n_i is large, s_i is used to represent the within stratum variance of *i*th stratum.

And the total mean of emission by stratified random sampling method, donated as $\bar{y}_{st}$, can be estimated from the sample mean of each stratum:

$$\bar{y}_{st} = \sum_{i=1}^{k} w_i \times \bar{y}_i \qquad (2.3)$$

$$w_i = \frac{Area_i}{Area_t} \qquad (2.4)$$

$Area_i$ is the area of *i*th stratum; $Area_t$ is the total area of Europe. We call w_i as the stratum weight.

The total variance of total mean $\overline{y}_{st}$ is donated as Var ($\overline{y}_{st}$), which can be calculated from the stratum weight and the standard deviation of each stratum:

$$Var\left(\overline{y}_{st}\right) = \sum_{i=1}^{k} \frac{w_i^2 \times s_i^2}{n_i} \qquad (2.5)$$

The accuracy of this method is the total variance of $\overline{y}_{st}$. According to Eq. 2.5, there are three elements calculating the Var ($\overline{y}_{st}$) – w_i, s_i, and n_i. w_i is determined by the area of *i*th stratum. s_i is controlled by the characteristic of N2O emission from *i*th stratum. For a certain total sample size n, the optimal allocation of n_i is determined by w_i and s_i:

$$n_i = \frac{w_i^2 \times s_i^2}{\sum_{i=1}^{k} w_i^2 \times s_i^2} \times n \qquad (2.6)$$

The Eq. 2.6 can be proved as below:

The optimal allocation of the sample size to each stratum means that the total variance needs to be smallest. According to Eq. 2.5, there is:

$$Var\left(\overline{y}_{st}\right) = \sum_{i=1}^{k} \frac{w_i^2 \times s_i^2}{n_i} \geq C \qquad (2.7)$$

C is a certain number which is the optimal total variance we want to obtain.

For

$$Var\left(\overline{y}_{st}\right) = \sum_{i=1}^{k} \frac{w_i^2 \times s_i^2}{n_i} = \sum_{i=1}^{k} \frac{1}{\left(\frac{n_i}{w_i^2 \times s_i^2}\right)} \qquad (2.8)$$

We set

$$x_i = \frac{n_i}{w_i^2 \times s_i^2} \qquad (2.9)$$

According to Handy et al. [1999], we get the inequality equation: $H \leq G \leq A$, for any positive numbers x_1, x_2, …, x_n, where

$$H = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} \qquad \text{(Harmonic mean)}$$

$$G = \left(\prod_{i=1}^{n} x_i \right)^{\frac{1}{n}} \qquad \text{(Geometric mean)}$$

$$A = \frac{\sum_{i=1}^{n} x_i}{n} \qquad \text{(Arithmetic mean)}$$

The equation holds if and only if $x_1 = x_2 = ... = x_n$.

Because of the transitivity of inequality equations, we simplify the above inequality equation as $H \le A$:

$$\frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} \le \frac{\sum_{i=1}^{n} x_i}{n} \qquad (2.10)$$

The equation holds if and only if $x_1 = x_2 = ... = x_n$.

Because all x_i are positive, we get:

$$\frac{n^2}{\sum_{i=1}^{n} x_i} \le \sum_{i=1}^{n} \frac{1}{x_i} \qquad (2.11)$$

The equation holds if and only if $x_1 = x_2 = ... = x_n$.

Replacing the x_i according to Eq. 2.9, Eq. 2.11 is changed as:

$$\frac{k^2}{\sum_{i=1}^{k} \frac{w_i^2 \times s_i^2}{n_i}} \le \sum_{i=1}^{k} \frac{1}{\left(\frac{n_i}{w_i^2 \times s_i^2} \right)} = Var(\bar{y}_{st}) \qquad (2.12)$$

The equation holds if and only if

$$\frac{n_1}{w_1^2 \times s_1^2} = \frac{n_2}{w_2^2 \times s_2^2} = ... = \frac{n_k}{w_k^2 \times s_k^2} = \frac{\sum_{i=1}^{k} n_i}{\sum_{i=1}^{k} w_i^2 \times s_i^2} .$$

Because $\sum_{i=1}^{k} n_i = n$

$$\frac{n_1}{w_1^2 \times s_1^2} = \frac{n_2}{w_2^2 \times s_2^2} = \ldots = \frac{n_k}{w_k^2 \times s_k^2} = \frac{n}{\sum_{i=1}^{k} w_i^2 \times s_i^2}$$

So Eq. 2.6 is proved.

To reduce the uncertainty with a limited sample size, sampling strategy plays an important role. Sampling design refers to the distribution of sampling locations over the study area. It has three attributes -- definition of strata, the total sample size, and the allocation of the sample size to each stratum.

2.2. Spatial factors controlling annual N2O emission

The most difficult part of sampling design in this research is the definition of strata. Spatial factors, which control the annual N_2O emission, are the theoretical basis of the definition of strata for 27 countries in European Union (EU27).

N_2O is produced through certain biogeochemical processes by microbes. Literature studying is an appropriate way to obtain an overview of the N_2O emission from agricultural and natural area. However, until now, there is no much research on the spatial factors influencing N_2O emission at the continental scale.

Recent research is focused on influencing factors at field scale, which control daily or seasonal N_2O emission at field scale. Some of the influencing factors do not have spatial and temporal difference between field scale and continental scale, which can be directly used in this research. By up scaling from daily to annual, some of the influencing factors at the field scale can be used as the spatial factor at the continental scale. Due to the large spatial and temporal variability, some of the influencing factors need to be up scaled with the spatial factors that can represent the annual characteristics of the phenomenon at continental scale.

The biogeochemical processes of N_2O production are the theory basic of field research on N_2O. No matter the spatial and temporal differences, the factors control N_2O emission by influencing the N_2O biogeochemical processes. The only difference between the factor at field scale and those at continental scale is the spatial or temporal level of the influencing factors in biogeochemical processes. For example, we use air temperature and rain full to describe the weather, while the annual average temperature and precipitation are used to describe the climate. Climate and weather describe the same influencing factors in biogeochemical processes, which are the temperature and the water content. This approach is the basic idea of obtaining spatial factors from literature studying.

The availability of spatial data for each corresponding spatial factor is an important criterion for up-scaling the influencing factors of N_2O emission at the field scale. The spatial data used in this research need high quality, in order to reduce the uncertainty of the estimation about them. Here are some criteria for choosing spatial data:

1. Accessibility. It means that the spatial data should be easy to access for other researchers through internet.
2. Reliability. The producer and publisher of the spatial data should be reliable. It is preferred that the spatial data is published by authorities in specific field.
3. Accuracy. It is difficult to obtain spatial data at continental scale with high spatial accuracy. We assume that the spatial data published by domain authorities has acceptable accuracy.
4. Completeness. The spatial data should at least cover the EU27 countries.
5. Temporal accuracy. To reduce the temporal uncertainty, we try to use the spatial data published in the same year.

2.3. Process spatial data to create strata

The strata, which are defined to measure annual N_2O emission samples in the EU 27 countries, are derived from the combination of spatial data, which represent the spatial factors controlling annual N_2O emission.

2.3.1. Prepare the spatial data

To prepare the data, several processing steps were required.

Reclassify: The spatial data are always plural, with their own attribute table or legend. The employed classes of their attributes or legends are not appropriate for N_2O emission research. The classification of each spatial factor controlling N_2O emission is not as exhaustive as the classification of corresponding spatial data. Because there is small quantity of N_2O emission from the earth, the difference of N_2O emission from the employed classes of spatial factors may be not significant. The specific classes of the spatial data are always too detailed to tell the different influence on N_2O emission, due to the different classification purpose. So reclassification is implemented to categorize the classes of each spatial factor to make them suitable for N_2O emission.

Projection: Because spatial data are published by different organizations, there is no unified reference system for all spatial data. All the spatial data were projected to the reference system – ETRS – Lambert Azimutal Equal Area, which is a standard reference system of spatial data published in European Environment Agency (EEA).

Conversion: All spatial data were converted to raster format, so that they could be combined together. The pixel size is determined as 250 meters.

Combination: After above processes, all spatial data are combined together.

The detailed flow chart of preparing spatial data can been found in Appendix A.

2.3.2. Dissimilarity tables

After preparing spatial data, they are all combined together to make the strata, which are designed for researching N_2O emission. The strata are created from the features of combination. After combination, there will be many features, which have large variability in their area. According to Eq. 2.4, the stratum weight w_i is determined by the stratum area. According to Eq. 2.6, if the stratum weight is too small and the total sample size is not large enough, it may happen that the stratum sample size n_i will be zero. To avoid this situation, all features of combination are clustered according their dissimilarity table, which we call the dissimilarity table of features.

Because each feature of combination is determined by the composition of the classes of spatial factors, we can calculate the dissimilarity table of features from the dissimilarity tables of all spatial factors.

We set the dissimilarity table of each spatial factor by our understanding of the different influencing powers of the N_2O emission among the classes, which this spatial factor has. We denote the dissimilarity tables of spatial factors as:
SF_1 [], SF_2 [], …, SF_n [].

For the dissimilarity table of *i*th spatial factor, each row and column represents one class in this spatial factor, which is denoted as SF_i [j]. "j" is the row number or the column number. Each value of the dissimilarity is denoted as SF_i [p, q]. It shows how large the difference between two classes, one of which is represented by the row number "p" and the other is by the column number "q". The value of dissimilarity table is from 0 – 1. 1 means that the difference between classes is the largest, while 0 means that there is no difference.

Similar to the dissimilarity table of a spatial factor, each row and column represents one class in the dissimilarity table of features. The value of the dissimilarity table is denoted as A [p, q]. "p" represents the *p*th feature in row and "q" represents *q*th feature in column. In the attribute table of combination, we can find that which classes of spatial factors the feature has. So for *p*th feature, we can find the list of composition of classes of its spatial factors as:
SF_1 [p_1], SF_2 [p_2], …, SF_n [p_n].
For *q*th feature, we can also have the list of composition of classes of its spatial factors as:
SF_1 [q_1], SF_2 [q_2], …, SF_n [q_n].
From the compositions of classes of spatial factors, which *p*th and *q*th features have, we can find the corresponding dissimilarity value in the dissimilarity table of each spatial factor, listed as below:
SF_1 [p_1, q_1], SF_2 [p_2, q_2], …, SF_n [p_n, q_n].

A [p, q] is from 0 to 1. 1 means that the difference between two features is the largest, while 0 means that there is no difference. A [p, q] is calculated from the weighted sum of dissimilarity values of spatial factors as:

$$A[p,q] = \sum_{i=1}^{n} m_i \times SF[p_i, q_i] \tag{2.13}$$

Where m_i is the weight of *i*th spatial factor, with

$$\sum_{i=1}^{n} m_i = 1 \tag{2.14}$$

We set the weight m_i to the spatial factors according to our understanding of the different influencing power of spatial factors. Because A [p, q] and SF [p, q] both range from 0 to 1, the sum of m_i is equal to 1.

2.3.3. Cluster and create strata

All features of combination were clustered by agglomerative hierarchical clustering based on the dissimilarity table of features. This process is implemented by the function of Agglomerative Nesting (agnes), which is part of the cluster library for R (http://stat.ethz.ch/R-manual/R-devel/library/cluster/html/agnes.html). We employed a Euclidean dissimilarity measure and average cluster distances. Average means that the distance between two clusters is the average of the dissimilarities between the points in one cluster and the points in the other cluster.

For *i*th feature in dissimilarity table, “agnes” uses V (i) to denote it. The function “agnes” calculates the agglomerative coefficient of V (i). It is denoted as m (i), which is the dissimilarity to the first cluster it is merged with, divided by the dissimilarity of the merger in the final step of the algorithm (Kaufman and Rousseeuw, 1990). The agglomerative coefficient of this clustering is denoted as “AC”, which is calculated as:

$$AC = \frac{\sum_{i=1}^{n}(1 - m(i))}{n} \tag{2.15}$$

All features of combination are reclassified according to the result of agglomerative hierarchical clustering. There is some theory to determine the cluster number (Catherine and Gareth, 2003; Lieti et al., 2004). Due to the lack of time, we set the range of the number of cluster between 10 and 20. The number of cluster is determined by evaluate the result of clustering in ArcGIS. In ArcGIS, we denote *i*th cluster as stratum *i*. According to the clustering of each feature set, we reclassify the features into strata in ArcGIS. The compositions of features in each stratum should have their meaning for N_2O emission.

2.4. Annual within stratum variance

Due to lack of field measurement data of the strata designed for N_2O emission, experts were consulted to estimate the distribution of annual N_2O emission in each stratum according to their knowledge and experience. A questionnaire was designed and distributed among 5 experts at Alterra. Finally, the results of the questionnaire were processed.

2.4.1. Questionnaire

Overall, the questionnaire was structured in two parts. The first part is the introduction of this research. It contained a description of this research and figures of 4 spatial factors and the final strata, which we created. The second part concerned the information and questionnaire of each stratum.

For each stratum, some information about its composition and spatial distribution was given, which has three parts. The first part was attribute table of this stratum. Attribute table contained all compositions, which this stratum has. Pie chart showed the percentage of the area, which the classes of each spatial factor take up in this stratum. For some stratum, there might be no difference of a spatial factor in the composition of this stratum, so there would not always be pie charts for all spatial factors. One European map is used to represent the location of this stratum.

A questionnaire table is presenting after the information of each stratum. The experts were asked to fill the best estimation of annual N_2O in the table. Because a stratum has different compositions and N_2O emission shows large spatial variability, the annual N_2O emission will not be the same in all locations. The distribution of v emission could be lognormal, normal or something else. In figure 2.1, we give two examples of the distribution of annual N_2O emission value. We let the experts to judge what kind of distribution of annual N_2O emission would be.

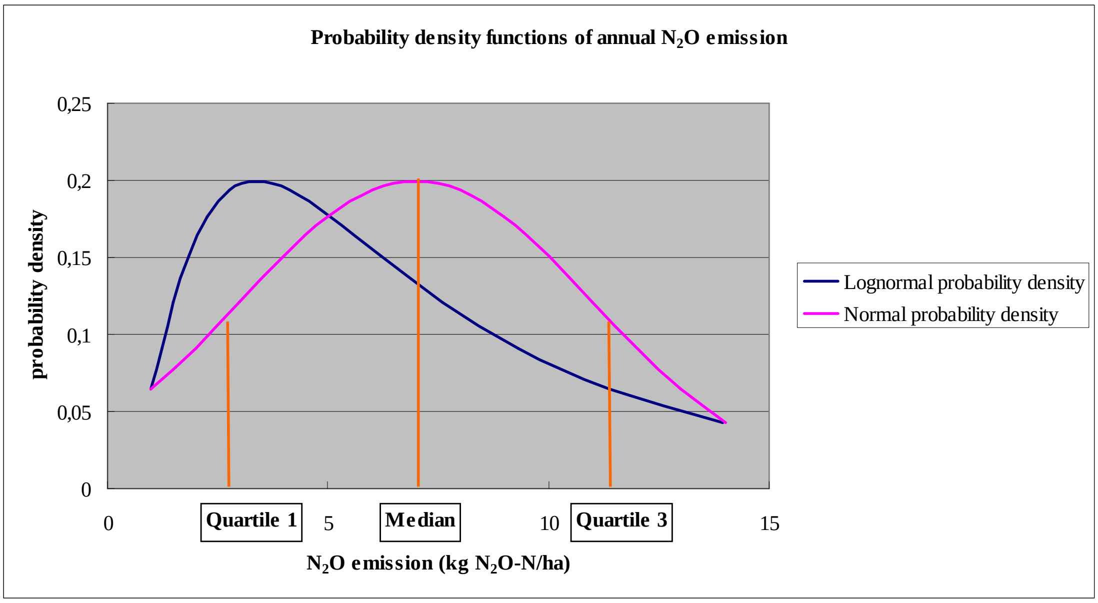


Figure 2.1, Example of probability density functions of annual N2O emission.

The questionnaire table presents the cumulative frequency table of annual N_2O emission cut off by quartiles. The first quartile corresponds to be that value for which the expert believes that there is 25% probability that the actual annual N_2O emission will be smaller, seeing the location of quartile 1 in figure 2.1. The third quartile corresponds to be that value for which there is 25% probability that the actual annual N_2O emission will be greater, see figure 2.1. The median (= 2^{nd} quartile) denotes the value for which there is 50% probability that the actual annual N_2O emission will be greater, see figure 2.1. The Quartile 1, Median and Quartile 3 only refer to the normal distribution in figure 2.1.

The content of the questionnaire is shown in Appendix B.

2.4.2. Estimate within stratum variance from questionnaire

For each stratum, experts have different judgments. It is difficult to compare how confident the experts feel about their answers. To simplify the analysis, the maximum distance and mean distance between the first quartile and the second quartile were calculated from answers for each stratum. These two kinds of distances are defined as below:

$$\text{Max distance} = \max(Q3) - \min(Q1) \quad (2.16)$$

$$\text{Mean distance} = \text{mean}(Q3) - \text{mean}(Q1) \quad (2.17)$$

Q3 is the value of the third quartile in the answers of questionnaire

Q1 is the value of the first quartile in the answers of questionnaire

Even though, the field measured N_2O emission can be expected to follow a lognormal distribution, for the sake of simplicity, in this research, we assume that the N_2O emissions are normally distributed. In normal distribution, the difference between the first quartile and the third quartile is the 1.34 times of the standard deviation. If the stratum sample size is large enough, we use the standard deviation of the sample mean, s_i, to represent the within stratum variance, which can be calculated from the equation below:

$$s_i = \text{distance}/1.34 \qquad (2.18)$$

Distance is the max distance or the mean distance calculated from Eq.2.16 or Eq.2.17.

Because there are two kinds of distances, we have two scenarios of the estimation of s_i. We call the scenario of the estimation of s_i, which is calculated from max distance as the pessimistic scenario. And the scenario of the estimation of s_i, which is calculated from mean distance as the optimistic scenario.

2.5. Variance of N_2O emission at a sample size

2.5.1. Calculate the variance of annual N_2O emission

Using Eq.2.18, the standard deviation s_i of each stratum can be calculated from the returned questionnaires. According to Eq.2.4, the stratum weight w_i was calculated. At a certain total sample size, after obtaining the standard deviation and the stratum weight, the optimal sample size of each stratum was calculated from Eq.2.6. However, for some strata, which have small weight and small standard deviation, the sample size for them could be zero. This was seen as unacceptable in this research.

To deal with this problem, at first, strata were forced to have at least sample size 2, as follows:
$n_i = n_{adjusted} = \min(n_{Eq.2.6}, 2)$

At a certain total sample size, after obtain s_i, w_i, n_i, for each stratum, the total variance was calculated according to Eq.2.5.

2.5.2. Trend of the variance curve

We have a set of the total sample size N.

$N = \{N_1, N_2, \ldots, N_k\}$, $k = 1, 2, \ldots, n$.

Using Eq.2.5 we can calculate the total variance V_i at a certain total sample size $N_i \in N$. A curve can be plotted with the X axis of N and Y axis of the corresponding total variance. We standardize the curves by calculating the decreasing rate at each sample size. For each set of total variance, it is denoted as R_i, and calculated as below:

$$R_i = -\frac{V_i - V_{i-1}}{V_1 \times (N_i - N_{i-1})} \times 100\% \qquad (2.21)$$

$N_i \in N$.

V_i is the total variance at N_i.

V_1 is the total variance at N_1.

3. Results

3.1. Results of Spatial factors

3.1.1. Biogeochemical process of N_2O production

N_2O is a product or intermedia of microbial transformations (nitrification and denitrification) of nitrogenous compounds. The same microbial processes of N_2O production could take place whether in soil, wastewater treatment plant, sediments or water bodies (Wrage et al., 2001).

Circumstances and specific microorganisms are the two factors controlling the pathway of N_2O production. Nitrification is the oxidation of NH_4^+ or NH_3 to NO_3^- via NO_2^- by autotrophic or heterotrophic nitrifiers, which is shown in figure 3.1. N_2O is formed during NH_3 oxidation through chemical decomposition of intermediates between NH_4^+ and NO_2^-. Different nitrifiers have their own preferable environment. Autotrophic nitrifiers, also called as Nitrobacteriaceae (Buchanan, 1917), are aerobes and many are obligate autographs. Fungi as a common heterotrophic nitrifer play an important role in the nitrification in soils with a low pH. No mater what kind of nitrifiers, they can be very important in terms of N transformations, even under circumstances where their population is not large (Wrage et al., 2001).

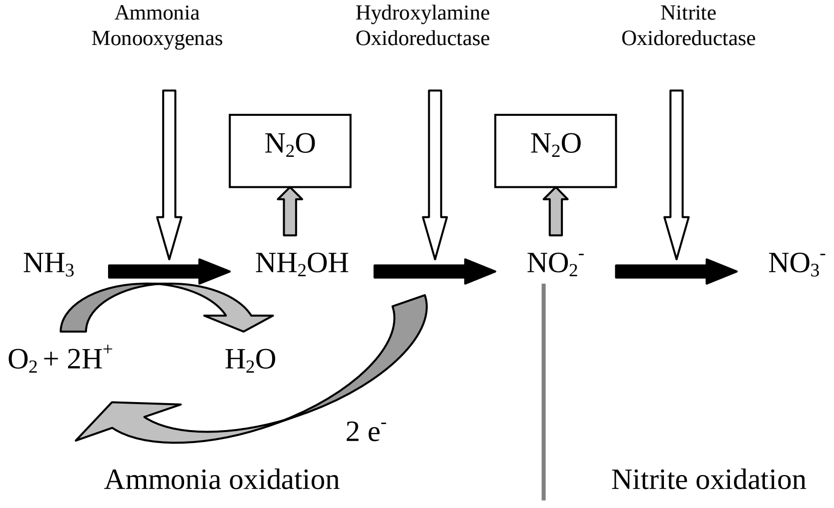


Figure 3.1, Nitrification: Outline of the pathway and enzyms involved (after Hynes and Knowles, 1984; Poth and Focht, 1985; Wood, 1986)

Denitrification is the stepwise reduction of NO_3^- to N_2, which is shown in figure 3.2. In contrast to nitrification, N_2O is a regular intermediate of denitrification, which can be released in high quantities in low-oxygen environments with sufficient NO_3^- and metabolizable organic C (Wrage et al., 2001). Denitrification is mostly confined to anaerobic deeper layers, waterlogged areas or the interior of soil aggregates (Tiedje et al., 1984; Leffelaar, 1986). And the production of nitrification, such as NO_3^- or NO_2^-, could be used in the denitrification by dinitrifiers. That is why the interfaces between these areas are the places where the production of N_2O is highest.

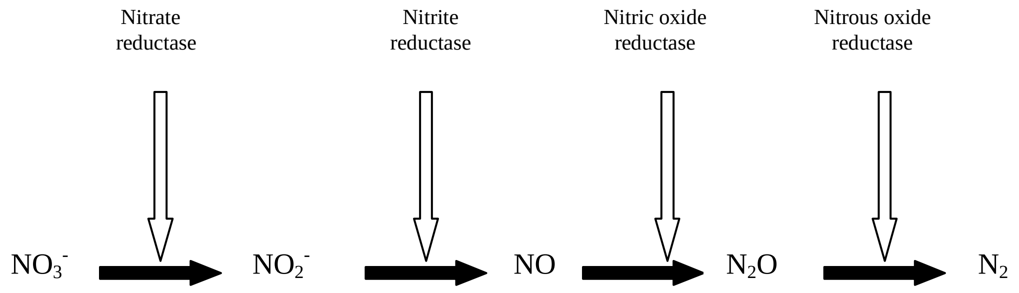


Figure 3.2, Denitrification: outline of the pathway and enzymes involved (after Hochstein and Tomlinson, 1988)

From biochemical aspect of N_2O production, the environmental factors influencing the nitrification and denitrification pathway in soil are N content, microbes, soil oxygen content, soil moisture, soil organic carbon content, pH, and temperature. These can be categorized into 3 groups: N input, N_2O producer and circumstance factors. N content of the soil is the input of the nitrogen in the environment. Different kinds of microbes are the nitrifers or denitrifers, which provide the enzymes for chemical reactions during the process of N_2O production. The rest of factors we call them circumstance factors, which are the most sensitive factors influenced by environment and also uncontrollable.

3.1.2. The influencing factors of N_2O emission at field scale

At the field scale, for instance, in a farm, all of the environmental factors mentioned in the biochemical process could also be measured, and analyzed for modeling. Figure 3.3 shows the influencing factors of N_2O emission in an intensive managed farm by G.L. Velthof [1997]

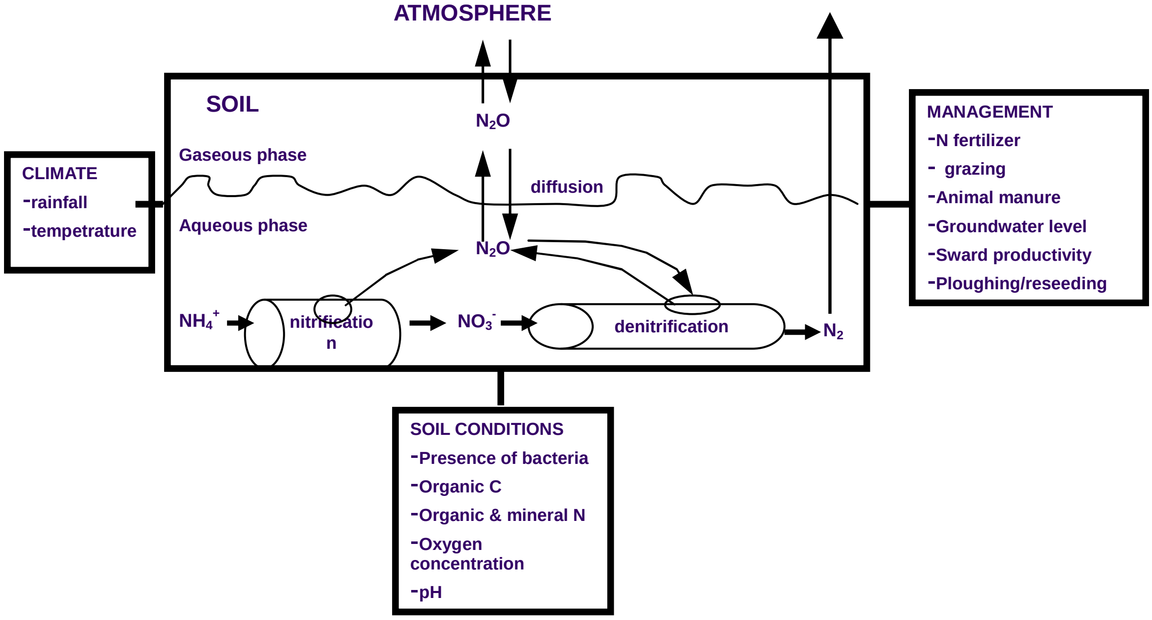


Figure 3.3,Major Factors controlling N2O emission from grass land soil (after Vetholf, 1997).

In figure 3.3, Vetholf [1997] classifies the influencing factors of N_2O emission into 3 categories: management style, soil condition, and climate. From the content of management style, the application of nitrogen for the farm is the most important factor, which not only determines the environmental factor of N input in the biogeochemical process, but also influences the N cycling in farm system. Though there is organic & mineral N factor in the soil condition, which also supply N content, this factor contributes much less than the application of fertilizers and manures in farm system. In the natural area, it could be a main source of N content.

The source of N input shows a very important factor for the researchers to distinguish different N cycles in biosphere. In pasture land, manure from animals is the major source of N input, which has a much higher emission factor than other fertilizers do. In arable land, the type of fertilizers influences the N_2O emission. In natural land, the main sources of N input are the deposition of N in the atmosphere and the biological fixation by plants. The amount of N input by deposition and biological fixation is much less than N input by fertilizer. That is why fertilized agricultural ecosystems emit more nitrous oxide than do most natural ecosystem (Mooney et al., 1987). An obvious conclusion obtained by researchers is that different ecosystem types exhibit distinctive patterns of N cycling (Groffman et al. 2000). This is also the basic understanding of this research. So the N input factor in the field scale is detailed as the type of ecosystem and the amount of N input.

The factor of presence of bacteria in the soil condition could be regarded as the N_2O producer in the N_2O production process. However, recent research shows that denitrifier and nitrifer diversity, as with all microorganisms in natural environments, is far greater than we have previously imagined (Coyne, 2009). This important factor is difficult to measure by current technology.

The other factors of soil condition and climate factors can be taken as the circumstance factors in the N_2O production process. Recent field experiments show that water-filled pore space (WFPS), soil temperature, and ground water level are important factors controlling N_2O emission ((Vetholf, 1997; Clayton, et al., 1997; Dobbie, et al., 1999). We use all these field measured factors to characterize the soil condition, instead of the ones listing in figure 3.3. Climate factors influence the N_2O emission by changing the soil condition. Unpredicted amount of rainfall and air temperature influence the soil temperature and soil water content of the N_2O production circumstance in a farm. Besides the climate factors, management style, such as the irrigation system and the way of cultivation, also influences the soil condition.

Overall, the influencing factor of N_2O emission in a farm scale could be the type of ecosystem, the amount of N input, the soil condition, climate and management style.

3.1.3. Spatial factors controlling annual N_2O emission at continental scale

Due to lack of long term research on N_2O emission at continental scale, the factors, that control N_2O emission at continental scale, are difficult to obtain directly from literatures. In this research, all the spatial factors chosen for annual N_2O emission at continental scale are based on the biogeochemical process of N_2O production and the field research. And the accessibility of the corresponding data is also taken into account for determining the spatial factors used in this research.

As mentioned before, the type of ecosystem, the amount of N input, the soil condition, the climate and the management style are five major factors for N_2O emission at field scale. Up scaling from field to continental scale and from daily to annual, the type of ecosystem and the management style would not change, and the amount of N input would be altered to the total annual N input. However, the spatial and temporal variability of the soil condition will be very large and the characteristic of climate will be altered as annual average temperature and annual cumulative precipitation, instead of air temperature and rain full.

There is no corresponding spatial data of ecosystem and management style for 27 countries in European Union (EU 27). Alternatively, we use land use type data to represent the ecosystem. And the differences of management style in land use types are significant. Though within same land use type, there are different management styles, the information of these is not complete for EU27. So in this research, we just take the land use type in the agricultural and natural area into account.

The total annual N input data is another important factor, which was obtained from the census data of European Union. Since different land use has its own source of N input, the total annual N input is calculated for each land use type respectively.

The soil condition data has many characteristics, such as soil temperature, water-fill pore space (WFPS), and soil organic content. However, due the large temporal and spatial variability, the corresponding data for the EU 27 is impossible to obtain. Soil type would be an alternative data, but there are more than thirty classes of soil types, which are too complex for the N_2O research. Most of the recent N_2O research just used the texture class of soil as a characteristic for the soil condition, because the soil texture could roughly show the differences of soil organic contend and the capability of water content. So we classify the soil type according to their texture types.

The climate data is the factor that will influence the soil condition. For the duration of one year, annual average temperature and annual cumulative precipitation are the main characteristics of climate data. For EU 27, there are data of bio-geographical regions, which not only take the climate factor into account but also the natural habitats, especially the species of plant. Because in natural area, climate and the species of plant are the main factors controlling N_2O emission, the bio-geographical region data is appropriate for this research.

These four factors have different influencing power in annual N_2O emission. Land use type representing the ecosystem types was assumed to have the largest influencing power, because the N cycling pattern is depends on the ecosystem type. Annual N input came as the second powerful influencing factor. The increase of N input to soil makes the soil N concentration rising, which means there are more N content that can be utilized for nitrification or denitrification. The concentration of N is the main drive of nitrification and denitrification in soil. The soil type and climate type were expected to have the lowest influencing power, since the circumstance factors have influence only after there are N_2O production processes in soil.

3.1.4. Source of corresponding spatial data

After determining the spatial factors controlling annual N2O emission, the corresponding spatial data were obtained yia internet or with help from other research groups at Wageningen University and Research centre.

Land use type—land use type data was derived from the Corine land cove data, published by the European Environment Agency (EEA) in 2007. Corine land cover 2000 (CLC2000) is an update for the reference year 2000 of the first CLC database which was finalized in the early 1990s as part of the European Commission program to Coordinate Information on the Environment (Corine). It covers 33 countries – 27 member states of European Union, and 6 countries in Europe, which are Albania, Bosnia and Herzegovina, Croatia, Liechtenstein, Macedonia FYROM, and Serbia Montenegro. Raster data of CLC2000 is derived from vector CLC2000 database by national teams within CLC2000 project, with pixel size of 100m and 250m respectively. Since this research is on continental scale, to reduce the time of processing the data, we use sthe grid with 250m pixel size.
(http://dataservice.eea.europa.eu/dataservice/metadetails.asp?id=1008)

N input— The Nomenclature of Territorial Units for Statistics (NUTS) was established by Eurostat more than 30 years ago in order to provide a single uniform breakdown of territorial units for the production of regional statistics for the European Union. Eurostat supply the scenario of annual fertilizer application and number of animals kept in farm in the agriculture area on the NUTS 2 region level. Based on the amount of animal we can calculate the manure produced in one year. So we obtained the annual N application for most of the regions in EU 27. This census data could be link to the spatial data of NUTS 2 region. The census data could be downloaded from soil group of Wageningen University and Research centre. (http://www.scammonia.wur.nl/UK/Database/)

Soil type—the Soil Geographical Database of Eurasia at 1:1,000,000 (SGDBE) is published by the Join Research Centre of European Union, which is a digitized European soil map. The newest version is from 2000. We used the vector data set, which includes detailed information with soil texture type.
(http://eusoils.jrc.ec.europa.eu/ESDB_Archive/ESDB/index.htm)

Climate region—the bio-geographical region dataset contains the official delineations used in the Habitats Directive. The map published by European Environment Agency in 2008. This version corrects the coast line of the EU 27 so that it spatially matches the Corine land cove data.
(http://dataservice.eea.europa.eu/dataservice/metadetails.asp?id=1054)

3.2. Results of processing spatial data and strata

3.2.1. The spatial data after preprocessing

The four spatial factors were land use type, annual N input, soil type and climate region. The source data of them can be found in Appendix. After preprocessing the source data, we obtained the four spatial data for combination

Land use type is shown in figure 3.4, with 3 classes, which are arable land, pasture land and natural land. It covers 33 countries – 27 member states of European Union, and 6 countries in Europe, which are Albania, Bosnia and Herzegovina, Croatia, Liechtenstein, Macedonia FYROM, and Serbia Montenegro. So the area covered by the land use map is larger than other 3 maps. Arable land covers about 42% of the whole area. Especially in the western, central and eastern Europe, arable land is the main land use type. Pasture land only takes up 9% of the whole map, which mostly occurs in the Western Europe and British Isles. Nature land covers the rest of the area, which occupy the large area in Sweden and Finland.

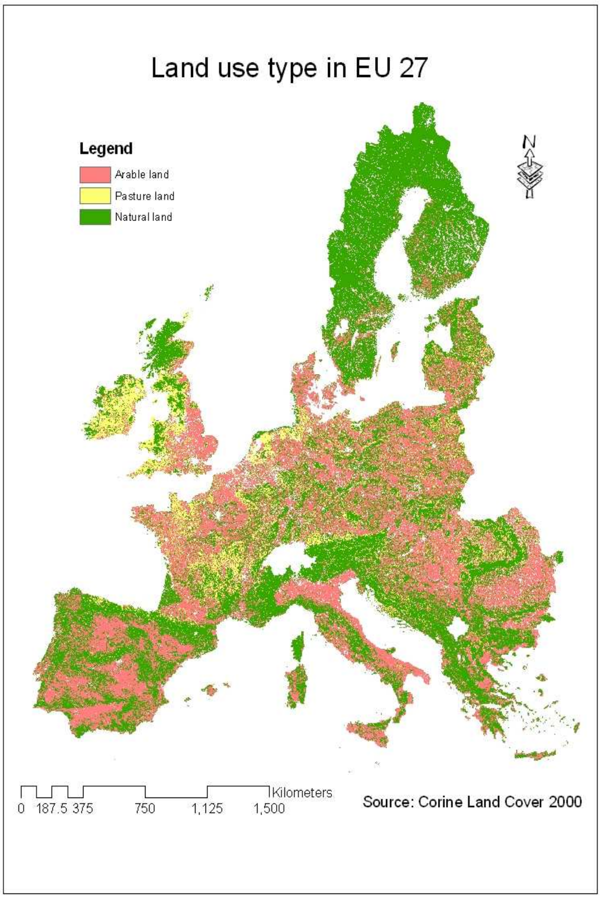


Figure 3.4,Simplified land use type from Corine land cover 2000.

Annual N input data is shown in figure 3.5, with five classes, which are 0 – 50, 50 – 100, 100 – 150, 150 – 200 and > 200. The unit is kg N/ha. The annual N input is also calculated according to these three land use types. The annual N input of agricultural area sums the N application to the field, biological fixation and deposition from atmosphere. The N application to field of arable land is different from that of pasture land, so for N input for these two land use type is calculated separately. And there is no application of nitrogen to natural area by human. The source of N input of natural area is only from biological fixation and deposition from atmosphere.

This map covers the study area – 27 countries in European Union. In about 90 % of the study area, the annual N input is below 100 kg N/ha. Because the annual N input of nature land is always smaller than 40 kg N/ha in Europe, 63% of the area has an annual N input below 50 kg N/ha. High annual N input is observed in Netherlands, Ireland, Denmark, France, some parts of U.K. and Germany.

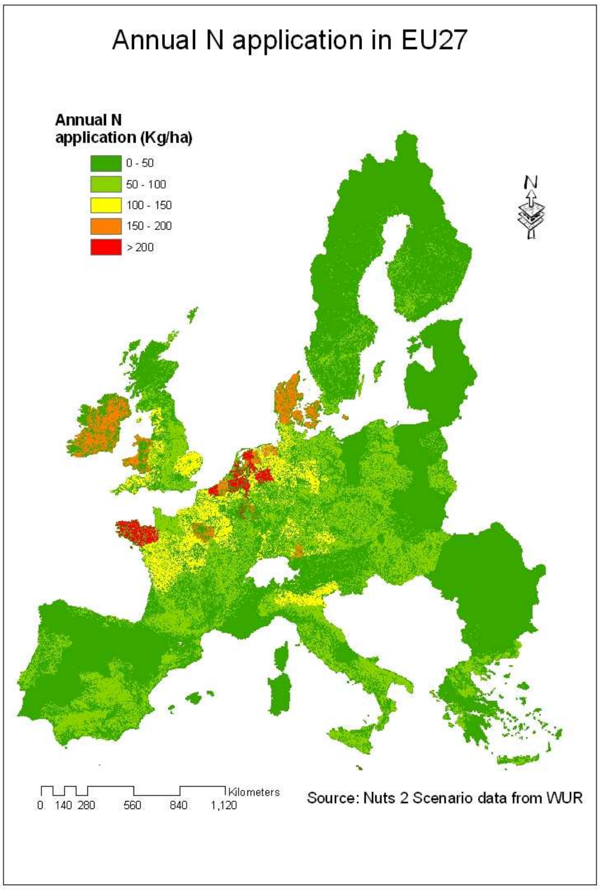


Figure 3.5, Annual N input classes in the EU 27.

The soil type is shown in figure 3.6, with 3 classes, which are peat, clay and sand & rock. Soil map covers EU27 countries. Clay soil takes up about 74% of the study area. Peat soil mainly occurs in U.K., Netherlands, Poland, Finland and Baltic states, with 4% of the total study area. There is large area with Soil & rock soil in the Finland, Denmark, some parts of Germany and Poland.

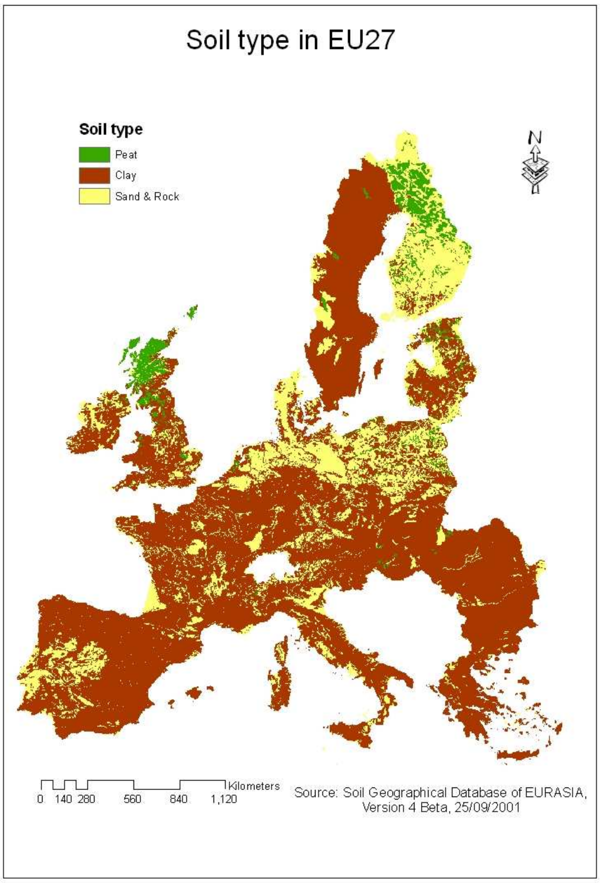


Figure 3.6, Generalized soil type in the EU 27.

The Climate regions are shown in figure 3.7, with 5 classes, which are Alpine, Atlantic, Boreal Mediterranean and Continental. It covers EU27 countries. Continental region covers the 33% of the total study area, which contains the biogeographically regions of Continental, Macaronesia, Pannonian and Steppic. The southern Europe has the typical Mediterranean region, with 21% of the total area. Sweden, Finland and Baltic states have Boreal region, with 19% of the total area. The area with high annual N input is observed with Atlantic region, which takes up 18 % of the study area. The rest is the Alpine region, which is mainly observed in the Alpine mountains and a part of Sweden.

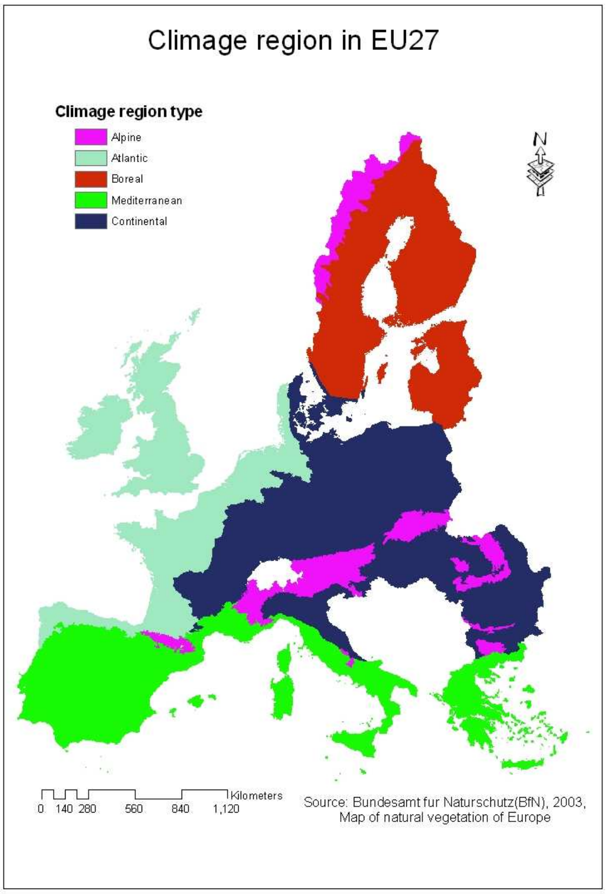


Figure 3.7, Climate regions in the EU 27.

3.2.2. Dissimilarity tables

Even after reducing the amount of classes for each spatial factor by reclassification, there are still 3 classes for land use type, 3 classes for soil type, 5 classes for climate region and 5 classes for annual N input. Each class of each spatial factor is given an ID number, in Figure3.8. After combination, there should be 3*3*5*5=225 features.

ID of Land use type:	ID of Soil type:	ID of annual N input:	ID of Climate region:
1. Arable land;	1. Peat;	1. 0 – 50;	1. Alpine;
2. Pasture land;	2. Clay;	2. 50 – 100;	2. Atlantic;
3. Nature land.	3. Sand & Rock.	3. 100 – 150;	3. Boreal;
		4. 150 – 200;	4. Mediterranean;
		5. >200.	5. Continental.

Figure 3.8, ID for classes of spatial factors.

To emphasize the importance of land use type, we categorize 225 features into 3 classes according to the land use type. For the classes of arable land and pasture land, there should be 3*5*5=75 features, because the remaining three spatial factors determine the amount of features. Because all features in Nature {} have the annual N input less than 50 kg N/ha, there is only 3*5=15 features in Nature {}. Finally, we only obtain 75+75+15=165 features.

We call denote three sets of strata as

Arable {} = 75 features which have the arable land use type;

Pasture {} = 75 features which have the pasture land use type;

Nature {} = 15 features which have the nature land use type.

For the remaining three spatial factors, we create the dissimilarity table for each one.

Dissimilarity table of annual N input: There are 5 classes – 0 – 50, 50 – 100, 100 – 150, 150 – 200, and > 200, with the unit of kg N/ha. We set the dissimilarity table as table 3.1. We denote the dissimilarity value in table 3.1 as N [i, j]. .

Table 3.1, Dissimilarity table of annual N input.

	0 - 50	50 - 100	100 - 150	150 - 200	> 200
0 - 50	0	0.5	0.8	1	1
50 - 100	0.5	0	0.5	1	1
100 - 150	0.8	0.5	0	0.5	1
150 - 200	1	1	0.5	0	1
> 200	1	1	1	1	0

Because of the high spatial variability of N application, with the increase of N input, the variance of N_2O emission also rises. So, the difference between classes, which have lower annual N input are smaller than that between classes, which have higher annual N input. Among the three remaining spatial factors, annual N input is the most important factor, so all dissimilarity value is not smaller than 0.5

Dissimilarity table of soil type: The second important factor is the soil type. It has 3 classes – peat, clay and sand & rock. We set the dissimilarity table as table 3.2. We denote the dissimilarity value in table 3.2 as S [i, j].

Table 3.2, Dissimilarity table of soil type.

	Peat	Clay	Sand & Rock
Peat	0	0.7	1
Clay	0.7	0	0.9
Sand & Rock	1	0.9	0

With the same amount of N input, N_2O emission from sand & rock area is lower than that from peat and clay. The N_2O emission from peat soil is the highest. Soil type is a very important factor, which controlling the environment of N_2O emission, so all the value of dissimilarity table is higher than 0.5.

Dissimilarity table of climate region: The last factor is the climate, which has 5 classes – Alpine, Atlantic, Boreal, Mediterranean, and Continental. We set the dissimilarity table as table 3.3. We denote the dissimilarity value in table 3.3 as C [i,j].

Table 3.3, Dissimilarity table of climate region.

	Alpine	Atlantic	Boreal	Mediterranean	Continental
Alpine	0	1	1	1	1
Atlantic	1	0	1	0.8	0.8
Boreal	1	1	0	1	0.8
Mediterranean	1	0.8	1	0	0.8
Continental	1	0.8	0.8	0.8	0

The climate factor effects the N_2O emission by influencing the environment of soil condition. So, all the value is above 0.5. The difference between classes of climate factor is difficult to measure, for there is no sufficient research on this. We assume that at the same amount of N input, Alpine area and boreal area has the lowest N_2O emission, while Atlantic has the highest. Continental area has the largest size, which means that the variability of temperature and precipitation in this area is very high. So, the difference between Continental area and other areas is smaller. Mediterranean area has high precipitation and also high annual temperature. High precipitation makes it close to Atlantic are, while during summer, high temperature in this area is close to Continental area.

The dissimilarity table of features: Because we categorized the features of combination into 3 classes – Arable {}, Pasture {}, and Nature {}, there should be 3 dissimilarity tables. However, all the features in these three classes are determined by the three remaining spatial factors. According to EQ.2.13, the value of dissimilarity table of features A [p, q] is calculated from the dissimilarity value of the dissimilarity table of spatial factors Sf_i [p_i, q_i] and their weights m_i. We set the same weight of spatial factors m_i for the dissimilarity table of Arable {} and Pasture {} in table 3.4, so the dissimilarity table of Arable {} and Pasture {} are the same, which are calculated as below:

$$A[p,q] = 0.5\times N[i,j] + 0.3\times S[i,j] + 0.2\times C[i,j] \qquad (3.1)$$

Table 3.4, Weight table of the remaining three factors for Arable {} and Pasture {}.

	Annual N input	Soil type	Climate
Weight	0.5	0.3	0.2

The annual N input determines how much nitrogen in soil. Soil condition is determined by soil type and climate. As the source of N_2O, the nitrogen in soil is much more important than soil condition, so annual N input has a weight of 0.5 in table 3.4. Soil type is a container, where N_2O is produced. Climate would influence the environment of soil condition, by wind, cloud, rain full and air temperature. The container – soil type is a little bit more important than the environment variable – climate. So we set the weight of soil type as 0.3 and that of climate as 0.2 in table 3.4.

Because all features in Nature {} have the annual N input less than 50 kg N/ha, the weight of annual N input is zero. We set the weight table of the soil type and climate in table 3.5. The dissimilarity table of Nature {} is calculated as below:

$$A[p,q] = 0.6 \times S[i,j] + 0.4 \times C[i,j] \qquad (3.2)$$

Table 3.5, Weight table of the remaining three factors for Nature {}.

	Annual N input	Soil type	Climate region
Weight	0	0,6	0,4

3.2.3. The cluster after agglomerating.

According to the three classes of features – Arable {}, Pasture {}, and Nature {}, there are should be 75+75+15=165 features.

For two sets of features – Arable {} and Pasture {}, we agglomerate them into 7 clusters, of which the clustering tree shows in figure 3.9. In figure 3.9, v_i represents ith feature in Arable {} and Pasture {}. From figure 3.9, we can see that v_i can be grouped into 7 clusters, which show in the red boxes.

According to the clustering tree in figure 3.10, we group the features in Nature {} into 3 clusters, which shows in the red boxes.

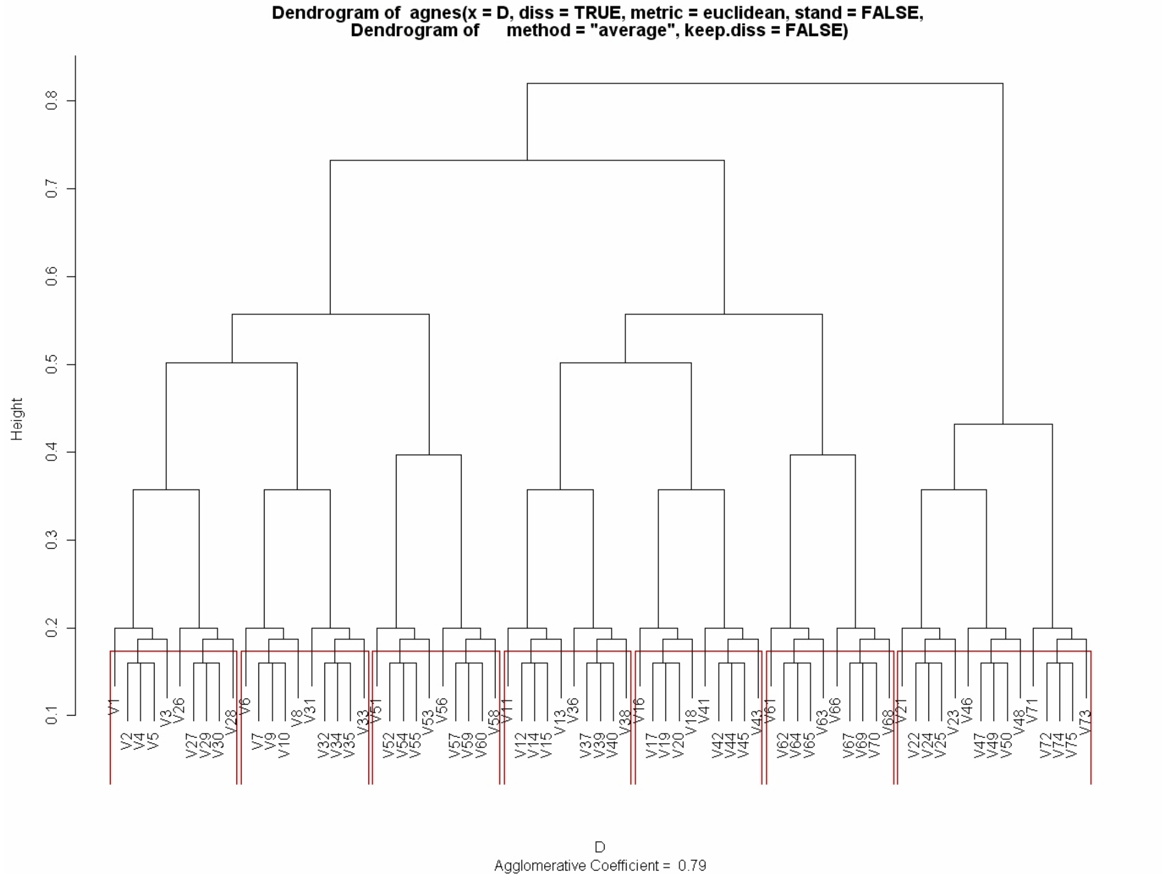


Figure 3.9, Tree of agglomerative hierarchical cluster for Arable {} and pasture {} .

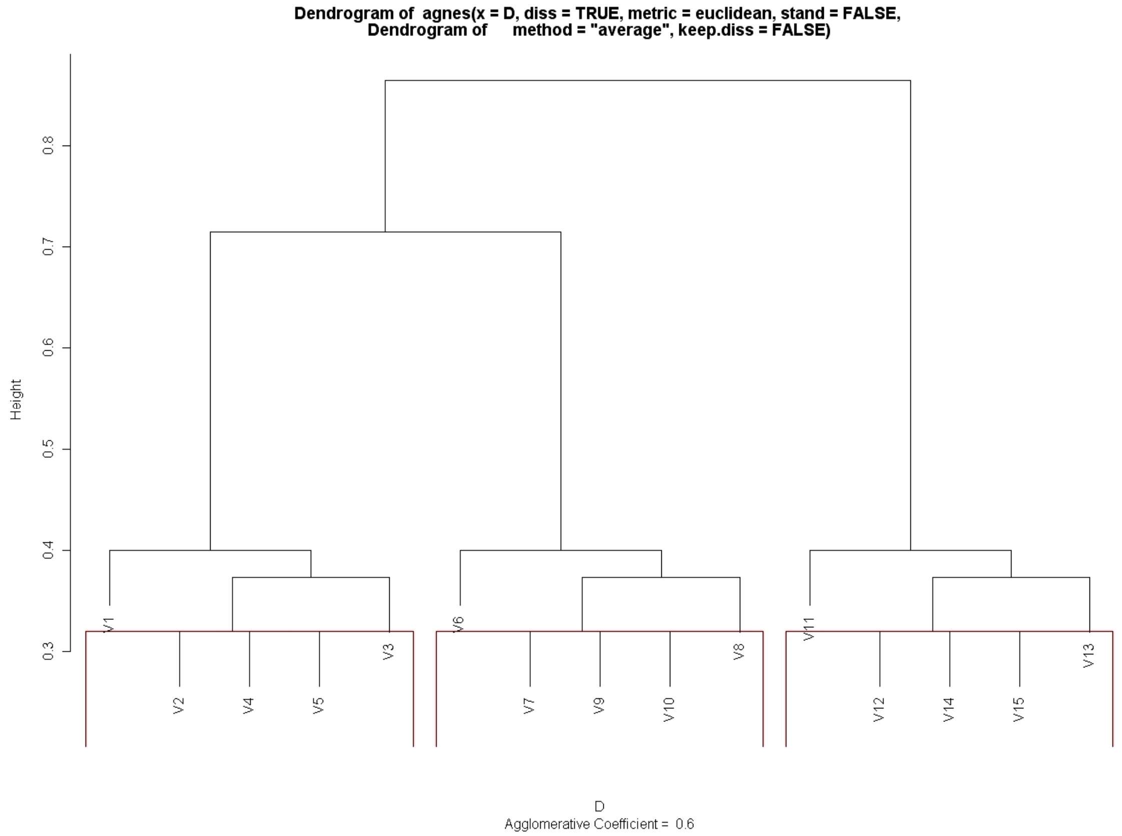


Figure 3.10, Tree of agglomerative hierarchical cluster for Nature {}.

At a certain size of n, the smaller is AC, the more acceptable is this clustering. From figure 3.9, we can see the AC of agglomerative hierarchical cluster for Arable {} and Pasture {} is 0.79. From figure 3.10, the AC of agglomerative hierarchical cluster for Nature {} is 0.6. But this value dose not gives the agglomerative coefficient of the clustering at a certain cluster number, due to the limitation of the function “agnes” in R. The efficiency of the clustering has to be evaluated in the next section.

3.2.4. The strata designed for annual N2O emission

We combine four spatial data in ArcGIS and obtained 129 features. Some of these features have other land use type than arable land, pasture land and nature land. We remove the features with other land use type and finally we only obtain 114 features. Within 114 features, according to the land use type they have, we categorize them into 3 sets, as Arable {}, Pasture {}, and Nature {}.

In figure 3.9 and 3.10, we agglomerate the features of each set into clusters. In ArcGIS, we reclassify the features to the corresponding strata according to the result of clustering. We make the criteria to make a decision what the number of clusters is:

1. The features with low annual N input are clustered into the same stratum as much as possible, according to the class of annual N input they have.
2. The features with high annual N input are clustered into different stratum according to their soil types.
3. The difference between peat and clay is much smaller than that between peat and sand and also smaller than that between clay and sand.
4. The difference between climate regions is smaller than that between soil types.

This criteria is implemented to the clustering of all three feature sets. Because we set the range of the total number of clusters between 10 and 20, after testing, we determine the cluster number for each feature sets as 7 clusters for Arable {}, 7 clusters for Pasture {} and 3 clusters for Nature {}. After clustering, the reclassification of three land use sets needs be merged to make the reclassification table for the combination of four spatial data. In reclassification table, we set the sequence of clusters as table 3.6

Table 3.6, Sequence of clusters in the reclassification table of combination.

	Stratum in reclassification table of combination
Arable land	stratum 1 - stratum 7
Pasture land	stratum 8 - stratum 14
Nature land	stratum 14 - stratum 17

After reclassification, we obtain the 17 strata and their attribute table. The final map of 17 strata is showing in figure 3.10. For each stratum, the area is shown in table 3.7. According to Eq.2.5, the stratum weight w_i is calculated in table 3.7. Even after agglomerative clustering, the weight of each stratum shows large variability, from 0.334 to 0.003. And there are only four strata, which have a weight over than 0.1. They are stratum 1, stratum 2, stratum 16 and stratum 17, which take up 74.3% of total weight. The information of each stratum could be found in the Questionnaire in Appendix.

Table 3.7, Information of strata area.

	Area (km^2)	w_i
stratum1	4.8E+05	0.121
stratum2	6.6E+05	0.165
stratum3	1.6E+05	0.039
stratum4	2.1E+04	0.005
stratum5	3.1E+04	0.008
stratum6	2.4E+05	0.061
stratum7	6.7E+04	0.017
stratum8	8.8E+04	0.022
stratum9	1.0E+05	0.026
stratum10	3.7E+04	0.009
stratum11	5.2E+04	0.013
stratum12	1.1E+04	0.003
stratum13	4.4E+04	0.011
stratum14	2.4E+04	0.006
stratum15	1.4E+05	0.036
stratum16	1.3E+06	0.334
stratum17	4.9E+05	0.123

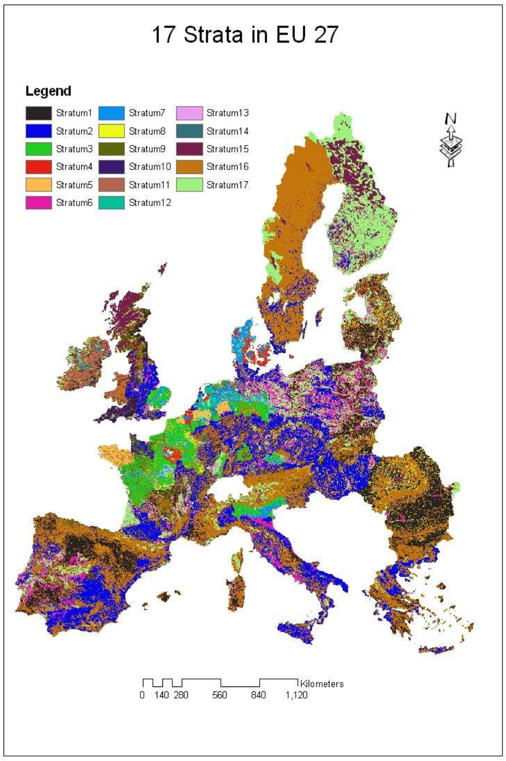


Figure 3.11, Strata designed for N2O emission sampling.

3.3. Results of annual within stratum variance

3.3.1. Answers of the questionnaire

After consulting experts, we get five answers of the questionnaire. The figures from 3.12 to 3.28 show the answers for each stratum. A_i represents the answers from the *ith* expert. The bottom number of each line is the value of estimate annual N_2O emission for Quartile 1, for which the expert believe that there is 25% probability that the actual annual N_2O emission will be smaller. The top number of each line is the value of estimate annual N_2O emission for Quartile 3, for which the expert believe that there is 25% probability that the actual annual N_2O emission will be greater. The middle value of each line is the value of estimate annual N_2O emission for Median, for which the expert believe that there is 50% probability that the actual annual N_2O emission will be greater.

The answers of Quartile 1, Median and Quartile 3 from experts for each stratum show large variability from figure 3.12 to figure 3.28. For all strata, except stratum 15, stratum 16 and stratum 17, A3 always has the lowest answers of Quartile 1, Median and Quartile 3, while A4 has the highest answers of them. The answer of Quartile 3 from A3 is much smaller than the answer of Quartile 1 from A4. The answers of Quartile 1, Median and Quartile 3 from experts A1 and A2 range between the answer of Quartile 1 from A3 and that from A4. The answers of Quartile 1, Median and Quartile 3 from experts A5 is also higher than them from experts A1, A2 and A3 in most of strata, but close to the answers from A4. So the experts can be separated into 2 groups according to the answers they give.

The length of each line shows the distance from Quartile 1 to Quartile 3. Because the scale of Y axis is logarithmic, ranging from 0.01 to 1000, so in different range, the unit distance is different. The top ranges have the higher unit distance. From figure 3.12 to 3.28, we see that the distances of answers from experts also show large variability. For all strata except stratum15, 16 and 17, the distance of answers from A4 is always the largest among answers from 5 experts. For the stratum 15, the stratum 16 and the stratum 17, A5 has the largest distance of answers. A3 always has the smallest distance of answers for all strata. The distances of answers from A1 and A2 are between that of A3 and A4

According to Eq.2.16 and Eq.2.17, we calculate the max distance of the answers and the mean distance for each stratum, which are also shown from figures from 3.12 to 3.28. Because of the large variability of answers from experts, the difference between max distance and mean distance are also large for each stratum.

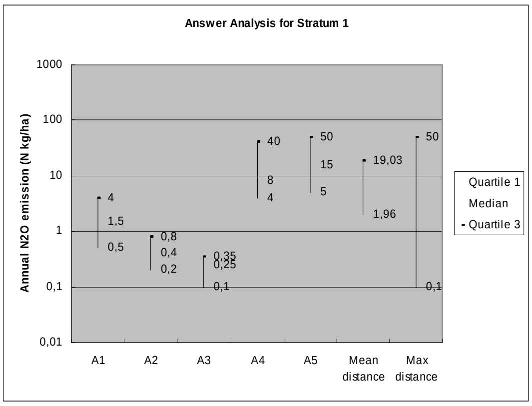


Figure 3.12, answer analysis for stratum 1

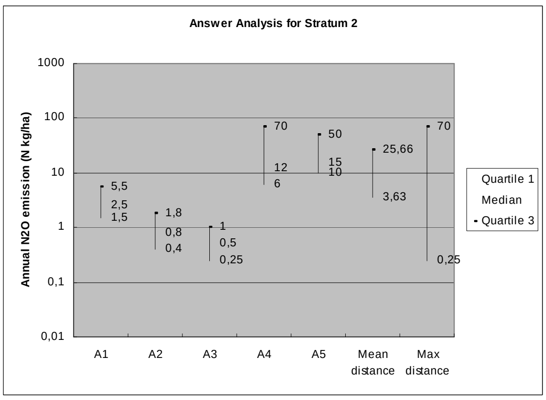


Figure 3.13, answer analysis for stratum2

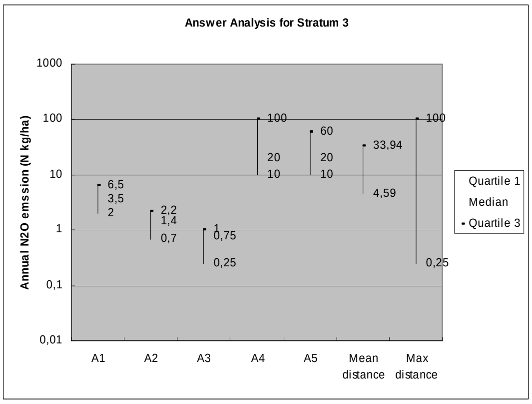


Figure 3.14, answer analysis for stratum 3.

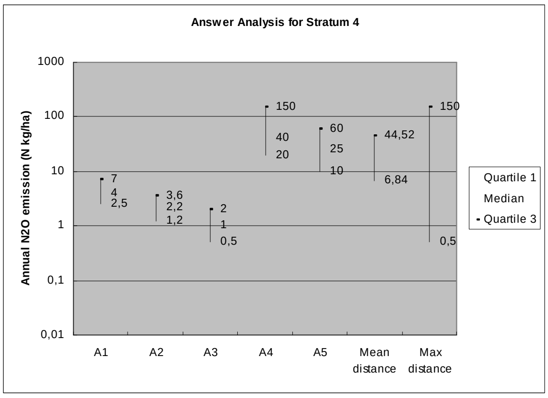


Figure 3.15, answer analysis for stratum 4.

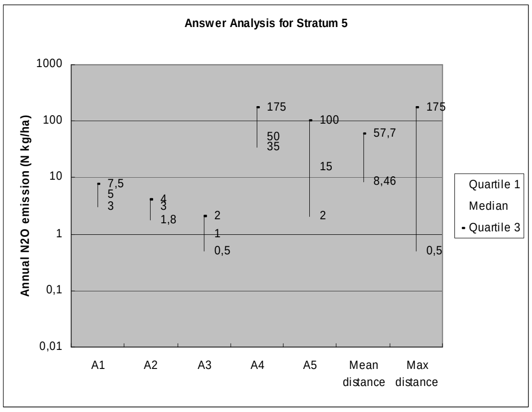


Figure 3.16, answer analysis for stratum 5

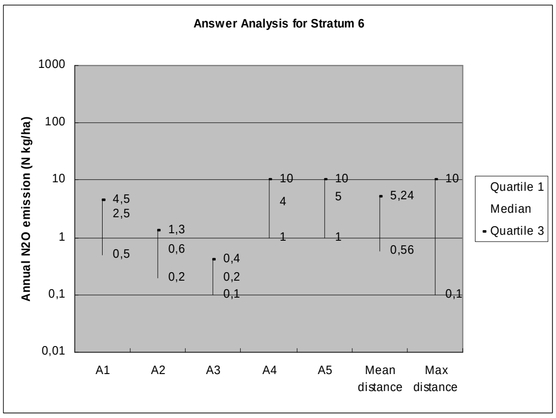


Figure 3.17, answer analysis for stratum 6.

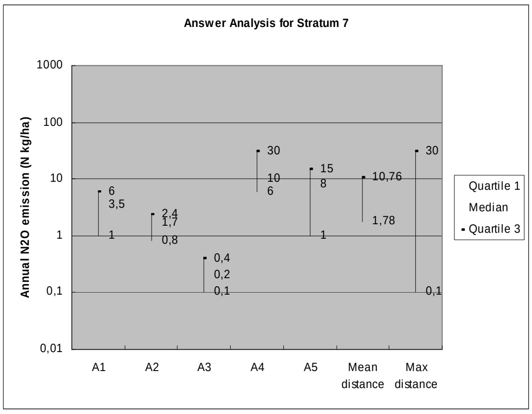


Figure 3.18, answer analysis for stratum 7.

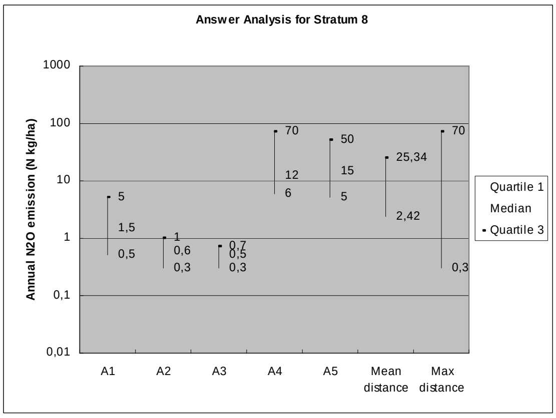


Figure 3.19, answer analysis for stratum 8.

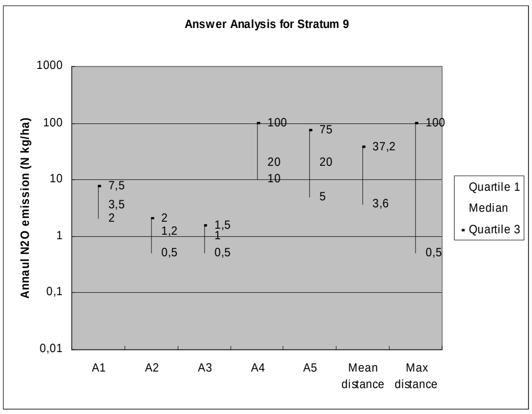


Figure 3.20, answer analysis for stratum 9.

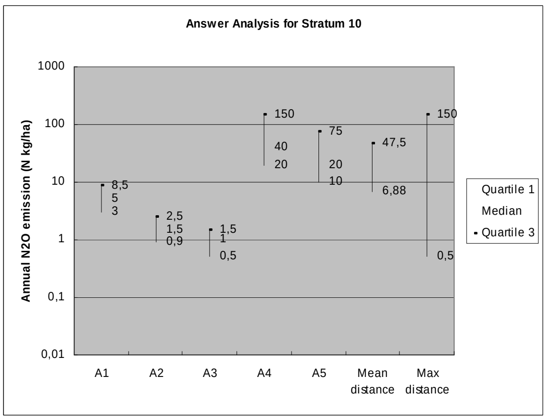


Figure 3.21, answer analysis for stratum 10.

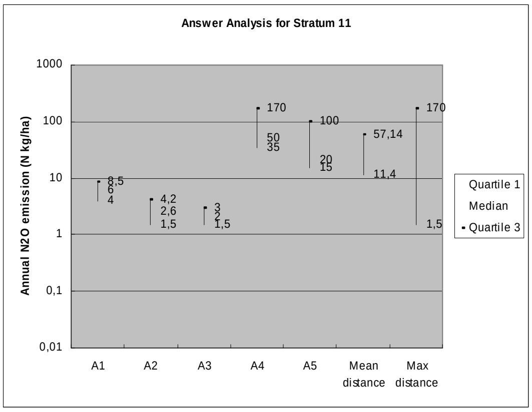


Figure 3.22, answer analysis for stratum 11.

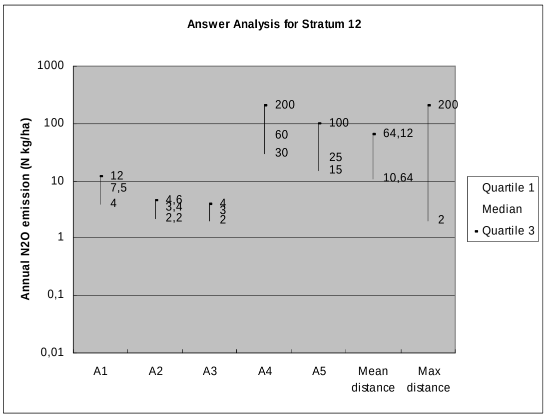


Figure 3.23, answer analysis for stratum 12.

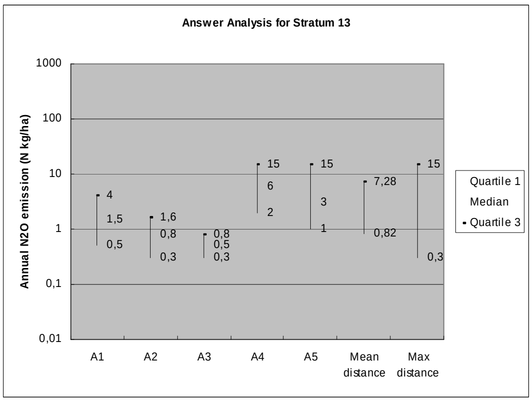


Figure 3.24, answer analysis for stratum 13.

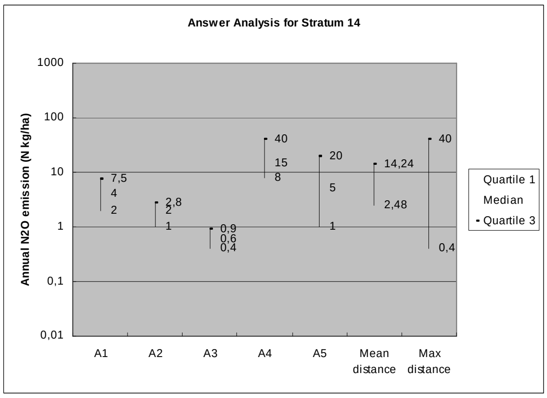


Figure 3.25, answer analysis for stratum 14.

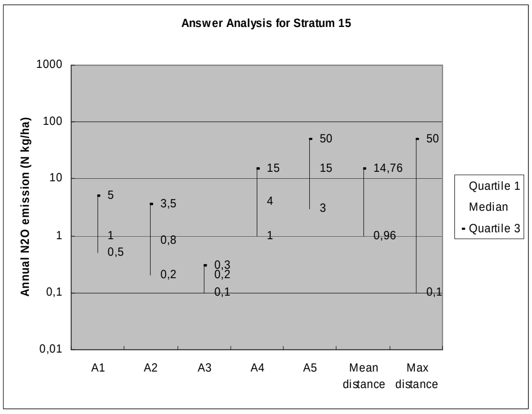


Figure 3.26, answer analysis for stratum 15.

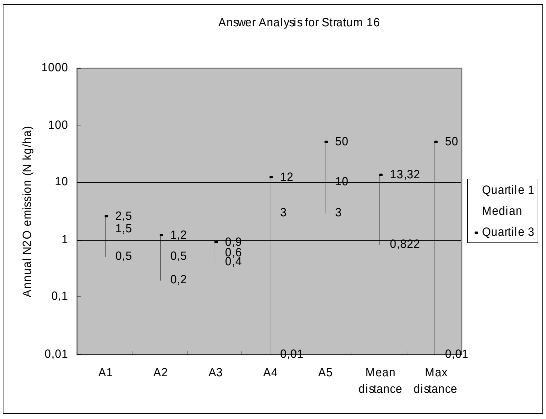


Figure 3.27, answer analysis for stratum 16.

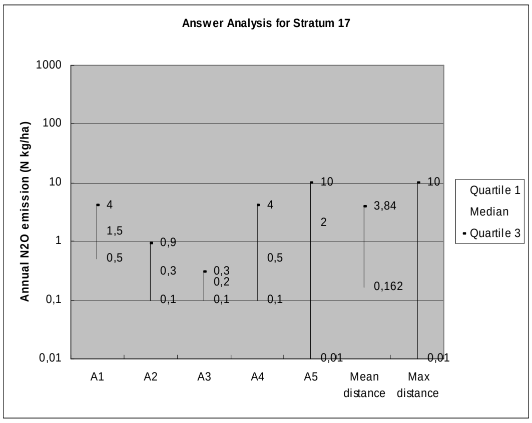


Figure 3.28, answer analysis for stratum 17.

3.3.2. Standard deviation for each stratum

The standard deviation for each stratum is calculated from the distance between Q3 and Q1. There are two sets of distance, which are max distance and mean distance, shown in figure 3.28. Though the values of two distances have large difference, the trends of these values are the same. The values of both distances increase to the top at stratum 5 and then suddenly drop to a very low value at stratum 6. From stratum 6, the values start to increase again, and reach another top at stratum 12 and drop down at stratum 13. From stratum 13 to 17, the fluctuation of values is not so large as before. The stratum6, stratum 13 and stratum 17 always have the minimum values of both distances. And the top values of both distances occur in stratum 5 and stratum 12.

All 17 strata can be categorized into 3 groups, according to table 3.6. The strata with arable land shows the same fluctuation of distance with the strata with pasture land.

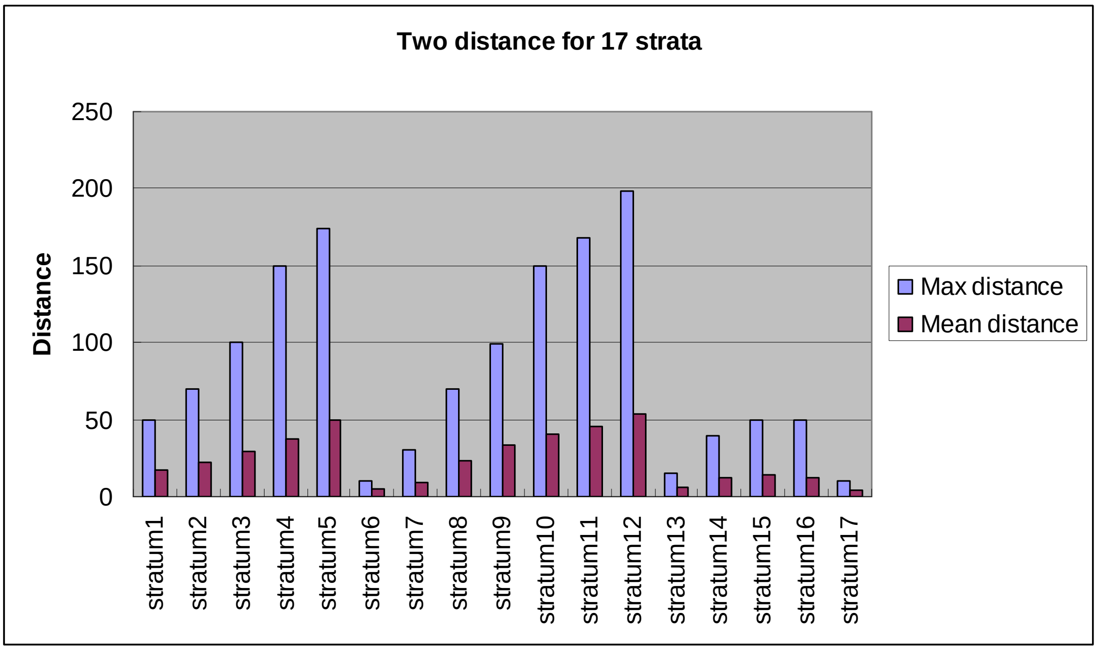


Figure 3.29, Max distance and mean distance for the 17 strata.

Because there are two kinds of distances, we have two scenarios of the estimation of standard deviation s_i. We call the scenario of the estimation of s_i, which is calculated from max distance as the pessimistic scenario. And the scenario of the estimation of s_i, which is calculated from mean distance as the optimistic scenario.

From the two sets of distances, we calculate the pessimistic s_i and optimistic s_i according to

2.20, respectively, which are shown in table 3.2. The standard deviations show the same fluctuation with pessimistic scenario and optimistic scenario of standard deviation. The stratum 6, stratum 13 and stratum 17 have the lowest standard deviations, while stratum 5 and stratum 12 have the highest standard deviations.

Table 3.8, Pessimistic scenario and optimistic scenario of standard deviation for 17 strata.

	Pessimistic SD	Optimistic SD
stratum1	37,24	12,74
stratum2	52,05	16,44
stratum3	74,44	21,90
stratum4	111,57	28,12
stratum5	130,22	36,75
stratum6	7,39	3,49
stratum7	22,31	6,70
stratum8	52,01	17,10
stratum9	74,25	25,07
stratum10	111,57	30,31
stratum11	125,75	34,13
stratum12	147,76	39,91
stratum13	10,97	4,82
stratum14	29,55	8,78
stratum15	37,24	10,30
stratum16	37,31	9,33
stratum17	7,46	2,74

3.4. Total variance at a sample size

3.4.1. The sample size for each stratum

There are two scenarios of standard deviation, the pessimistic s_i and the optimistic s_i, are obtained according to the answers of questionnaire from experts, which is shown in table 3.8. For each stratum, according to its size of area, we calculate the stratum weight w_i in table 3.7.

In table 3.9 and 3.10, we show the pessimistic scenario and optimistic scenario for each stratum. From table 3.9 and table 3.10, we can find that stratum 4, stratum 5, stratum10, stratum 11 and stratum12 have the highest standard deviation for both scenarios, but they do not have very high weight of stratum. So the $w_i^2 * s_i^2$ for them are no more than 3. Stratum 16, stratum 1 and stratum 2, which have the highest weight, also have the highest $w_i^2 * s_i^2$ among all strata in both tables.

Table 3.9, Pessimistic scenario for 17 strata.

	Area (km²)	w_i	Pessimistic SD	Pessimistic scenario of $w_i^2*s_i^2$
stratum1	4,82E+05	0,12	37,24	20,34
stratum2	6,56E+05	0,16	52,05	73,73
stratum3	1,57E+05	0,04	74,44	8,62
stratum4	2,13E+04	0,01	111,57	0,36
stratum5	3,10E+04	0,01	130,22	1,03
stratum6	2,41E+05	0,06	7,39	0,20
stratum7	6,74E+04	0,02	22,31	0,14
stratum8	8,85E+04	0,02	52,01	1,34
stratum9	1,05E+05	0,03	74,25	3,83
stratum10	3,74E+04	0,01	111,57	1,10
stratum11	5,16E+04	0,01	125,75	2,66
stratum12	1,07E+04	0,00	147,76	0,16
stratum13	4,38E+04	0,01	10,97	0,01
stratum14	2,38E+04	0,01	29,55	0,03
stratum15	1,44E+05	0,04	37,24	1,81
stratum16	1,33E+06	0,33	37,31	155,61
stratum17	4,88E+05	0,12	7,46	0,84

Table 3.10, Optimistic scenario for 17 strata.

	Area (km²)	w_i	Optimistic SD	Optimistic scenario of $w_i^2*s_i^2$
stratum1	4,82E+05	0,12	12,74	2,38
stratum2	6,56E+05	0,16	16,44	7,35
stratum3	1,57E+05	0,04	21,90	0,75
stratum4	2,13E+04	0,01	28,12	0,02
stratum5	3,10E+04	0,01	36,75	0,08
stratum6	2,41E+05	0,06	3,49	0,04
stratum7	6,74E+04	0,02	6,70	0,01
stratum8	8,85E+04	0,02	17,10	0,14
stratum9	1,05E+05	0,03	25,07	0,44
stratum10	3,74E+04	0,01	30,31	0,08
stratum11	5,16E+04	0,01	34,13	0,20
stratum12	1,07E+04	0,00	39,91	0,01
stratum13	4,38E+04	0,01	4,82	0,00
stratum14	2,38E+04	0,01	8,78	0,00
stratum15	1,44E+05	0,04	10,30	0,14
stratum16	1,33E+06	0,33	9,33	9,73
stratum17	4,88E+05	0,12	2,74	0,11

According to Eq.2.6, at a certain total sample size n, ni for each stratum is determined by the percentage of $w_i^2*s_i^2/\sum w_i^2*s_i^2$, for which there are also two scenarios, shown in figure 3.30. From figure 3.30, we can see that the distribution of sample size for each stratum has large difference. For both scenarios of $W_i^2*s_i^2/\sum W_i^2*s_i^2$, stratum 16 has the highest percentage of the sample size in total. The stratum1, stratum 2, stratum3 and stratum 16 occupy 90 % of the total sample size.

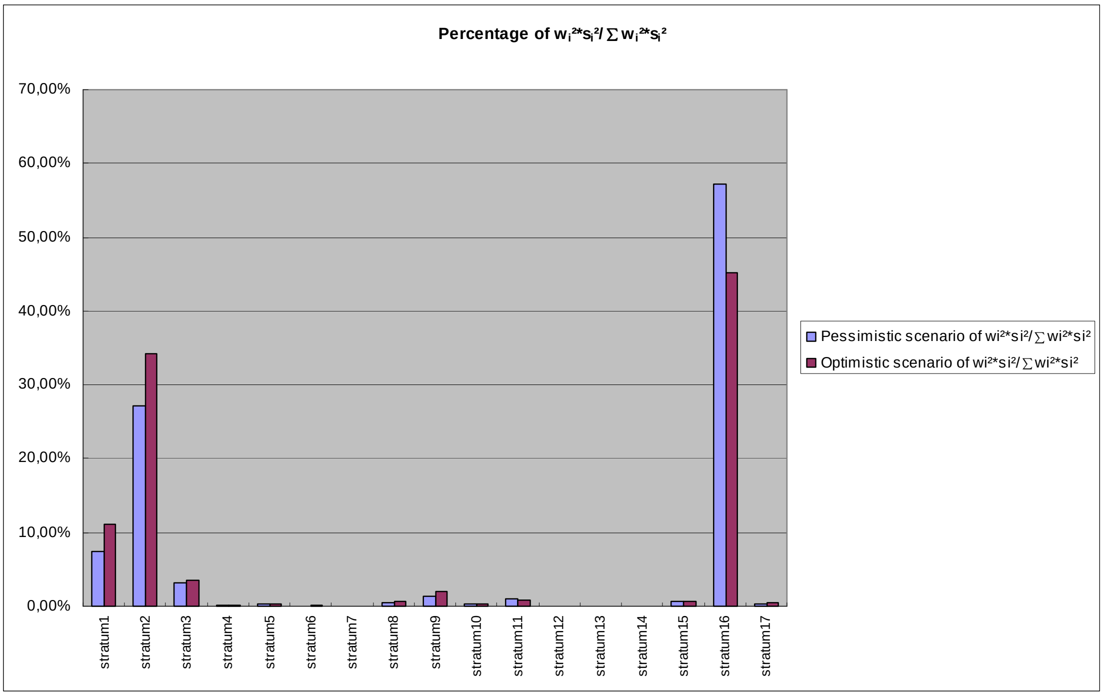


Figure 3.30, Percentage of $wi^2*si^2/\sum wi^2*si^2$ for pessimistic and optimistic scenarios.

If we set the total sample size scenario as:

$N = \{N_1, N_2, \ldots, N_k\}$, $k = 1, 2, \ldots, n$.

At a certain total sample size $N_i \in N$, we can calculate the ni for each stratum according to the pessimistic scenario and optimistic scenario of $w_i^2*s_i^2/\sum w_i^2*s_i^2$. If the total sample size is not large enough, the stratum4, stratum5, stratum 6, stratum 7, stratum 8, stratum 10, stratum 12, stratum13, stratum 14 and stratum 17would have no sample size, due to the tiny percentage they have in figure 3.30.

After we calculate the sample size ni for each stratum according two scenarios, we have two scenario of the total sample size as below:

$N_p = \{N_{p1}, N_{p2}, \ldots, N_{pk}\}$

$N_o = \{N_{o1}, No_2, \ldots, N_{ok}\}$

$k = 1, 2, \ldots, n$

N_p is the total sample size set calculated from pessimistic scenario. N_o is the total sample size calculated from optimistic scenario.

In table 3.11, we give an example of the scenario of total sample size N, and its corresponding N_p and N_o. From table 3.11, We can find that the difference between the elements of N_p and N_o is very small. With the increase of Ni in N, the difference among N_i, Npi, and Noi becomes very small.

Table 3.11, Scenarios of total sample size N, and its corresponding Np and No.

N	100	200	350	500	700	1000	1500	2000	3000	4000	5000	6000	7000	8000	10000
Np	120	217	362	510	710	1010	1508	2006	3003	4002	5004	6003	7004	8002	10002
No	119	215	361	509	709	1007	1507	2005	3001	4002	5003	6003	7003	8002	10001

3.4.2. The total variance at a certain total sample size

After calculating the n_i, s_i, and w_i for each stratum, we can calculate the total variance V_i at a certain total sample size according to Eq.2.3. Because there are pessimistic scenario and optimistic scenario, there are also two sets of total variance.

In figure 3.31, we plot the two sets of total variance at each total sample size. The total variance calculated from pessimistic scenario has a higher start point than that from optimistic scenario does. The total variances calculated from both scenarios are decreased fast when the total sample size is smaller than 1000. This means that the accuracy of estimation will increase quickly when the sample size is raised from 100 to 1000. However, if the sample size is bigger than 1000, the total variance calculated from optimistic scenario will not decrease significantly. So does the total variance calculated from pessimistic scenario.

From figure 3.31, we can also estimate the range of total sample size need for certain accuracy. If we wan to reduce the total variance of mean annual N_2O emission to 2 kg N/ha, we need the total sample size ranging from less than 100 to almost 800. If we want to increase the accuracy to 1 kg N_2O-N/ha, the range of total sample size will be from almost 200 to 4000.

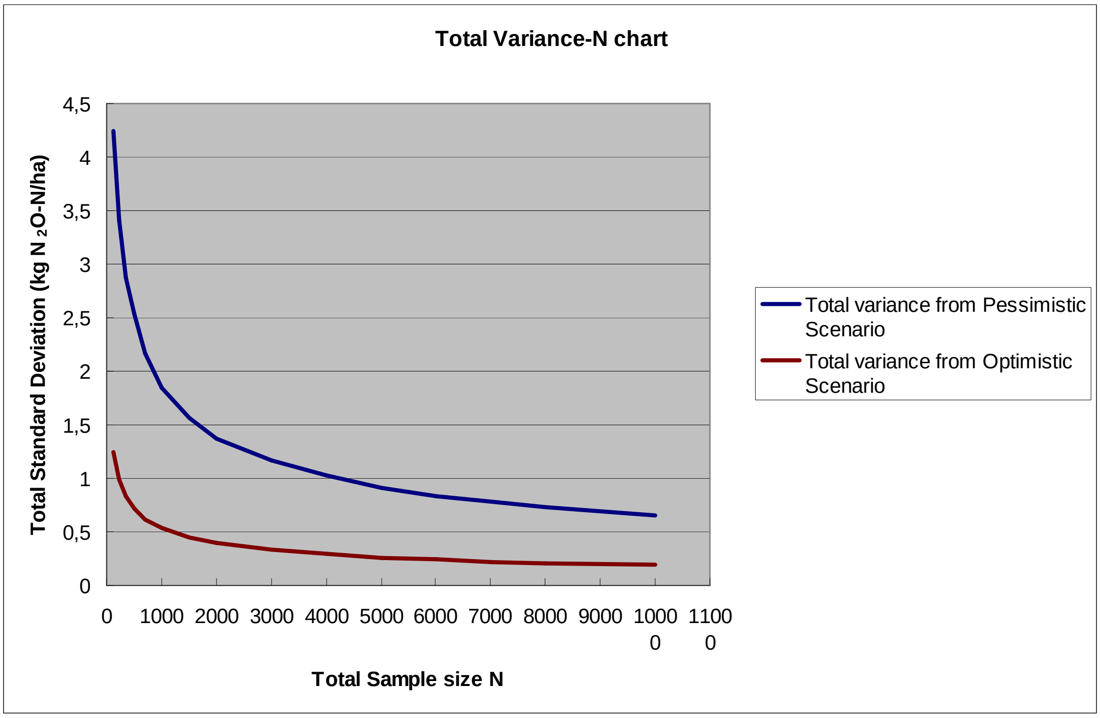


Figure 3.31, Total variance calculated for pessimistic scenario and optimistic scenario.

3.4.3. Trend of total variance

According to Eq.2.21, we can calculate the trend R for the curves in figure 3.30. We plot the rate Ri at its total sample size in figure 3.31. The value R represents the decreasing rate of total variance at a certain sample size, compared to the total variance at a total sample size of 100. It removes the influence of different scenarios and shows the efficiency of accuracy increase by raising the total sample size.

From figure 3.32, we can see that at certain sample size, the rate Ri of variance calculated from pessimistic scenario is almost the same with that calculated from optimistic scenario. They have a sharp decrease when the total sample size is smaller than 1000. This means that with the increase of sample, the decrease rate of the total variance is very fast. When the total sample size increases to more than 2000, the decrease rate of the total variance would not change much. This means that by increasing the sample size, the total variance would not be reduced significantly.

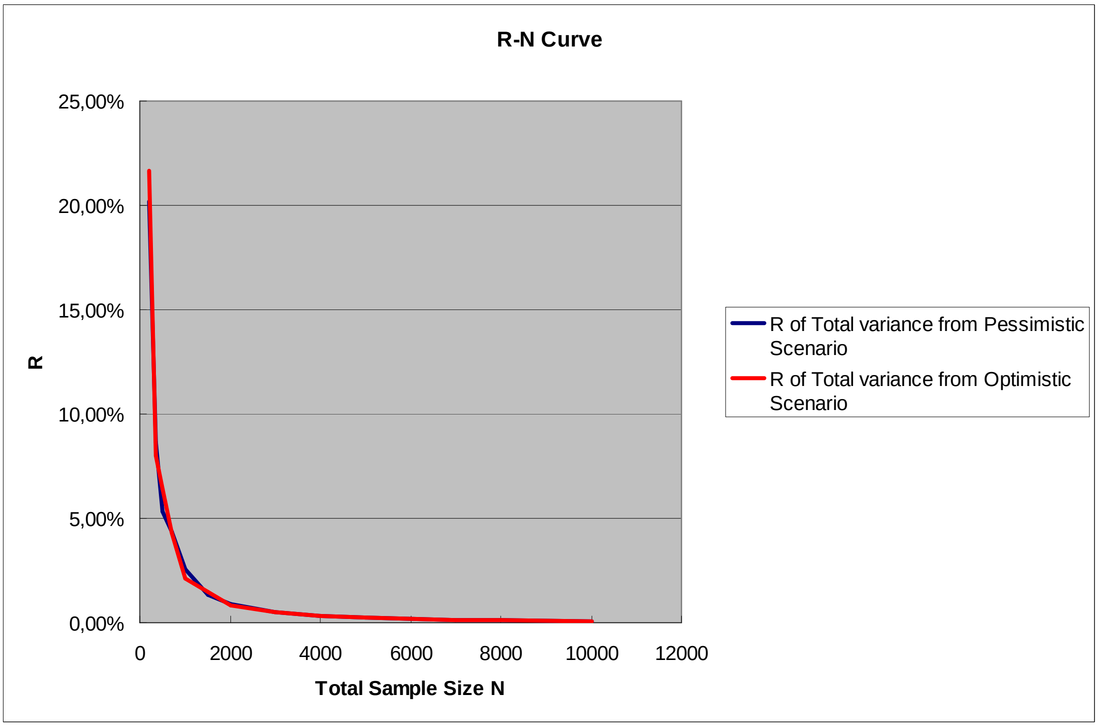


Figure 3.32, , the curve of decreasing rate at each sample size for pessimistic and optimistic scenario.

Because the R-N curves of pessimistic and optimistic scenario are the same, we can obtain an R-N function for all the scenario with the same designed strata by regression, which is

$$R = f(N) \tag{3.3}$$

Replace the R in Eq.2.21, we can get:

$$f(N_i) = -\frac{V_i - V_{i-1}}{V_1 \times (N_i - N_{i-1})} \tag{3.4}$$

Because $\partial V = V_i - V_{i-1}$ and $\partial N = N_i - N_{i-1}$, Eq.3.4 can be changed as:

$$-V_1 \times f(N) \times \partial N = \partial V \tag{3.5}$$

We assume that V_1 is a constant number in Eq.2.5, which can be obtained from field measurement. Calculating the integration for both sides of Eq.3.5, we obtained that

$$-V_1 \int_{N=N_1}^{N_i} f(N) = V_1 - V_i \tag{3.6}$$

We can obtain function F (N) with

$$\int_{N=0}^{N} f(N) = F(N) - F(0) \tag{3.7}$$

Eq.3.6 can be changed as

$$-V_1[F(N_i) - F(N_1)] = V_1 - V_i \tag{3.9}$$

So,

$$F(N_i) = F(N_1) - \frac{V_1 - V_i}{V_1} \qquad (3.10)$$

If we know the N_1, V_1 and V_i, using Eq.3.10, we can calculate the Ni by the R-N function R=f (N) and its integration function F (N), which we obtained in figure 3.31.

4. Discussion

4.1. The complexity and number of spatial factors

Due to the complexity of the biogeochemical processes of N_2O production, it is difficult to fully model the N_2O emission from soil. Researchers use field measurement data to study the environment factors influencing N_2O emission.

The spatial factors controlling annual N_2O emission at continental scale were analyzed from the influencing factors of N_2O emission at field scale. The spatial factors – land use type, annual N input, soil type and climate region, are widely accepted as important factors controlling annual N_2O emission in N_2O research. However, there are also other important factors influencing N_2O emission, such as management style, the type of nitrogen fertilizer, ground water level and soil pH. Because the corresponding spatial data of these factors are difficult to obtain for the EU27 countries, we had to ignore their influence in this research. The ignorance of some spatial factors simplifies the stratification of study area, but also enlarges the variability of N_2O emission in the strata we created, because the uncertainty of these spatial factors is not taken into the account when we made the sample strategy.

The number of spatial factors directly affects the complexity of stratification. The more spatial factors we had used, the more features we obtained after combining their corresponding spatial data. Increasing the number of spatial factors will make the stratification of these features difficult, because there are more factors considered. The number of strata will be raised with the increase of features. This causes the high risk that there are many strata that have very small sample size, because of the small area they covers in the study area. From figure 3.30, we can see that most of strata that we created in this research take up very small percentage of the total sample size. In table 3.7, we can find that these strata all have small area so that the stratum weight w_i becomes very small. This causes the small sample size they have. The increase of the number of spatial factors can make this situation more serious, because the mean percentage of area taken up by each stratum will be decreased if the number of strata increased.

From the result of stratification using these four spatial factors, we can conclude that it would not always true that the more spatial factors we had used, the nicer result of stratification we obtained. The number of spatial factor used to create the strata need be controlled so that the number of strata would not be too big.

4.2. The Definition of strata

After choosing the spatial factors, the definition of strata is determined by value of dissimilarity tables and the number of strata.

In this research, we reclassify the features of the result of combining four spatial factors to the strata by agglomerative hierarchical clustering. There are two reasons for clustering. The first one is to smooth the large variability of area of features. Even though we clustered all 114 features into 17 strata, the variability of strata area is still large. The second reason is to make the questionnaire simple.

4.2.1. The value of dissimilarity tables

There is no enough measurement data to quantify the difference between the classes of each spatial factor. All the values in the dissimilarity tables of spatial factors only tell the rough difference. We determined them by our understanding of the spatial factors. The largest difference and the smallest difference between the classes of each spatial factor are emphasized in the dissimilarity table. In all dissimilarity tables of the spatial factors, the table value is above 0.5 in order to limit the range of dissimilarity.

The dissimilarity tables of spatial factors are created to calculate the dissimilarity table of the features. By testing, we find that the weight of each spatial factor is more important than the value in dissimilarity tables of the spatial factors, when we calculated the dissimilarity table of the features. Because the weight we give to each spatial factor determines the contribution that this spatial factor has in the dissimilarity table of the features. For instance, the weight of annual N input is set as 0.5. The contribution of annual N input for each value in the dissimilarity table of features range from 0 to 0.5.

The importance of spatial factors for N_2O emission is easier to obtain than the difference between classes of each spatial factor. Land use type is the major influencing factor of N_2O emission, so according to the land use types, we select three sets of features – Arable {}, Pasture {} and Nature {}. But there is no field measurement data we can use to quantify the weights of the remaining three spatial factors either. We tested the agglomerative coefficient (AC) of clustering the features calculated from different composition of weights. AC is calculated according to Eq.2.15. The weights in table 3.4 and table 3.5 are accepted.

4.2.2. The number of strata

We set the range of the number of strata from 10 to 20. By testing the different number of strata, we find that if the number of strata was less than 10, the composition of each stratum would be too complex to interpret by experts; if the number of strata was more than 20, the area of some strata would be too small.

We determined the size of clusters for each set of features according to the clustering tree. The function "agnes" in software R could not evaluate the agglomerative efficient (AC) with a certain number of clusters. Though we can find the number of clusters from the clustering trees in figure 3.9 and 3.10, the height in the figures is meaning less, just for plotting. So we have to evaluate the result of clustering in ArcGIS. Luckily, the result of 7 clusters for Arable {} and Pasture {} is appropriate in this research. The result of 3 clusters for Nature {} is also accepted.

4.3. The confidence of experts' answers

Due to the lack of measurement data, we consult experts to guess range of the annual N2O emission for each stratum. It is also a difficult question for all experts, because of the diverse compositions each stratum has. That is why the high variability is shown from figure 3.11 to 3.27. Also due to the diverse compositions each stratum has, the distribution of mean annual N2O emission of each stratum could be normal or lognormal, though most of field measurement shows it should be normal.

The confident of expert's answers is not taken into account in this research, because of the large variability they show. Some experts would tell us they are not so sure about the answers of some strata. But they may be quite sure about the answers of other strata. This makes it difficult to determine which answers should be included and which should not.

That is why we finally put all the answers from experts and calculate the max distance and mean distance for each stratum. This alternative way has the risk of over estimation of the mean annual N_2O emission for each stratum. This is acceptable, because we only evaluate the efficiency of the stratified random sampling methods. If this method could significantly reduce the variance at a situation of over estimation, why not use it?

4.4. The total variance

There are two interesting parts discussed here. The first one is the percentage of the total sample size taken by each stratum. The second one is the variability of total variance with the increasing total sample size.

4.4.1. The percentage of stratum sample size

At a certain total sample size, the optimal stratum sample size is determined by the percentage of $w_i^2 * s_i^2 / \sum w_i^2 * s_i^2$. From the proving of Eq.2.6, we can find that the optimal allocation of stratum sample size equalizes the contribution of each stratum to the total variance. So the percentage of stratum sample size – $w_i^2 * s_i^2 / \sum w_i^2 * s_i^2$ is the key issue during calculating the total variance.

From figure 3.30, we can see that the stratum1, stratum 2, and stratum 16 occupy almost 90 % of the total sample size and the percentage of sample size for stratum 16 is the largest. However, in table 3.9 and table 3.10, we can see these strata do not have the largest standard deviation in both pessimistic scenario and optimistic scenario. In this research, the major factor controlling the allocation of sample size is the stratum weight w_i. The stratum that has the largest w_i takes up the largest stratum sample size. The reason is that the strata designed in this research have large variability of area. Again, this proves the importance of clustering.

Otherwise, the variability of stratum area would be much larger.

4.4.2. The variability of total variance with the increasing total sample size

At the total sample size of 100, the total variance calculated from pessimistic scenario is 4.24 kg N_2O-N/ha and that from optimistic scenario is 1.24 kg N_2O-N/ha. The variance of estimation of mean annual N_2O emission by stratification random sampling way could be in the range from 1.24 kg N_2O-N/ha to 4.24 kg N_2O-N/ha.

The mean annual N_2O emission from cultivated area in the EU 15 countries in 1996 is 5.6 N_2O -N /ha (Pascal & Oswarld, 2001). According to Lim et al. [1999], the uncertainty of estimation of N_2O emission by IPCC 1997 ranges 70% to 100% of the N_2O emission. So the uncertainty of N_2O emission from EU 15 could be ranging from 3.92 kg N_2O-N/ha to 5.6 kg N_2O-N/ha. Over estimated uncertainty of IPCC 1997 is larger than the variance by stratification random sampling calculated from the pessimistic scenario at the total sample size of 100. But the lowest variance of estimation by stratified random sampling is smaller than that by IPCC 1997.

If we increase the sample size to 200, the range of total variance by stratification random sampling will be from 0.98 kg N/ha to 3.41 kg N/ha. This range is much lower than that using IPCC 1997. Increasing sample size will reduce the total variance. Using stratification random sampling methods, it could control the uncertainty of N2O emission. This is very attractive for political decision making. By using 200 sample points, we could get a more accuracy estimation than that by IPCC 1997.

From figure 3.32, we can see that for the strata designed in this research, the variability of total variance with the increasing total sample size shows a constant trend, no matter the total variance is calculated from pessimistic scenario or optimistic scenario.

With the development of model-based methods, its accuracy will also be improved. We suppose that the stratified random sampling with the same designed strata as this research has already implemented in the EU 27. We obtained the total variance V_1 at the total sample size N_1 = 200, which is smaller than the uncertainty of IPCC 1997 method. Some years later, a new method of IPCC 2010 is applied and its uncertainty is reduced to V_i. Because the stratified random sampling has the same designed strata with this research, the R-N function f (N) is the same for all the scenario of strata standard deviations, which is already known in figure 3.32. Using Eq.3.10, We can very easily obtain the total sample size N_i, more than which the total variance of estimation from the stratified random sampling is smaller than the new method of IPCC 2010. Using this way, we can easily compare the efficient of the stratified random sampling and the new model-based method

5. Conclusion

1. By referring to the statistic books, we answered the first research question – *What are the theory and the practice of stratified random sampling to estimate nitrous-oxide gas emission in EU 27?*

The mean annual N_2O emission from agricultural and natural area in 27 countries in European Union could be calculated from the mean of annual N_2O emission, $\overline{y}_i$, from sample points in each stratum as below:

$$\overline{y}_{st} = \sum_{i=1}^{k} w_i \times \overline{y}_i$$

With the total variance of $\overline{y}_{st}$

$$Var\left(\overline{y}_{st}\right) = \sum_{i=1}^{k} \frac{w_i^2 \times s_i^2}{n_i}$$

$$w_i = \frac{Area_i}{Area_t}$$

$Area_i$ is the area of *i*th stratum; $Area_t$ is the total area of Europe. We call w_i as the stratum weight.

The optimal sample size for each stratum n_i, is calculated as below:

$$n_i = \frac{w_i^2 \times s_i^2}{\sum_{i=1}^{k} w_i^2 \times s_i^2} \times n$$

2. By literature studying and analyzing the spatial factor at the field scale, we obtained the conclusion of research question 2 –*What are important spatial factors controlling annual nitrous-oxide gas emissions from natural and agricultural land?*

The important factors controlling annual N_2O emission from agricultural and natural land are land use type, annual N input, soil type and climate region.

There are three classes in land use type, which are arable land, pasture land, and nature land.

The annual N input is categorized into 5 classes, which are 0 – 50, 50 – 100, 100 – 150, 150 – 200, and > 200, with the unit kg N /ha. The soil type has three classes, which are peat, clay and sand & rock. We use 5 climate regions in this research, which are Alpine, Atlantic, Boreal, Mediterranean and Continental.

3. Using agglomerative hierarchical clustering to the features after combining four spatial data, we obtained the conclusion of research question 3 – *How to process the spatial data to create strata, which are designed for N_2O emission?*

All spatial data are reclassified, projected and converted to raster data. To emphasize the importance of land use type, we group the features with the same land use type into 3 sets, after combining all spatial data. We denote them as Arable {}, Pasture {} and Nature {}.

Dissimilarity table of each spatial factor except land use type is made, denoted as S[i, j], N [i, j] and C [i, j]. According to the dissimilarity table of each spatial factor and its weight, the dissimilarity table of features in Arable {} and Pasture {} is calculated as below:

$$A[p,q] = 0.5 \times N[i,j] + 0.3 \times S[i,j] + 0.2 \times C[i,j]$$

We agglomerate Arable {} and Pasture {} into 7 clusters according to the dissimilarity table of strata.

The dissimilarity table of features in Nature is calculated as below:

$$A[p,q] = 0.6 \times S[i,j] + 0.4 \times C[i,j]$$

Nature {} is grouped into 3 clusters.

The result of reclassification strata is sequenced in table below:

	Stratum in reclassification table of combination
Arable land	stratum 1 - stratum 7
Pasture land	stratum 8 - stratum 14
Nature land	stratum 14 - stratum 17

According to the reclassification table of cluster, we can group the result of combination into 17 strata, which we design for N2O emission research in EU 27.

4. After creating the strata, we make a questionnaire to obtain the conclusion of Research question 4 – *What are the annual within strata variances for nitrous-oxide by consulting expert?*

After consulting the experts by questionnaire, experts give their guess for Quartile 1, Median and Quartile 3 of annual N2O emission for each stratum. We calculate pessimistic scenario and optimistic scenario of standard deviation from max distance and mean distance respectively as below:

Max distance = max (Q3) – min (Q1)

Mean distance = mean (Q3) – mean (Q1)

Si = Distance/1.34

Q1 is the answer of Quartile 1 in the questionnaire;

Q3 is the answer of Quartile 3 in the questionnaire;

Distance is the max distance or mean distance.

5. Using the pessimistic scenario and optimistic scenario of stratum weight and standard deviation, we answered the research question 5 – *What is the total variance at certain total sample size?*

The pessimistic scenario of standard deviation is calculated from the max distance from the experts' answers. The optimistic scenario of standard deviation is calculated from the mean distance from the experts' answers. The stratum weight is calculated from the stratum area. Total variance at a certain total sample size could be calculated from the stratum weight w_i, stratum standard deviation s_i, and the sample size n_i for *i*th stratum. At the total sample size of 100, the total variance calculated from pessimistic scenario is 4.24 kg N_2O-N/ha and that from optimistic scenario is 1.24 kg N_2O-N/ha.

We can calculate the decreasing rate R_i for each pair of the total sample size N_i and its corresponding total variance V_i as below:

$$R_i = -\frac{V_i - V_{i-1}}{V_1 \times \left(N_i - N_{i-1}\right)} \times 100\%$$

We can plot the R-N curve, and obtain an R-N function R=f (N) by regression from the R-N curve.

If we know the total variance V_1 at the total sample size N_1, when we want to the total sample size N_i with the total variance V_i. We can use the equation below

$$F(N_i) = F(N_1) - \frac{V_1 - V_i}{V_1}$$

F(N) is the integration function of R = f (N).

Finally we achieve the objective of this research – *to asses the accuracy of design-based estimation of mean annual nitrous-oxide emission by natural and agricultural land in EU 27 by applying stratified random sampling.*

With the sample size 200, the range of total variance by stratified random sampling will be from 0.98 kg N/ha to 3.41 kg N/ha. At the sample size of 200, the accuracy of estimation by this method is already better than that by IPCC 1997. In this research, the accuracy of estimation of mean annual N2O emission by stratification random sampling method will increase largely with the increase of the total sample size.

For the strata designed in this research, the variability of total variance with the increasing total sample size shows a constant trend, no matter the total variance is calculated from pessimistic scenario or optimistic scenario. This characteristic is shown as the decreasing rate R_i. From the regression of R-N curve, we can obtain a R-N function R=f (N). Using this function we can simply calculate the total sample size at a certain total variance if we know the total variance at one total sample size.

6. Reference

Buchanan, R.E., 1917. Studies on the nomenclature and classification of the bacteria Ⅲ the families of the *Eubacteriales*. *Journal of Bacteriology* 2, 347-350.

Barnett, V.,1993. Sample survey principles and methods, p105-139. British Library Cataloguing in Publication Data, Great Britain.

Catherine A.S., and Gareth M.J., 2003. Finding the number of clusters in a data set: An information theoretic approach. Journal of the American Statistical Association 98, 750 – 763.

Clayton, H., I.P., McTaggart, J., Parker, L., Swan, K.A., Smith, 1997. Nitrous oxide emissions from fertilised grassland: A 2-year study of the effects of N fertiliser form and environmental conditions. *Biol Fertil Soils* 25, 252-260.

Coyne, M.S. (2009). Biological Denitrification page 207. Nitrogen in agricultural Systems, editor:J.S. Schepers and W.R. Raun, Agronomy Monograph 49, USA, 2009.
Hardy, G., J.E., Littlewood, G.., Polya, 1999. Inequalities. Cambridge Mathematical Library, Cambridge University Press.

Dobbie, K.E., I.P, McTaggart, K.A., Smith, 1999. Nitrous oxide emissions from intensive agricultural systems: Varations between crops and seasons, key driving variables, and mean emission factors. *Journal of Geophysical Research* 104, 26,891-26, 899.

Groffman, P.M., R., Brumme, K. Butterbach-Bahl, K.E., Dobbie, A.R., Moiser, D., Ojima, H., Papen, W.J., Parton, K.A., Smith, C., Wagner-Riddle, 2000. Evaluating annual nitrous oxide fluxes at the ecosystem scale. *Global Biogeochemical Cycles*, 14, 1061-1070.

Hansen, H.M., W.N., Hurwitz, W.G., Madow, 1993. Sample survey Methods and Theory ,volum 1 Methods and applications, p 179-234. John Wiley & Sons, Inc, New York.

Hochestein, L.I., Tomlinson, G.A., 1988. The enzymes sddociated with denitrification. *Annual Review of Microbiology* 42, 231-261.

Hynes, R.K., Knowles, R., 1984. Production of nitrous oxide by *Nitroso-monas europaea*: effects of acetylene, pH, and oxygen. *Canadian Journal of Microbiology* 30, 1397-1404.

Kaufman, L. and Rousseeuw, P.J., 1990. Finding Groups in Data: An Introduction to Cluster Analysis. Wiley, New York.

Leffelaar, P.A., 1986. Dynamics of partial anaerobiosis, denitrification, and water in a soil aggregate: experimental. *Soil Science* 142, 352-366.

Levy, P.S., and Lemeshow, S., 2008. Sampling of Populations Methods and Applications, p121-142. John Wiley & Sons, Inc, New York.

Lieti R, M.C., Ortiz, L.A, Sarabia., M.S., Sanchez, 2004. Seletcting Variabels for k-Means cluster Analysis by Using a Genetic algorithm that Optimises the Sihouettes. Analytica Chimica Acta 515, 87 – 100.

Lim, B., Boileu P., Bonduki Y., Van Amstel, A.R., Janssen L.H.J.M>, Olivier J.G.J., Kroeze C., 1999. Improving the quality of national greenhouse gs inventories. *Enviromental Science & Policy* 2, 335-346.

Matern, B., 1986. Spatial Variation, 2nd ed. p155. Springer-Verlag, New York,

Mooney, H. A., P.M., Vitousek, and P. A., Maston, 1987. Exchange of materials between terrestrial ecosystems and the atmosphere. *Science*, 238, 926-932.

Olea, R. A., 1984. Sampling design optimization for spatial functions. *Math* 16, 369-392,

Pascal B. & Oswald V.C., 2001. Estimates of N2O and CH4 fluxes from agricultural lands in various regions in Europe. *Nutrient Cycling in Agroecosystems* 60, 35-47.

Poth, M., D.D., Focht, 1985. ^{15}N kinetic analysis of N_2O production by Nitrosomonas europaea: an examination of nitrifier denitrification. *Applied & Environmental Microbiology* 49, 1134-1141.

Prather, M., R. Derwent, D. Thhalt, P. Fraser, E. Sanhueza, and X. Zhou, Other trace gases and atmospheric chemistry, in Climate Change 1994: Radiative Forcing of Climate Changes and an Evaluation of the IPCC IS92 Emission Scenarios, edited by J. Houghton, et al., pp, 77-126, Camridge Univ. Press, New York, 1995.

Schimel, D.S., B.H., Braswell, W., Emanuel, B., Rizzo, T., Smith, F.I., Woodward, H., Fisher, T.G.F., Kittel, R., McKeown, T., Painter, N., Rosenbloom, D.S., Ojima, W.J., Parton, D.W., Kicklighter, A.D., McGuire, J.M., Melillo, Y., Pan, A., Haxeltine, C., Prentice, S., Sitch, K., Hibbard, R., Nemani, L., Pierce, S., Running, J., Borchers, J., Chaney, R Neilson, 1997. Continental scale variability in ecosystem process: models, data, and the role of disturbance. Ecological Monograph 67, 251-271.

Tiedje, J.M., Sexstone, A.J., Parkin, T.B., revsbech, N.P., 1984. Anaorobic processes in soil. *Plant and Soil* 76, 197-212.

Thompson, T.K., 2002. Sampling, p117-128. John Wiley & Sons, Inc, New York.

Velthof, G.L.,1997. Nitrous oxide emission from intensively managed grasslands. Doctoral thesis, Wageningen Agricultural university, the Netherlands.

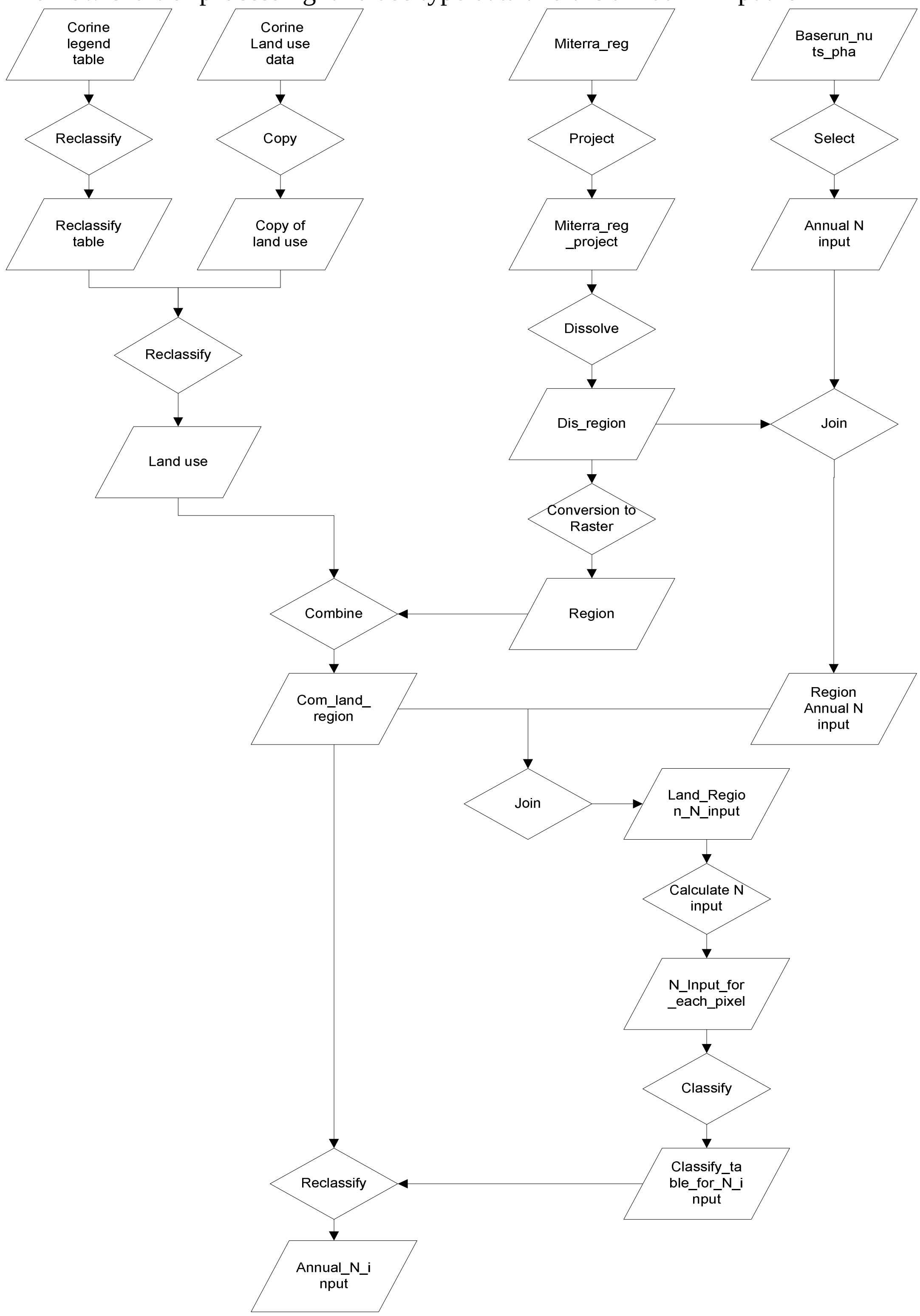
Velthof, G.L., Oudendag, D.A., Oenema, O., 2007. Development and application of the integrated nitrogen model MITERRA-EUROPE. Alterra-report 1663.1, Wageningen.

Wrage N., Velthof G.L., Van Beusichem M.L., Oenema O., 2001. Role nitrifier denitrification in the production of nitrous oxide. *Soil Biology & Biochemistry*, 33 1723-1732.

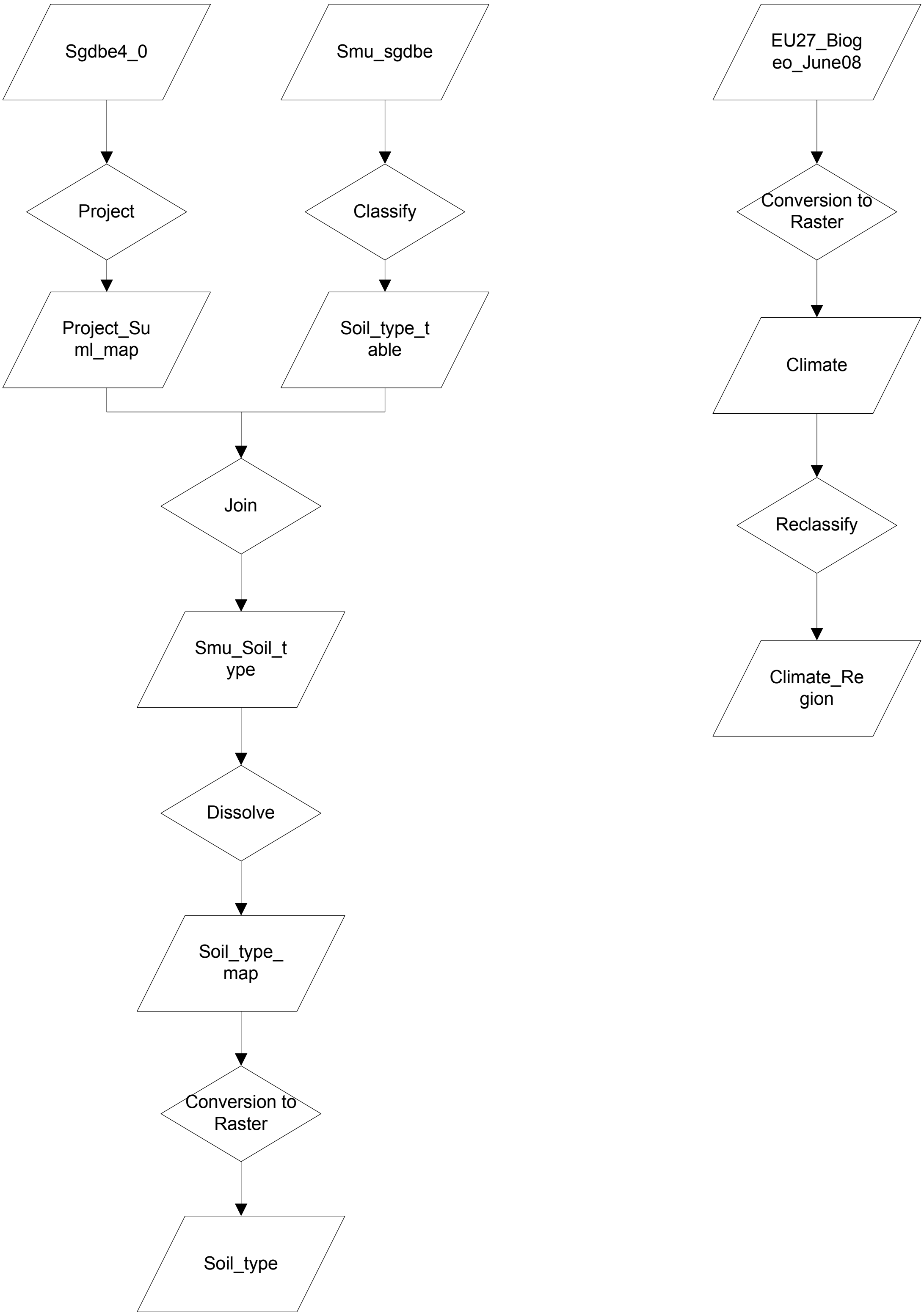
Wood, P.M., 1986. Nitrification as a bacterial energy source. In: Prosser, J.I.(Ed.), Nitrification. Special Publications of the Society for General Microbiology 20, 39-62.

7. Appendix A

The flow chart of processing land use type data and the annual N input is



The flow chart of processing soil type data and climate region data is shown below:



8. Appendix B

Questionnaire for the annual N2O emission in EU27

By Haolu shang

Supervisors: dr. ir. Sytze de Bruin
dr. ir. Gerard Heuvelink

Geo-Information Science and Remote Sensing
Wageningen University

In my MSc thesis research, I am investigating a sampling method for estimation of the mean annual N2O emission from agriculture and nature areas in EU-27 countries.

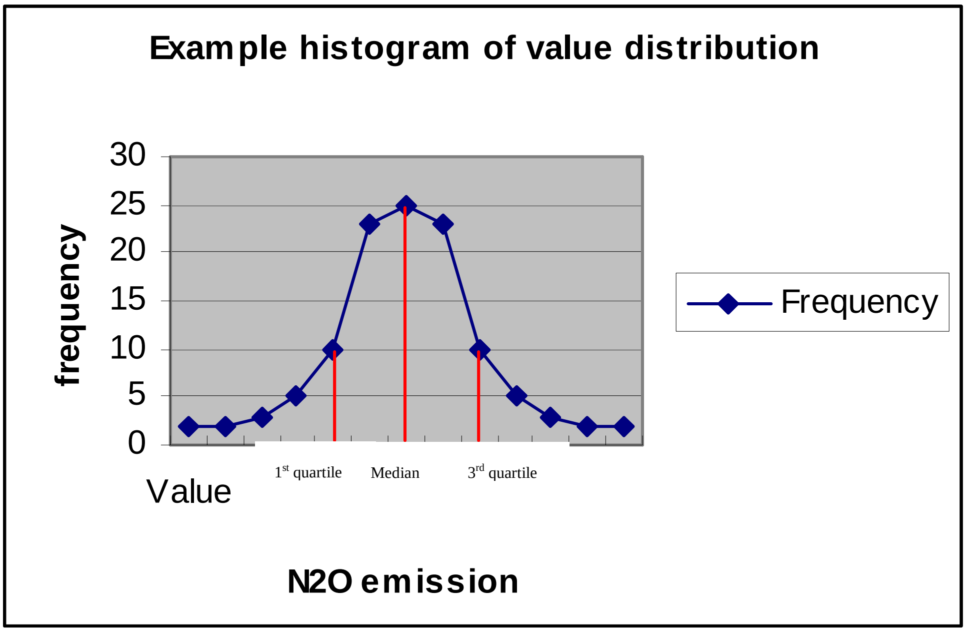
I propose a stratified random sampling method to estimate the mean of annual N2O emission. The basic idea of this method is that we stratify the area according to the control factors of annual N2O emission. The variability of annual N2O emission between each stratum is larger than that within each stratum. The purpose of this research is to evaluate the accuracy of this method.

In each stratum, I randomly choose some sample locations. Annual N2O emission is measured at each location with five chambers – four of them making a 100m×100m square and one in the center. We assume continuous measurements over the entire year. To calculate the variance of annual N2O emission from the whole area, the variance in each stratum needs to be estimated from the annual N2O emissions from the sample points. For lack of field measurement data, we consult experts to estimate the variability of the annual N2O emission in each stratum.

After consulting the literatures and experts at WUR, I chose four spatial factors for controlling N2O emission in EU27. These are land use type, soil type, N application and climate region. These factors are combined to stratify the study area. Due to the limited information, we classified the whole area into three land use types, which are natural land, agricultural land, and pasture land. The map is shown in figure 1. Built-up area has been removed from this map. The soil type map, in figure 2 and climate region map, figure 3, for EU-27 are shown below. In figure 4, we present the N application strata as derived from Nuts 2 Census data.

An overlay of the four factors gives 129 existing combinations. I used a clustering approach to group similar combinations(judged on the basis of a dissimilarity matrix) and reduce the number of classes to 17. The final map is shown in figure 5

For each stratum, I present an attribute table and pie charts with compositional information.
As described above, a location is chosen at random within a stratum and the annual N2O emission is measured by taking the average of the continuous measurements at five points within the 100m×100m square (the center of square is at this location). For a stratum has different compositions and N2O emission shows spatial variability, the annual N2O emission will not be the same in all locations within a stratum. Instead, it will be a distribution of value, of which one example histogram is shown below.



The questionnaire table presents the cumulative frequency table of annual N2O emission cut off by quartiles. The first quartile should be that value for which you believe that there is 25% probability that the actual emission will be smaller. The third quartile should be that value for which there is 25% probability that the actual emission will be greater. I kindly ask you enter your best guesses for the quartiles of N2O emissions in the bottom questionnaire table on each sheet.

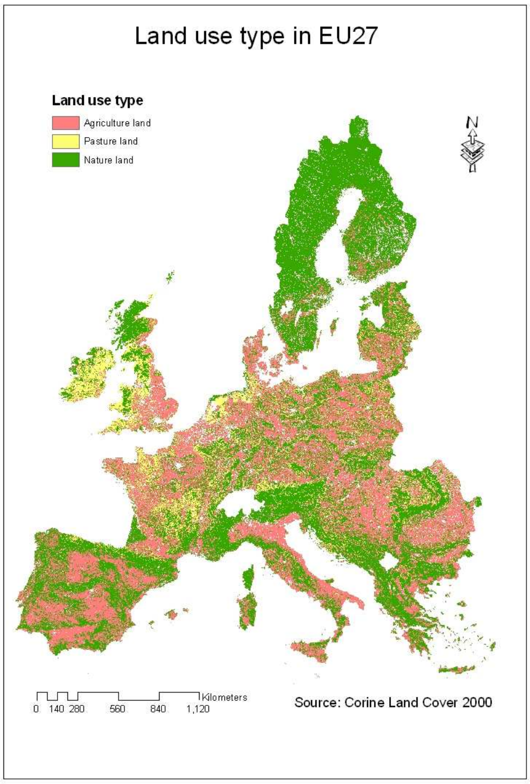


Figure 1: land use type in EU 27

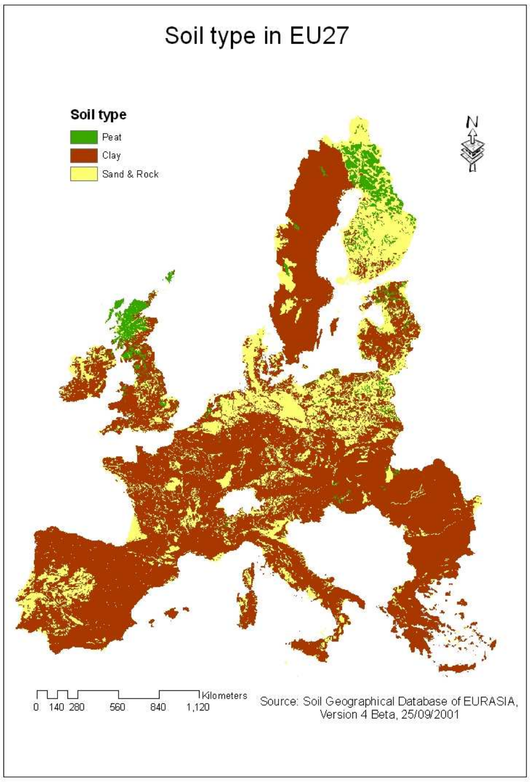


Figure 2: Soil type in EU 27

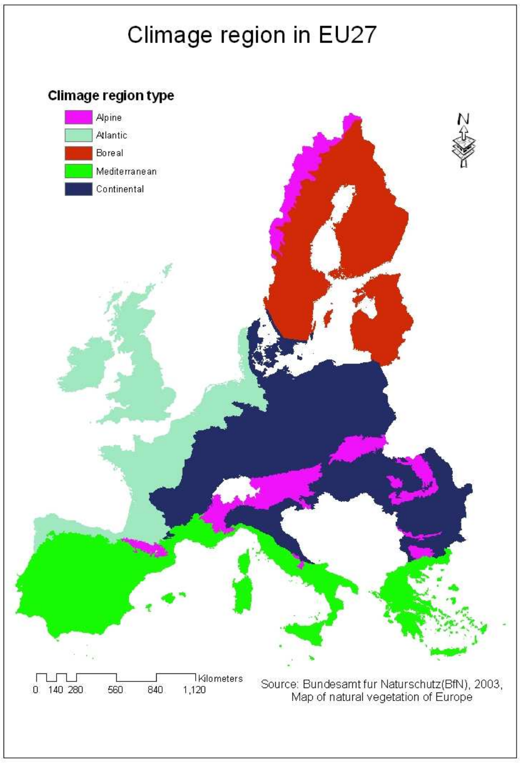


Figure 3: Climate region in EU 27

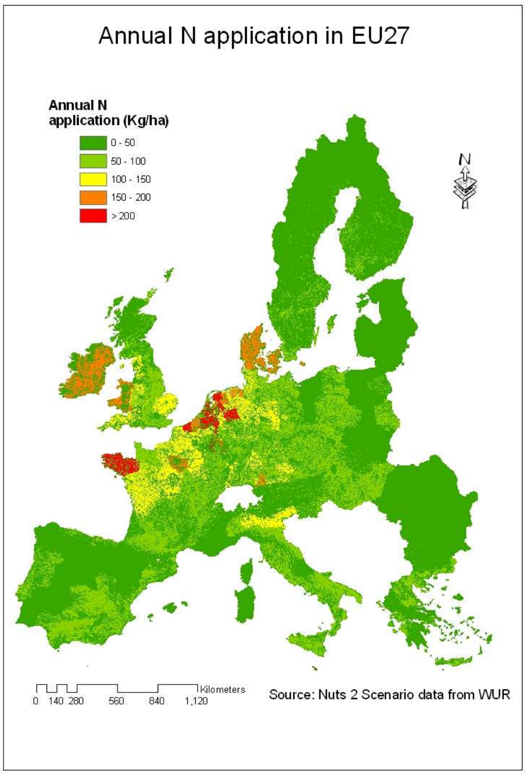


Figure 4: Annual N application in EU 27

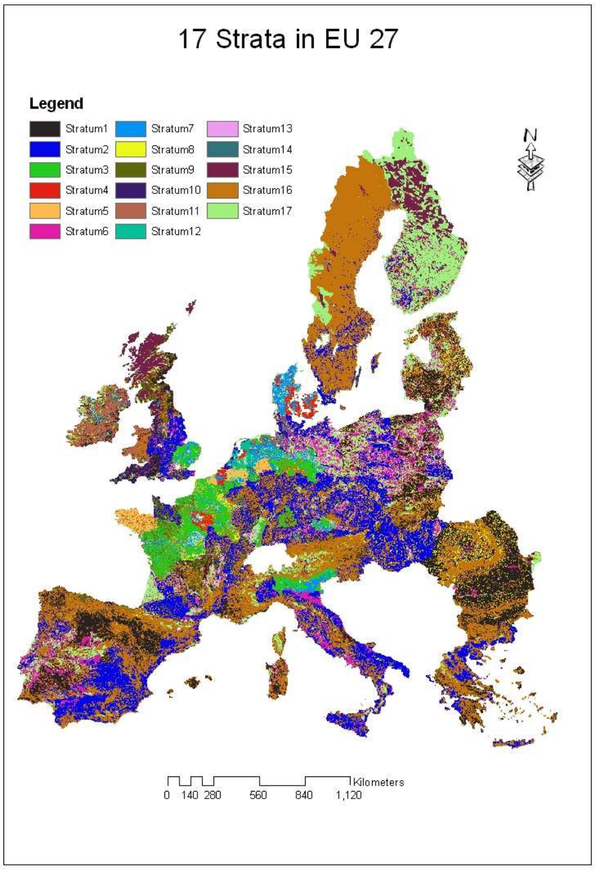


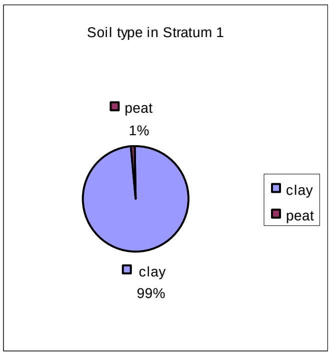
Figure 5: 17 strata for N2O emission in EU 27

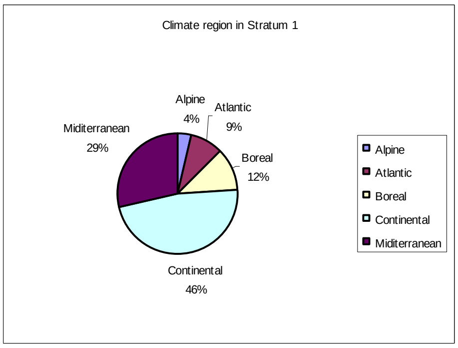
Stratum 1

The attribute table of stratum 1 is shown in table 1. Two pie charts show the statistical information of the soil type and climate region in stratum 1 respectively. The figure 6 shows the area stratum1 covers in EU27. The questionnaire is in table2.

Table 1, attribute table of stratum 1.

Class	Area (km²)	Land type	Soil type	N input(kg/ha*year)	Climate region
1	1,8E+04	agriculture	clay	0 - 50	Alpine
2	4,2E+04	agriculture	clay	0 - 50	Atlantic
3	5,4E+04	agriculture	clay	0 - 50	Boreal
4	2,2E+05	agriculture	clay	0 - 50	Continental
5	1,4E+05	agriculture	clay	0 - 50	Mediterranean
6	1,1E+01	agriculture	peat	0 - 50	Alpine
7	6,2E+02	agriculture	peat	0 - 50	Atlantic
8	2,5E+03	agriculture	peat	0 - 50	Boreal
9	3,2E+03	agriculture	peat	0 - 50	Continental
Total area (km²)		4,8E+05			





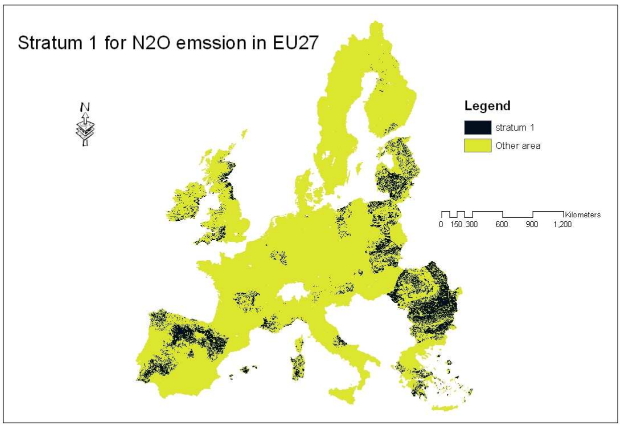


Figure 6, the area stratum1 coves in EU 27

Table 2, Questionnaire for stratum 1

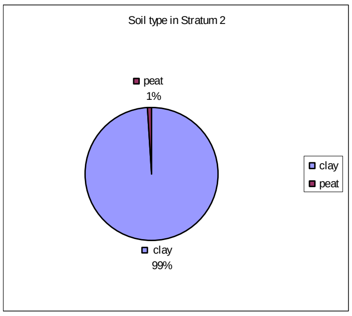
	Annual N2O emission (kg N ha^{-1} yr^{-1})
Quartile 1	
Median	
Quartile 3	

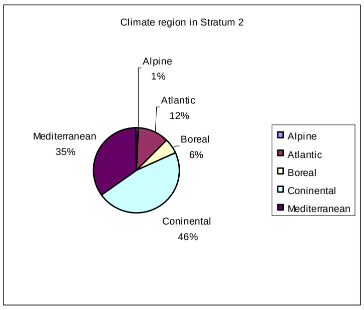
Stratum 2

The attribute table of stratum 2 is shown in table 3. Two pie charts show the statistical information of the soil type and climate region in stratum 2 respectively. The figure 7 shows the area stratum2 covers in EU27. The questionnaire is in table4.

Table 3, attribute table of stratum 2.

Class	Area (km²)	Land type	Soil type	N input(kg/ha*year)	Climate region
1	7,2E+03	agriculture	clay	50 - 100	Alpine
2	7,7E+04	agriculture	clay	50 - 100	Atlantic
3	3,5E+04	agriculture	clay	50 - 100	Boreal
4	3,0E+05	agriculture	clay	50 - 100	Continental
5	2,3E+05	agriculture	clay	50 - 100	Mediterranean
6	1,5E+00	agriculture	peat	50 - 100	Alpine
7	6,0E+02	agriculture	peat	50 - 100	Atlantic
8	1,7E+03	agriculture	peat	50 - 100	Boreal
9	4,8E+03	agriculture	peat	50 - 100	Continental
10	2,0E+01	agriculture	peat	50 - 100	Mediterranean
Total area (km²)		6,6E+05			





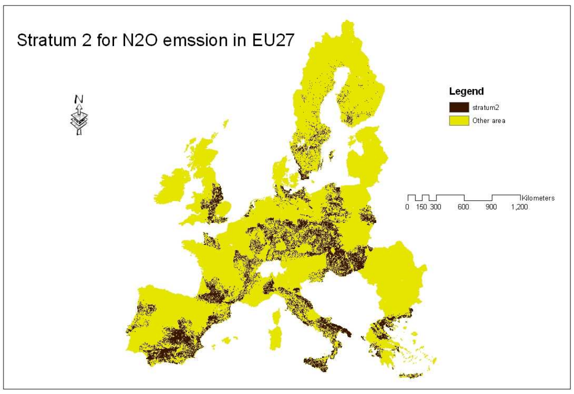


Figure 7, the area stratum2 coves in EU27.

Table 4, Questionnaire for stratum 2

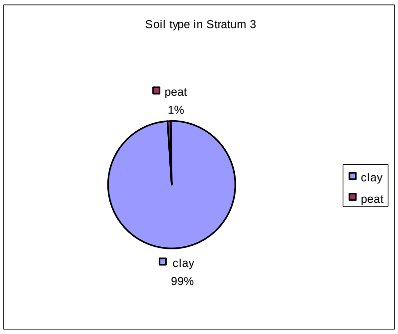
	Annual N2O emission (kg N ha^{-1} yr^{-1})
Quartile 1	
Median	
Quartile 3	

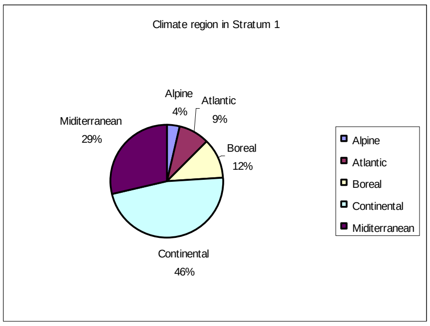
Stratum 3

The attribute table of stratum 3 is shown in table 5. Two pie charts show the statistical information of the soil type and climate region in stratum 3 respectively. The figure 8 shows the area stratum3 covers in EU27. The questionnaire is in table 6.

Table 5, attribute table of stratum3.

Class	Area (km²)	Land type	Soil type	N input(kg/ha*year)	Climate region
1	1,1E+03	agriculture	clay	100 - 150	Alpine
2	1,0E+05	agriculture	clay	100 - 150	Atlantic
3	5,4E+04	agriculture	clay	100 - 150	Continental
4	1,1E+03	agriculture	peat	100 - 150	Atlantic
5	7,8E+01	agriculture	peat	100 - 150	Continental
Total area (km²)		1,6E+05			





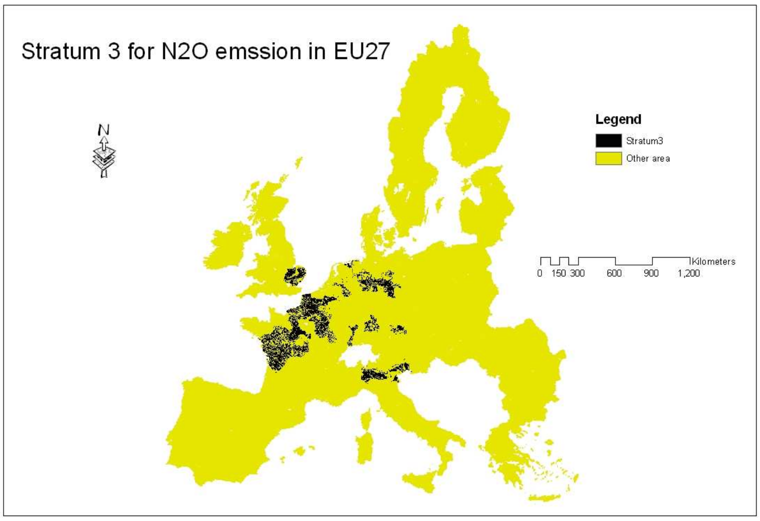


Figure 8, area stratum 3 covers in EU27

Table 6 Questionnaire for stratum 3

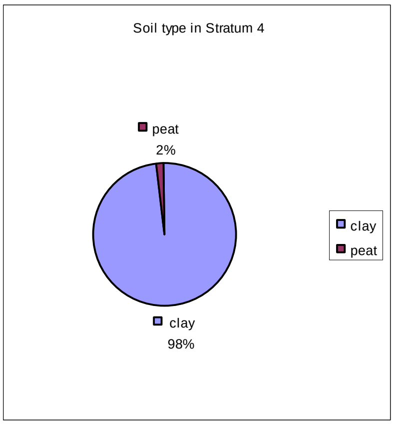
	Annual N2O emission (kg N ha^{-1} yr^{-1})
Quartile 1	
Median	
Quartile 3	

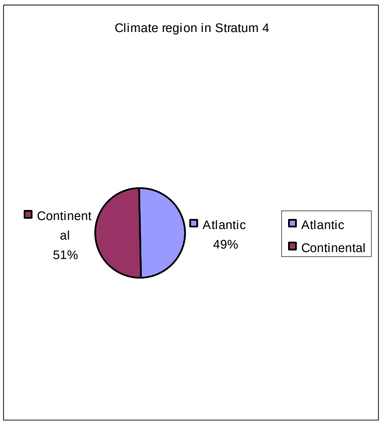
Stratum 4

The attribute table of stratum 4 is shown in table 7. Two pie charts show the statistical information of the soil type and climate region in stratum 4 respectively. The figure 9 shows the area stratum4 covers in EU27. The questionnaire is in table 8.

Table 7, attribute table of stratum 4

Class	Area (km²)	Land type	Soil type	N input(kg/ha*year)	Climate region
1	1,0E+04	agriculture	clay	150 -200	Atlantic
2	1,1E+04	agriculture	clay	150 -200	Continental
3	1,4E+02	agriculture	peat	150 -200	Atlantic
4	1,9E+02	agriculture	peat	150 -200	Continental
Total area (km²)		2,1E+04			





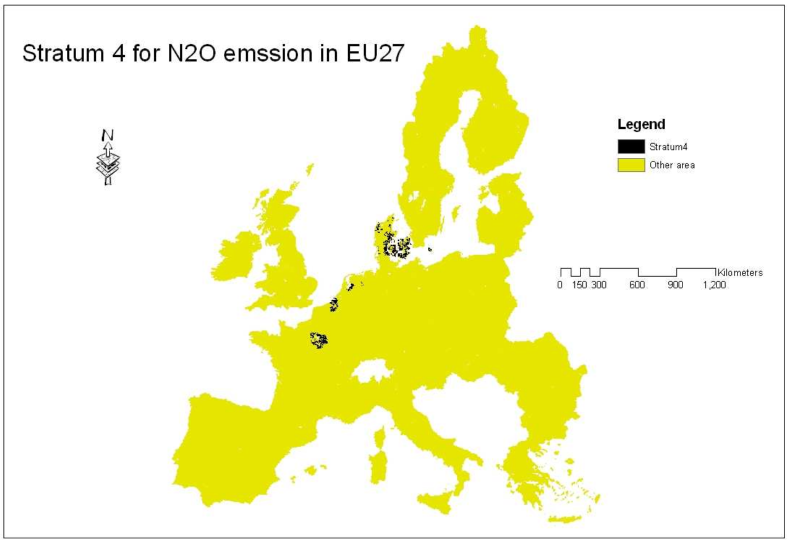


Figure 9, area stratum 4 covers in EU27.

Table 8, Questionnaire for stratum 4

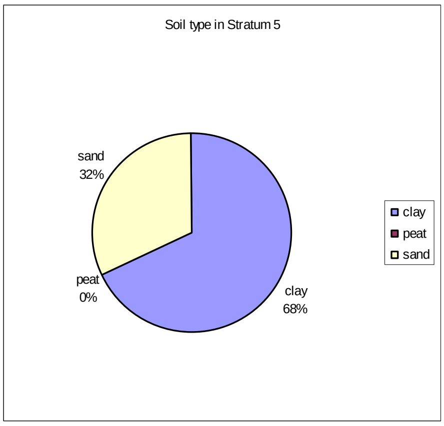
	Annual N2O emission (kg N ha^{-1} yr^{-1})
Quartile 1	
Median	
Quartile 3	

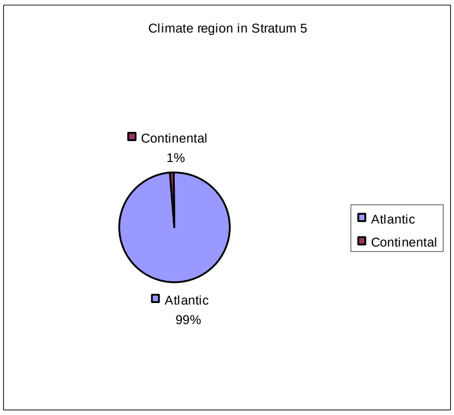
Stratum 5

The attribute table of stratum 5 is shown in table 9. Two pie charts show the statistical information of the soil type and climate region in stratum 5 respectively. The figure 10 shows the area stratum5 covers in EU27. The questionnaire is in table 10.

Table 9, attribute table of stratum 5

Class	Area (km²)	Land type	Soil type	N input(kg/ha*year)	Climate region
1	2,1E+04	agriculture	clay	>200	Atlantic
2	1,4E+02	agriculture	clay	>200	Continental
3	4,0E+01	agriculture	peat	>200	Atlantic
4	9,8E+03	agriculture	sand	>200	Atlantic
5	1,2E+02	agriculture	sand	>200	Continental
Total area (km²)		3,1E+04			





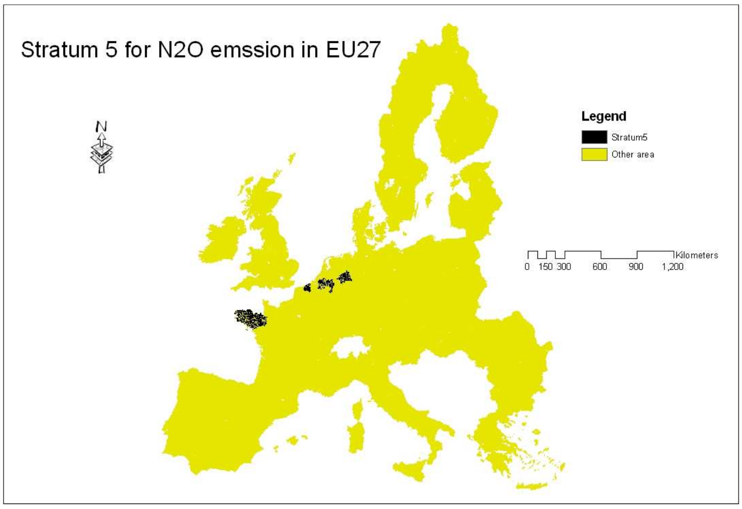


Figure 10, area stratum 5 covers in EU27

Table 10, Questionnaire for stratum 5

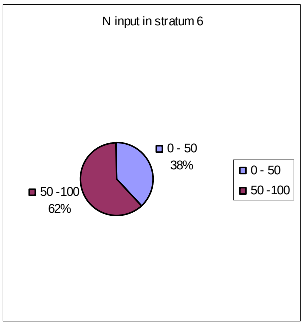
	Annual N2O emission (kg N ha^{-1} yr^{-1})
Quartile 1	
Median	
Quartile 3	

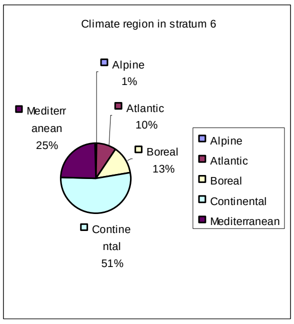
Stratum 6

The attribute table of stratum 6 is shown in table 11. Two pie charts show the statistical information of the annual N application and climate region in stratum 6 respectively. The figure 11 shows the area stratum6 covers in EU27. The questionnaire is in table 12.

Table 11, attribute table of stratum 6

Class	Area (km²)	Land type	Soil type	N input(kg/ha*year)	Climate region
1	1,4E+03	agriculture	sand	0 - 50	Alpine
2	4,8E+03	agriculture	sand	0 - 50	Atlantic
3	1,6E+04	agriculture	sand	0 - 50	Boreal
4	2,6E+04	agriculture	sand	0 - 50	Mediterranean
5	4,2E+04	agriculture	sand	0 - 50	Continental
6	3,1E+02	agriculture	sand	50 - 100	Alpine
7	1,8E+04	agriculture	sand	50 - 100	Atlantic
8	1,5E+04	agriculture	sand	50 - 100	Boreal
9	3,4E+04	agriculture	sand	50 - 100	Mediterranean
10	8,2E+04	agriculture	sand	50 - 100	Continental
Total area (km²)		2,4E+05			





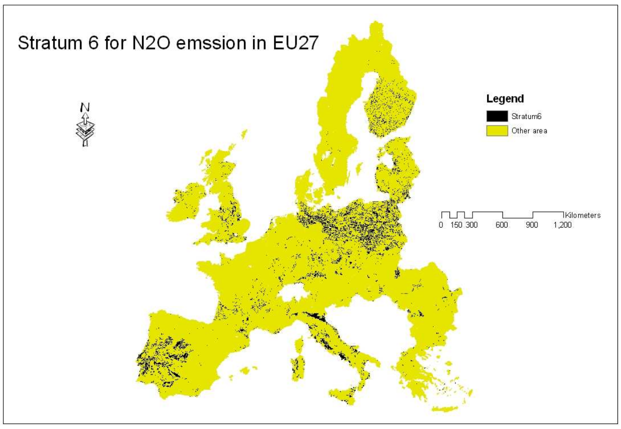


Figure 11, area stratum 6 coves in EU27

Table 12, Questionnaire for stratum 6

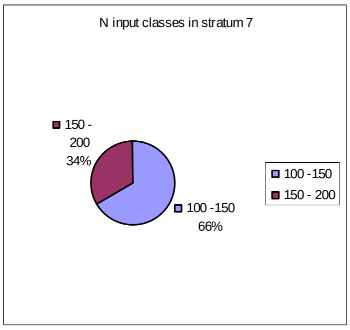
	Annual N2O emission (kg N ha^{-1} yr^{-1})
Quartile 1	
Median	
Quartile 3	

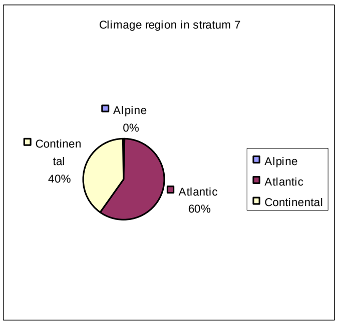
Stratum 7

The attribute table of stratum 7 is shown in table 13. Two pie charts show the statistical information of the annual N application and climate region in stratum 7 respectively. The figure 12 shows the area stratum7 covers in EU27. The questionnaire is in table 14.

Table 13, attribute table of stratum 7

Class	Area (km²)	Land type	Soil type	N input(kg/ha*year)	Climate region
1	2,9E+02	agriculture	sand	100 - 150	Alpine
2	2,8E+04	agriculture	sand	100 - 150	Atlantic
3	1,6E+04	agriculture	sand	100 - 150	Continental
4	1,2E+04	agriculture	sand	150 -200	Atlantic
5	1,1E+04	agriculture	sand	150 -200	Continental
Total area (km²)		6,7E+04			





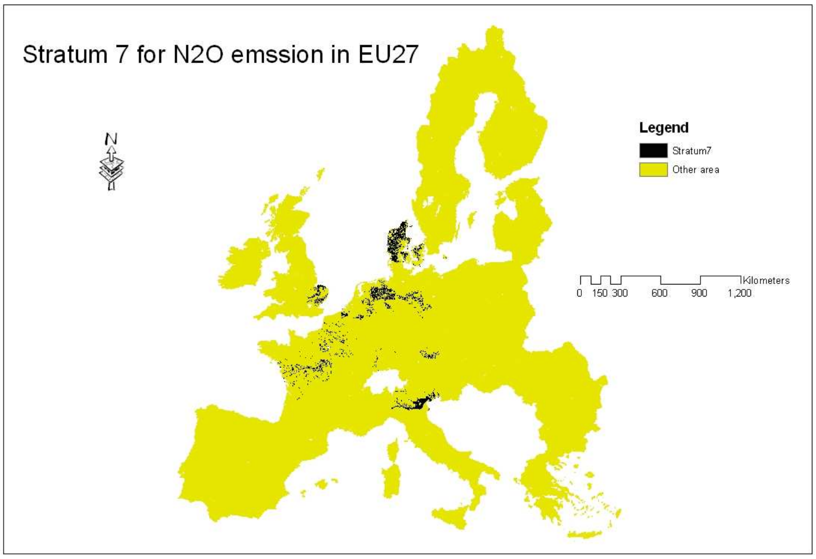


Figure 12, area stratum 7 covers in EU27

Table 14, Questionnaire for stratum 7

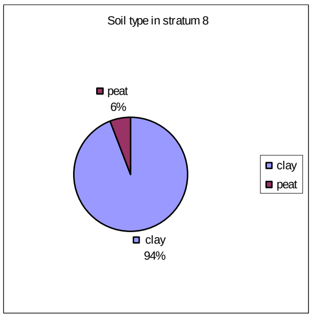
	Annual N2O emission (kg N ha^{-1} yr^{-1})
Quartile 1	
Median	
Quartile 3	

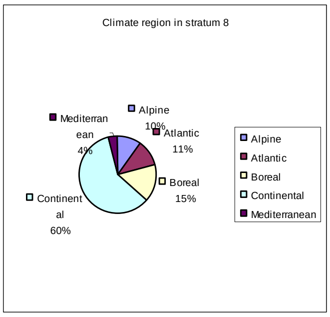
Stratum 8

The attribute table of stratum 8 is shown in table 15. Two pie charts show the statistical information of soil type and climate region in stratum 8 respectively. The figure 13 shows the area stratum8 covers in EU27. The questionnaire is in table 16.

Table 15, attributes table of stratum 8

Class	Area (km²)	Land type	Soil type	N input(kg/ha*year)	Climate region
1	9,1E+03	pasture	clay	0 - 50	Alpine
2	9,5E+03	pasture	clay	0 - 50	Atlantic
3	3,5E+03	pasture	clay	0 - 50	Mediterranean
4	4,8E+04	pasture	clay	0 - 50	Continental
5	1,3E+04	pasture	clay	0 - 50	Boreal
6	1,3E+01	pasture	peat	0 - 50	Alpine
7	7,5E+01	pasture	peat	0 - 50	Atlantic
8	4,0E+03	pasture	peat	0 - 50	Continental
9	1,2E+03	pasture	peat	0 - 50	Boreal
Total area (km²)		8,8E+04			





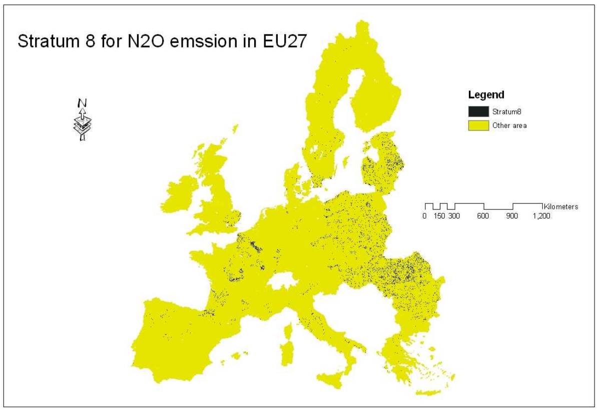


Figure 13, area stratum 8 covers in EU27.

Table 16, Questionnaire for stratum 8

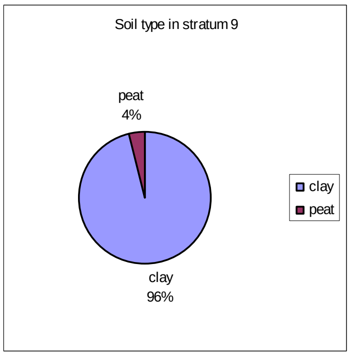
	Annual N2O emission (kg N ha^{-1} yr^{-1})
Quartile 1	
Median	
Quartile 5	

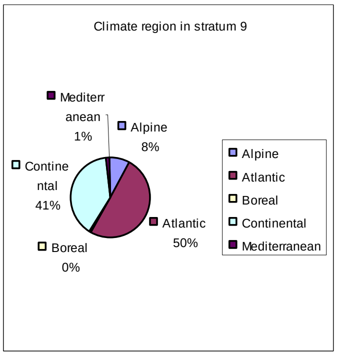
Stratum 9

The attribute table of stratum 9 is shown in table 17. Two pie charts show the statistical information of soil type and climate region in stratum 9 respectively. The figure 14 shows the area stratum9 covers in EU27. The questionnaire is in table 18.

Table 17, attributes table of stratum 9

Class	Area (km²)	Land type	Soil type	N input(kg/ha*year)	Climate region
1	8,4E+03	pasture	clay	50 - 100	Alpine
2	4,9E+04	pasture	clay	50 - 100	Atlantic
3	1,2E+03	pasture	clay	50 - 100	Mediterranean
4	4,1E+04	pasture	clay	50 - 100	Continental
5	4,8E+02	pasture	clay	50 - 100	Boreal
6	2,9E+03	pasture	peat	50 - 100	Atlantic
7	1,2E+03	pasture	peat	50 - 100	Continental
Total area (km²)		1,0E+05			





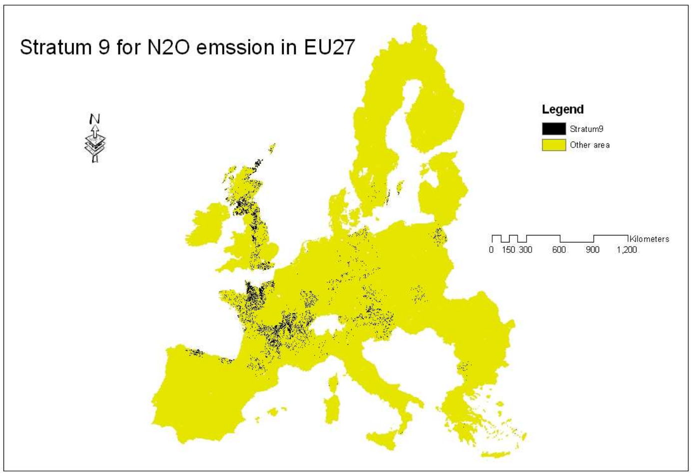


Figure 14, area stratum 9 covers in EU 27

Table 18, Questionnaire for stratum 9

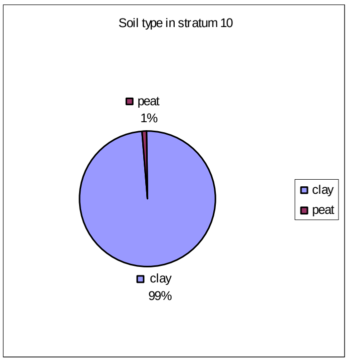
	Annual N2O emission (kg N ha^{-1} yr^{-1})
Quartile 1	
Median	
Quartile 2	

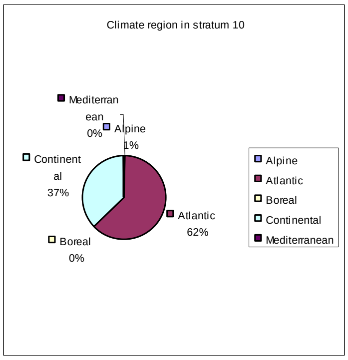
Stratum 10

The attribute table of stratum 10 is shown in table 19. Two pie charts show the statistical information of soil type and climate region in stratum 10 respectively. The figure 15 shows the area stratum10 covers in EU27. The questionnaire is in table 20.

Table 19, attribute table of stratum 10

Class	Area (km²)	Land type	Soil type	N input(kg/ha*year)	Climate region
1	3,1E+02	pasture	clay	100 - 150	Alpine
2	2,3E+04	pasture	clay	100 - 150	Atlantic
3	8,7E+01	pasture	clay	100 - 150	Mediterranean
4	1,4E+04	pasture	clay	100 - 150	Continental
5	1,0E+01	pasture	clay	100 - 150	Boreal
6	3,7E+02	pasture	peat	100 - 150	Atlantic
7	7,5E-01	pasture	peat	100 - 150	Continental
8	1,6E+00	pasture	peat	100 - 150	Boreal
Total area (km²)	3,7E+04				





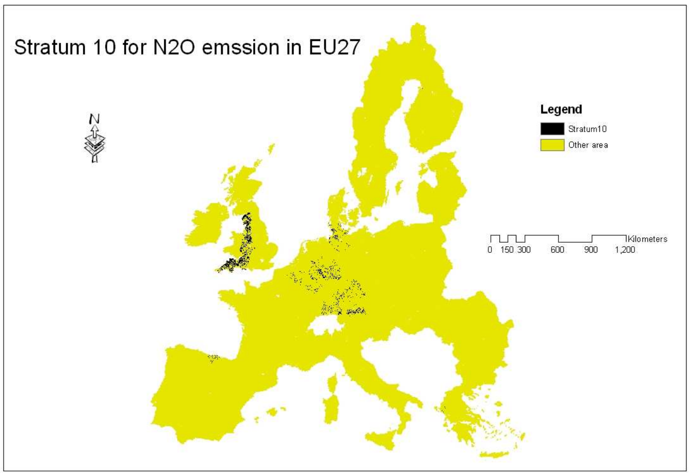


Figure 15, area stratum 10 covers in EU27

Table 20, Questionnaire for stratum 10

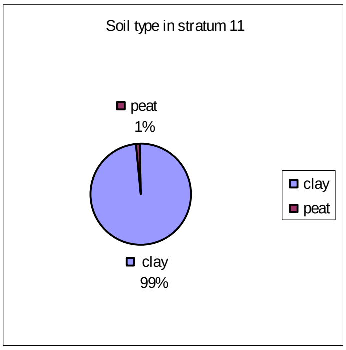
	Annual N2O emission (kg N ha^{-1} yr^{-1})
Quartile 1	
Median	
Quartile 2	

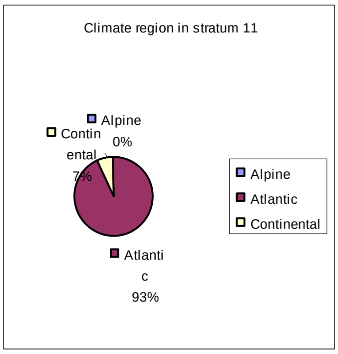
Stratum 11

The attribute table of stratum 11 is shown in table 21. Two pie charts show the statistical information of soil type and climate region in stratum 11 respectively. The figure 16 shows the area stratum11 covers in EU27. The questionnaire is in table 22.

Table 21, attribute table of stratum 11

Class	Area (km²)	Land type	Soil type	N input(kg/ha*year)	Climate region
1	1,6E+02	pasture	clay	150 -200	Alpine
2	4,7E+04	pasture	clay	150 -200	Atlantic
3	3,4E+03	pasture	clay	150 -200	Continental
4	7,5E+02	pasture	peat	150 -200	Atlantic
5	8,1E+00	pasture	peat	150 -200	Continental
Total area (km²)		5,2E+04			





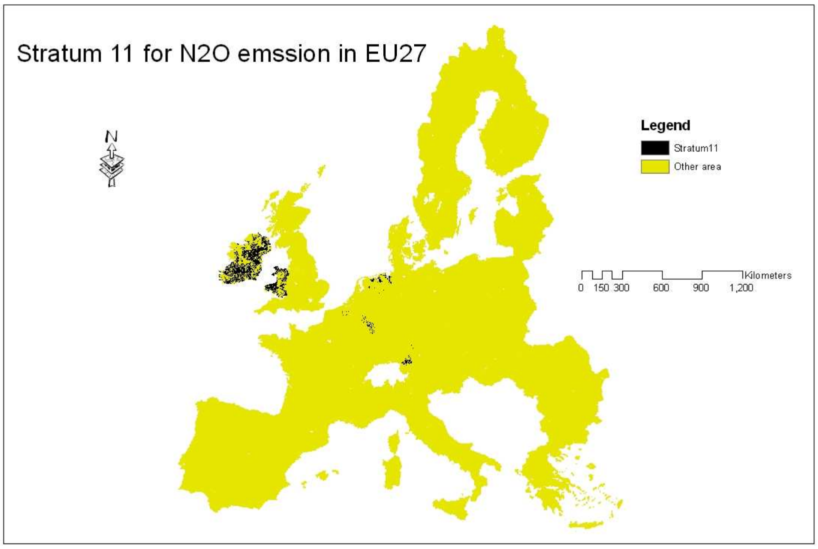


Figure 16, area stratum 11 covers in EU27

Table 22, Questionnaire for stratum 11

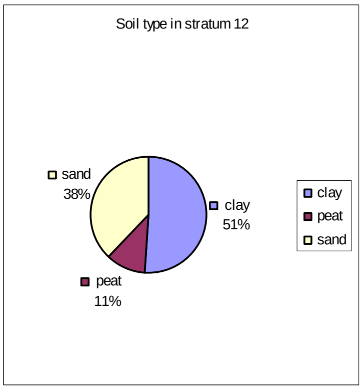
	Annual N2O emission (kg N ha^{-1} yr^{-1})
Quartile 1	
Median	
Quartile 2	

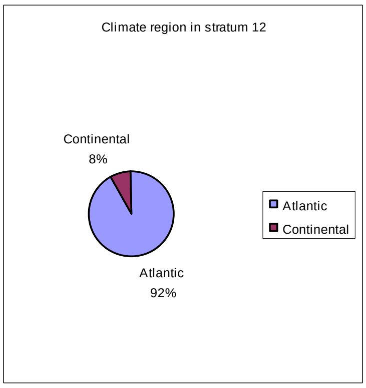
Stratum 12

The attribute table of stratum 12 is shown in table 23. Two pie charts show the statistical information of soil type and climate region in stratum 12 respectively. The figure 17 shows the area stratum12 covers in EU27. The questionnaire is in table 24.

Table 23, attributes table of stratum 12

Class	Area (km²)	Land type	Soil type	N input(kg/ha*year)	Climate region
1	4,7E+03	pasture	clay	>200	Atlantic
2	8,0E+02	pasture	clay	>200	Continental
3	1,2E+03	pasture	peat	>200	Atlantic
4	4,0E+03	pasture	sand	>200	Atlantic
5	3,7E+01	pasture	sand	>200	Continental
Total area (km²)		1,1E+04			





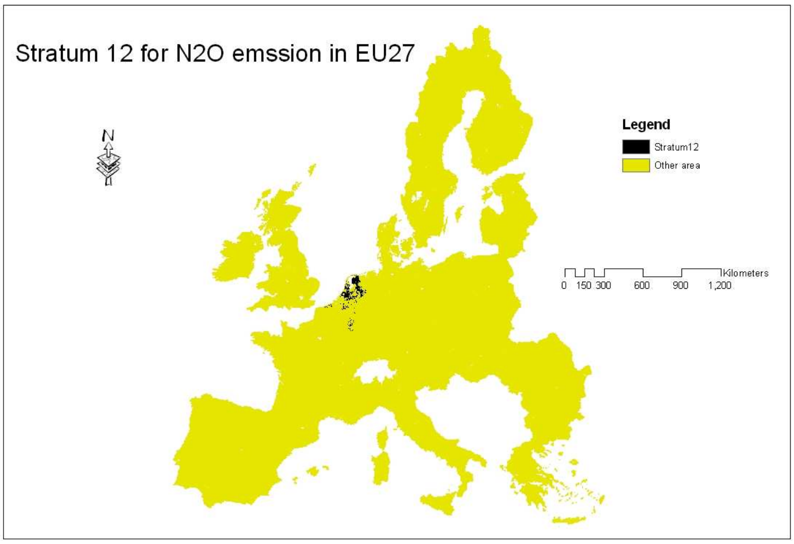


Figure 17, area stratum 12 covers in EU27

Table 24, Questionnaire for stratum 12

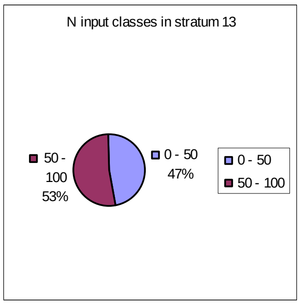
	Annual N2O emission (kg N ha^{-1} yr^{-1})
Quartile 1	
Median	
Quartile 2	

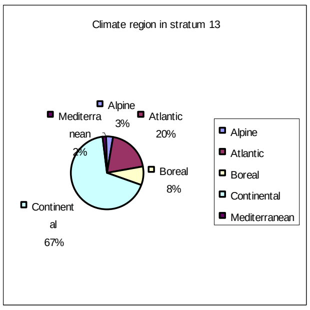
Stratum 13

The attribute table of stratum 13 is shown in table 25. Two pie charts show the statistical information of annual N application and climate region in stratum 13 respectively. The figure 18 shows the area stratum13 covers in EU27. The questionnaire is in table 26.

Table 25, attribute table of stratum 13

Class	Area (km²)	Land type	Soil type	N input(kg/ha*year)	Climate region
1	2,3E+02	pasture	sand	0 - 50	Alpine
2	2,9E+03	pasture	sand	0 - 50	Atlantic
3	3,5E+03	pasture	sand	0 - 50	Boreal
4	5,0E+02	pasture	sand	0 - 50	Mediterranean
5	1,3E+04	pasture	sand	0 - 50	Continental
6	1,1E+03	pasture	sand	50 - 100	Alpine
7	5,7E+03	pasture	sand	50 - 100	Atlantic
8	3,1E-01	pasture	sand	50 - 100	Boreal
9	1,8E+02	pasture	sand	50 - 100	Mediterranean
10	1,6E+04	pasture	sand	50 - 100	Continental
Total area (km²)		4,4E+04			





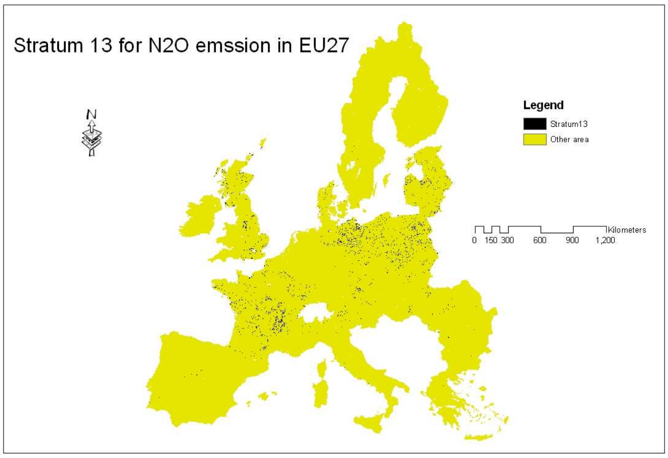


Figure 18, area stratum 13 covers in EU27

Table 26, Questionnaire for stratum 13

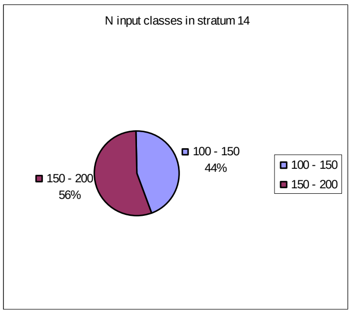
	Annual N2O emission (kg N ha^{-1} yr^{-1})
Quartile 1	
Median	
Quartile 2	

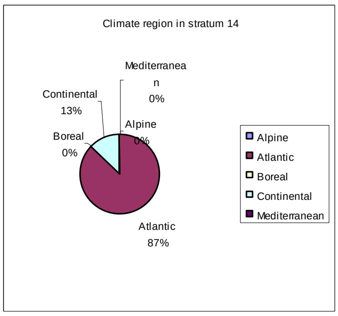
Stratum 14

The attribute table of stratum 14 is shown in table 27. Two pie charts show the statistical information of annual N application and climate region in stratum 14 respectively. The figure 19 shows the area stratum14 covers in EU27. The questionnaire is in table 28.

Table 27, attribute table of stratum 14

Class	Area (km²)	Land type	Soil type	N input(kg/ha*year)	Climate region
1	4,4E+00	pasture	sand	100 - 150	Alpine
2	8,5E+03	pasture	sand	100 - 150	Atlantic
3	4,6E+00	pasture	sand	100 - 150	Boreal
4	1,1E+00	pasture	sand	100 - 150	Mediterranean
5	2,0E+03	pasture	sand	100 - 150	Continental
6	1,2E+04	pasture	sand	150 -200	Atlantic
7	9,9E+02	pasture	sand	150 -200	Continental
Total area (km²)		2,4E+04			





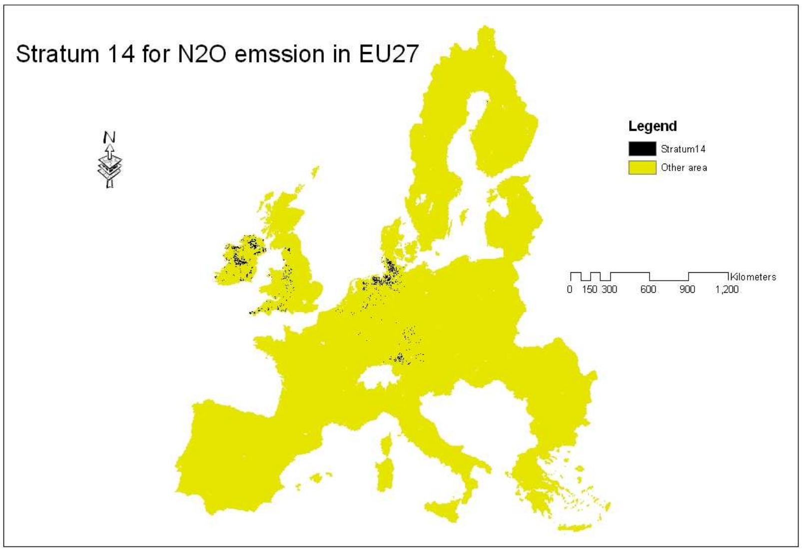


Figure 19, area stratum 14 covers in EU27

Table 28, Questionnaire for stratum 14

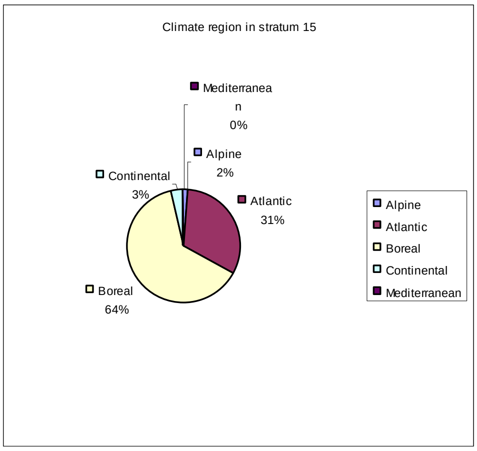
	Annual N2O emission (kg N ha^{-1} yr^{-1})
Quartile 1	
Median	
Quartile 2	

Stratum 15

The attribute table of stratum 15 is shown in table 29. Two pie charts show the statistical information of climate region in stratum 15 respectively. The figure 20 shows the area stratum15 covers in EU27. The questionnaire is in table 30.

Table 29, attributes table of stratum 15

Class	Area (km²)	Land type	Soil type	N input(kg/ha*year)	Climate region
1	2,4E+03	nature	peat	0 - 50	Alpine
2	4,5E+04	nature	peat	0 - 50	Atlantic
3	9,1E+04	nature	peat	0 - 50	Boreal
4	4,9E+03	nature	peat	0 - 50	Continental
5	7,2E+00	nature	peat	0 - 50	Mediterranean
Total area (km²)	1,4E+05				



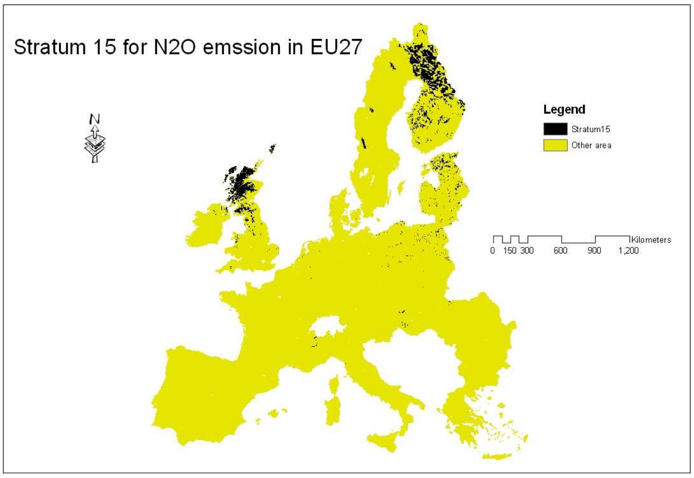


Figure 20, area stratum 15 covers in EU27

Table 30, Questionnaire for stratum 20

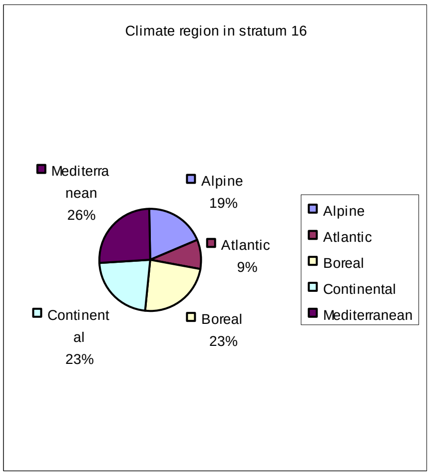
	Annual N2O emission (kg N ha^{-1} yr^{-1})
Quartile 1	
Median	
Quartile 2	

Stratum 16

The attribute table of stratum 16 is shown in table 31. Two pie charts show the statistical information of climate region in stratum 16 respectively. The figure 21 shows the area stratum16 covers in EU27. The questionnaire is in table 32.

Table 31, attribute table of stratum 16

Class	Area (km²)	Land type	Soil type	N input(kg/ha*year)	Climate region
1	2,6E+05	nature	clay	0 - 50	Alpine
2	1,2E+05	nature	clay	0 - 50	Atlantic
3	3,1E+05	nature	clay	0 - 50	Boreal
4	3,0E+05	nature	clay	0 - 50	Continental
5	3,5E+05	nature	clay	0 - 50	Mediterranean
Total area (km²)		1,3E+06			



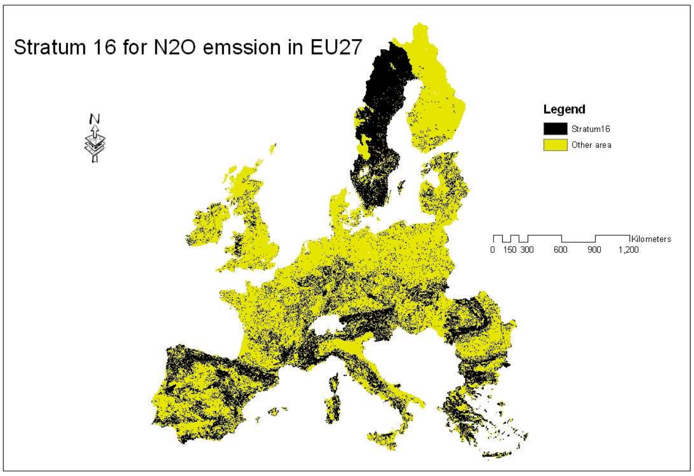


Figure 21, area stratum 16 covers in EU27

Table 32, Questionnaire for stratum 16

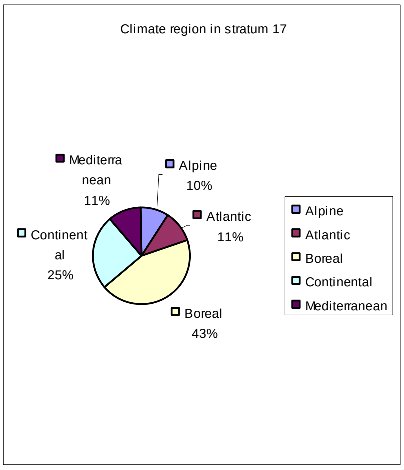
	Annual N2O emission (kg N ha^{-1} yr^{-1})
Quartile 1	
Median	
Quartile 2	

Stratum 17

The attribute table of stratum 17 is shown in table 33. Two pie charts show the statistical information of climate region in stratum 17 respectively. The figure 22 shows the area stratum17 covers in EU27. The questionnaire is in table 34.

Table 33, attribute table of stratum 17

Class	Area (km²)	Land type	Soil type	N input(kg/ha*year)	Climate region
1	4,6E+04	nature	sand	0 - 50	Alpine
2	5,3E+04	nature	sand	0 - 50	Atlantic
3	2,1E+05	nature	sand	0 - 50	Boreal
4	1,2E+05	nature	sand	0 - 50	Continental
5	5,5E+04	nature	sand	0 - 50	Mediterranean
Total area (km²)		4,9E+05			



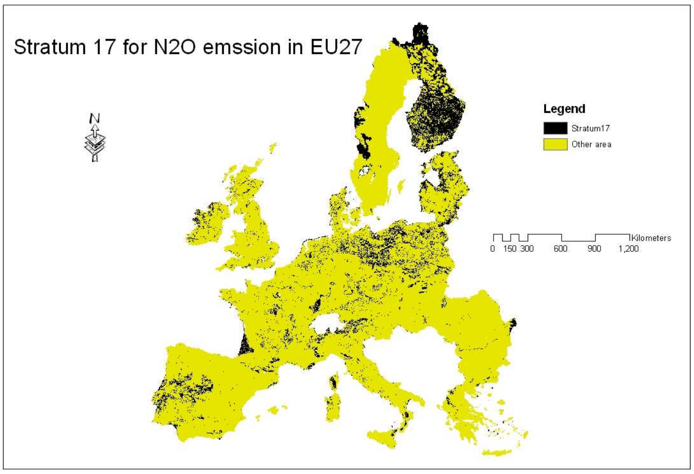


Figure 22, area stratum 17 covers in EU27

Table 34, Questionnaire for stratum 17

	Annual N2O emission (kg N ha^{-1} yr^{-1})
Quartile 1	
Median	
Quartile 2	

Thank you very much for your answering and we really want to receive some remarks on the category of stratums in the EU27 from you.

Remarks: