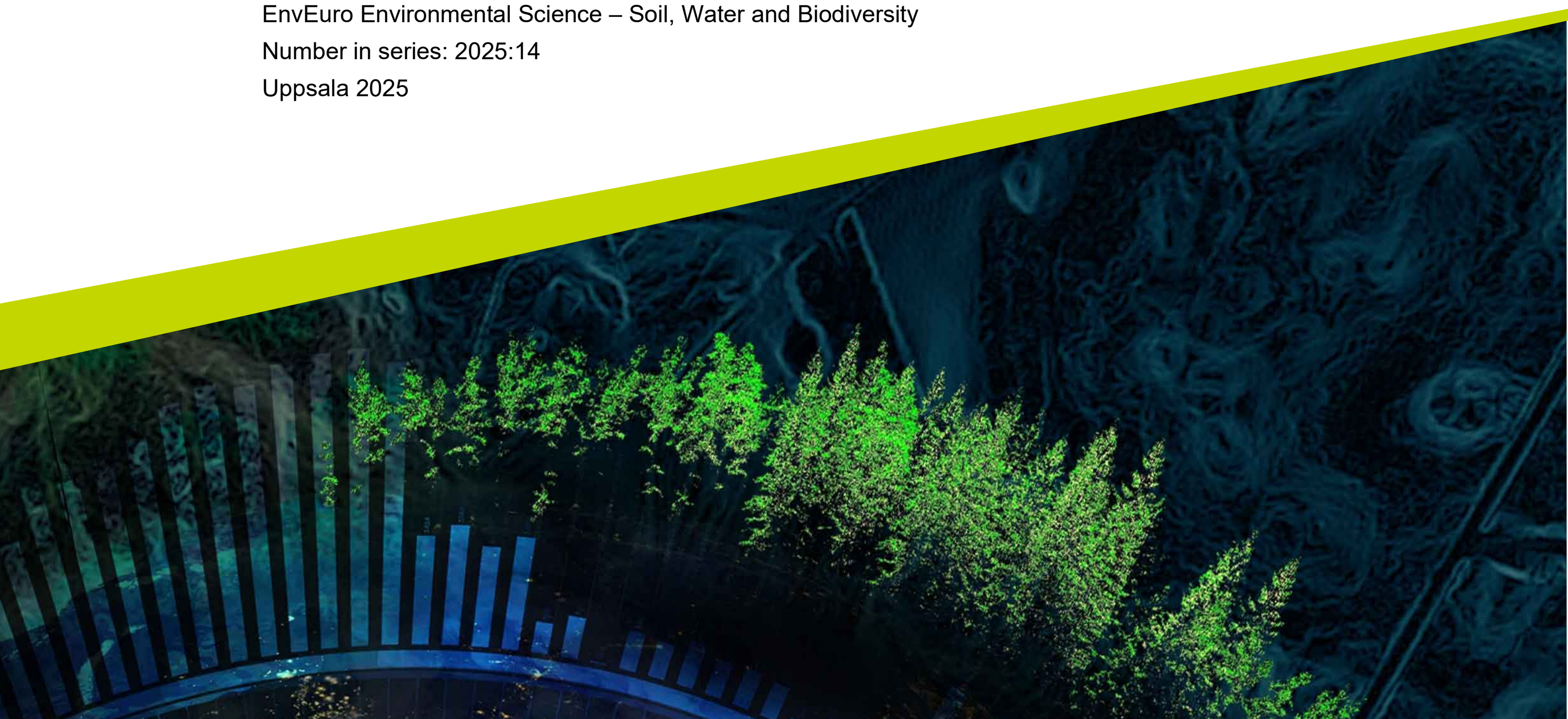
Swedish Forest Soils and the Soil Monitoring Law

Comparing selected soil properties in the SFSI and the LUCAS Soil Survey and assessing the requirements laid down by upcoming EU legislation

Merlin Mayer

Degree project/Independent project • 30 credits
Swedish University of Agricultural Sciences, SLU
Department of Soil and Environment
EnvEuro Environmental Science – Soil, Water and Biodiversity
Number in series: 2025:14
Uppsala 2025



Swedish Forest Soils and the Soil Monitoring Law. Comparing selected soil properties in the SFSI and the LUCAS Soil Survey with the requirements laid down in upcoming EU legislation

Merlin Mayer

Supervisor:	**Carina Josefsson Ortiz, SLU, Department of Soil and Environment**
Assistant supervisor:	Johan Stendahl, SLU, Department of Soil and Environment
Assistant supervisor:	Robert Jandl, BOKU, Institute of Soil Research
Examiner:	Johanna Wetterlind, SLU, Department of Soil and Environment

Credits:	30 credits
Level:	Master's level (A2E)
Course title:	Master thesis in Environmental science
Course code:	EX0897
Programme/education:	EnvEuro Environmental Science – Soil, Water and Biodiversity
Course coordinating dept:	Department of Aquatic Science and Assessment
Place of publication:	Uppsala
Year of publication:	2025

Keywords:	Soil Monitoring Law, forest soils, Sweden, soil sampling, LUCAS Soil

Swedish University of Agricultural Sciences
Faculty of Forest Sciences
Department of Soil and Environment

Acknowledgements

I would like to express my deepest gratitude to my supervisor, Carina Josefsson Ortiz, who came up with the idea for the project, supported me in every step along the way and was always eager to help with questions and discuss problems. I want to thank my co-supervisors, Johan Stendahl and Robert Jandl, for aiding me by providing their expert advice on challenging issues and feedback on my work, which I benefitted from greatly. Furthermore, I am grateful to Johanna Wetterlind for being my examiner, providing important inputs at the start of the project and putting me in touch with experienced colleagues.

A big thank you goes to Mattias Lundblad, for aiding me in finding a thesis topic and connecting me with Carina. I also want to thank Nicolas Saby for his helpful guidance on spatial sampling theory, volunteering to proof-read and providing access to the R scripts without which this project would not have been possible. Special thanks go out to Christofer Engberg Hydén for always being available for questions regarding R programming and the SFSI, as well as Johannes Larsson for providing me with data.

I would like to mention my appreciation for my fellow master students at the office, for creating a fun, supportive and productive working environment. I want to thank my flatmates for the continued support and the fun we had along the way and especially Markus, for exchanging ideas, fuelling my work with chocolate and jokes and volunteering to be my opponent.

Lastly, I would like to thank my family for supporting me throughout my studies and my decision to go abroad, and most of importantly my partner, Lena, whose unwavering support, trust, and confidence in me provides stability and constantly motivates me to perform to the best of my abilities.

Abstract

In this project the Swedish Forest Soil Inventory (SFSI) and the LUCAS Soil Survey were analysed in order to compare them regarding their methodologies and results for selected soil properties (OC, N, pH, soil texture) and assess their compliance with the soil monitoring system requirements laid down in the EU's upcoming Soil Monitoring Law. To achieve this, the two soil inventories were compared to a stratified random sampling scheme which was generated using an approach recommended by the EU's Joint Research Centre, to design and optimize a soil monitoring system in line with the new directive. It was shown that the SFSI has a good spatial coverage of Swedish Forest Soils that may fulfil the requirements, except for certain soil properties such as soil texture which will need to be measured in more detail. The LUCAS Soil Survey's sampling methodology is not well suited to the characteristics of forest soils due to only considering the upper 20 cm of the mineral soil and ignoring humus layers. But it could serve as a supplementary data source to the SFSI in reporting under the Soil Monitoring Law for soil properties where the SFSI data is not at the required level.

Variance in OC and N was found to be very high in both the SFSI and LUCAS, with coefficients of variation around 1.0. It was shown that dividing the country in soil units based on administrative borders and soil regions as proposed in the Soil Monitoring Law does not result in less variability in the data. Instead, a considerable reduction in variance could be achieved by analysing soils with and without peat layer separately.

Keywords: Soil Monitoring Law, forest soils, Sweden, soil sampling, LUCAS Soil

Table of contents

List of tables

List of figures

Abbreviations

Abbreviation	Description
CLC	Corine Land Cover
CV	Coefficient of Variation
ESDAC	European Soil Data Centre
EU	European Union
FAO	Food and Agriculture Organization of the United Nations
JRC	Joint Research Centre of the European Union
LUCAS	Land Use/Cover Area frame statistical Survey
LULUCF	Land Use, Land Use Change, and Forestry
N	Nitrogen
NFI	National forest inventory
NMD	National Land Cover Database (Nationella Marktäckedata)
NUTS	Nomenclature of territorial units for statistics; classification system for regions in the EU for statistical purposes
OC	Organic carbon
SD	Standard Deviation
SFSI	Swedish Forest Soil Inventory
SML	Soil Monitoring Law
WRB	World Reference Base for Soil Resources

1 Introduction

On 5th July 2023, the European Commission (2023) published their *Proposal for a Directive on Soil Monitoring and Resilience*, more commonly known as the Soil Monitoring Law (SML). The objective of the directive is "to put in place a solid and coherent soil monitoring framework for all soils across the EU" and achieving "healthy soils" by 2050. According to Veerman et al. (2020) at least 60-70 % of soils in the EU are unhealthy because of current unsustainable management practices and soil pollution. Although forests are less affected than other land use classes like grassland or cropland (Romero et al. 2024), Swedish forests have seen a reduction in growth rate in recent years. While climate change is most likely the driving factor (Laudon et al. 2024), healthy soils might be important to the resilience of forests to droughts of increasing frequency and intensity (Gazol et al. 2017).

Forest soils are essential to the functioning of the ecosystems built on top of them and provide ecosystem services, such as timber production, carbon storage, regulation of water flow and water filtration, and a habitat for organisms living within the soil (Paré et al. 2024). Especially boreal soils are a huge carbon reservoir. Deluca and Boisvenue (2012) cite up to ~ 1700 Pg total carbon stored in soils in the circumpolar region (including wetlands and permafrost soils). However, this carbon pool is at risk of becoming a major source of emissions as a result of climate change (Bradshaw & Warkentin 2015).

1.1 The soil monitoring law

The proposal for the soil monitoring law by the European Commission (2023) aims to improve soil health across the European Union and "achieve healthy soils by 2050". The proposal includes the following definition of soil health:

> '[S]oil health' means the physical, chemical and biological condition of the soil determining its capacity to function as a vital living system and to provide ecosystem services.

In addition to establishing a system for monitoring and assessing soil health, the text includes measures on sustainable soil management and contaminated sites. The final text of the directive will include some changes to the original proposal as a result of the trilogue negotiations between the European Commission, the European Parliament, and the Council of the European Union. At the time of writing (May 2025), a provisional agreement between the institutions has been announced but the agreed upon text has not been published yet. However, the resolution adopted by the European Parliament (2024) and the general approach agreed upon by the ministers in the Council of the European Union (2024) are available. Together with the proposal by the Commission, these two documents form the basis for the analysis of the monitoring requirements in this thesis. Because it is difficult to gauge which passages from the three versions will be included in the final text, some of the presented results (see chapter 3) might require different interpretation once the law has entered into force.

1.1.1 Member States' responsibilities under the Soil Monitoring Law

Soil districts and soil units

The text agreed by the Council, requires Member States to establish soil districts throughout their territories. These are administrative regions that provide governance structures, managing one or more soil units. These soil units, that must also be established by the Member States, should reflect "a certain degree of homogeneity" and at the very least consider the soil type based on the map of soil regions of the European Union and Adjacent Countries, and the land use categories defined in the LULUCF regulation (Regulation (EU) 2018/841) and the IPCC Guidelines (IPCC 2006). A soil unit may consist of multiple areas that are not spatially interconnected but share the characteristics mentioned above. Member States are encouraged to use the most detailed or updated information available for their soil unit definition, also taking into account climatic and environmental conditions.

The differentiation in soil districts and soil units is an advancement from the original proposal, where only soil districts were mentioned, serving as both administrative and environmentally homogeneous entity. The Parliament largely kept the soil district definition introduced by the Commission but added river basins and water bodies, islands representing individual soil districts, and the use of remote sensing data from the Copernicus programme as further considerations for Member States when delineating soil districts.

Soil monitoring system

Member States shall monitor several soil descriptors to assess soil health, and soil sealing and destruction, including but not limited to soil organic carbon stocks and concentrations, electrical conductivity, bulk density, extractable phosphorus, soil erosion rate, concentration of heavy metals, total nitrogen, and pH.

The requirements for the sampling scheme to be implemented by the Member States are laid out by the Council as follows:

> The sampling scheme shall be a stratified random sampling optimised on the best available information on the variability of soil health descriptors and the stratification shall be based on the soil units […]. The number and location of the sampling points shall represent the variability of the chosen soil descriptors within the soil units with a

> maximum percent error (or Coefficient of Variation) of 5 %. (Council of the European Union 2024)

The last sentence deviates from the original proposal where the 5 %-requirement is phrased like this:

> The size of the national sample shall meet the requirement of a maximum percent error (or Coefficient of Variation) of 5% for the estimation of the area having healthy soils. (European Commission 2023)

The most significant difference between the two passages is that in the Council's text the maximum percent error or coefficient of variation of 5 % refers to the "variability of the chosen soil descriptors", whereas in the Commission's proposal it relates to the "estimation of the area having healthy soils". The possible interpretations of these two wordings are discussed in chapter 4.3.1.

Furthermore, Member States should apply "appropriate procedures" to determine the allocation and size of the sample to correctly account for the required maximum estimation error (Council of the European Union 2024). The Bethel algorithm (Bethel 1989), which is cited as an example of such an appropriate procedure, was developed as a tool to determine the optimum allocation of samples in multivariate surveys while minimising sampling costs.

1.1.2 JRC assistance to Member States

The Council's amendments also call for the Commission to support Member States in designing a stratified random sampling scheme by providing "relevant maps of soil descriptors, the initial starting sample and the relevant data […] collected under previous European soil surveys" (Council of the European Union 2024). In fact, this process has already begun, and the Commission's Joint Research Centre (JRC) has supplied Member States with a set of EU-wide maps of soil properties (Ballabio et al. 2016, 2019) and a proposed methodology to design an optimized sampling scheme. The JRC suggests to use a script in the statistical programming language *R* (R Core Team 2024) and a specialized package (Barcaroli 2014) that applies the Bethel algorithm and a genetic algorithm[1] to determine optimal stratification and sample allocation based on input variables which are drawn from the provided soil maps.

1.2 Current state of monitoring of Swedish forest soils

Currently, Swedish forest soils are covered by two soil inventory programmes with repeated observations at constant monitoring plots: the Swedish Forest Soil

[1] A genetic algorithm can be used to optimize large samples and applies evolutionary concepts, such as reproduction, crossover, and mutation, to identify the solution with the highest "fitness" (de Gruijter et al. 2006).

Inventory (SFSI) (SLU 2023) and the Land Use/Cover Area frame statistical Survey (LUCAS) Soil (Orgiazzi et al. 2018) carried out across the European Union.

The SFSI takes place in conjunction with the National Forest Inventory (NFI), using a subset of the NFI's plots, and begun in 1963 with a consistent methodology since 1973 (Olsson 1999). Sampling happens annually and the completion of one cycle takes ten years (SLU 2023). LUCAS Soil has been first carried out in 2009 across all EU Member States (except Bulgaria and Romania which were first surveyed in 2012) and repeated in 2015, 2018, and 2022. Contrary to the SFSI, LUCAS also covers cropland and grassland in addition to forest land. The methodology has remained largely the same since 2009, although there were multiple improvements introduced in the most recent repetition (2022). For a detailed comparison of the methodology of the two inventories see chapter 2.2.

Another forest soil monitoring system that is in place in many European Countries is ICP Forests, with around 5500 plots across Europe (Wellbrock et al. 2024). Sweden's contribution to the ICP Forests monitoring network was based on SFSI plots sampled in the period 1993-2002, but since then reporting of soil data to ICP Forests has stopped. There were however repeated soil surveys in other European countries (Wellbrock et al. 2024).

1.2.1 Characteristics of Swedish forest soils

In Sweden, forests cover 68 % of the country (Statistics Sweden 2023) – a much higher share than the 39 % forest cover in the EU as a whole (Eurostat 2024). Therefore, forest soils will make up the majority of sampling plots required to fulfil the reporting requirements under the SML.

Forest soils in general have several characteristics that distinguish them from arable soils as pointed out by Wellbrock et al. (2024), who argue that the Commission's proposal for the SML does not appropriately cover the special circumstances of forest soils. In many cases, they are not fertile enough for agriculture because they are lacking nutrients, are affected by waterlogging, or located on steep slopes. This results in different management practices, which in turn lead to forest soils having other physical and chemical properties and biogeochemical dynamics than agricultural soils. Especially the accumulation of organic matter in forest soils needs to be addressed in the SML (Wellbrock et al. 2024).

In total, 60.4 % of Swedish forest soil can be classified as Podzols (Olsson et al. 2009), which are characterized by a spodic horizon, that forms following the translocation of organic substances, iron-, and aluminium-oxides from the topsoil downwards, where they precipitate. Podzols usually have a thick humus layer over an acidic bleached ash-grey E (eluvial) horizon and a comparatively thin illuvial (spodic) horizon (B) divided in a dark brownish black (Bh with humus accumulation) and a rust-brown (Bs with Al and Fe accumulation) section (Blume et al. 2016). Other common soil types (especially in the south) include Cambisols, Arenosols, Leptosols and Histosols (SLU 2024b).

Histosols (or peat soils) are especially relevant in context of soil monitoring, because the special conditions under which they are formed may cause challenges to soil sampling in the field (Dettmann et al. 2022). They are also amongst the most important carbon pools globally (Yu et al. 2010), which makes effectively protecting them highly important in climate change mitigation. Ågren et al. (2022) estimate that 18-24 % of the Swedish landmass is covered by peatlands, which means that an appropriate soil monitoring network for Sweden must include suitable methods to sample peat soils.

1.3 Scope of the project

The goal of this project was to analyse the SFSI and the LUCAS Soil Survey in order to compare them regarding their methodologies and results for selected soil properties and assess their compliance with the soil monitoring system requirements laid down in the SML.

The soil properties analysed for the comparison between the two inventories were the organic carbon (OC) and nitrogen (N) concentrations, pH and soil texture. The R script and soil property maps developed by the JRC were utilized to design a theoretically optimal stratified sampling scheme for Swedish forest soils against which the SFSI and LUCAS were evaluated. In addition, an alternative method to delineate soil units was proposed.

2 Material & Methods

2.1 Overview of data sets used and data availability

In Sweden two soil monitoring programmes exist that cover forest soils – the SFSI (SLU 2023) and the LUCAS Soil Survey (Orgiazzi et al. 2018). Both have fixed sample sites and surveys are repeated at regular intervals (ten years for the SFSI; between three and six years for LUCAS). In addition to varying interval lengths, the two programmes show considerable differences in sampling design and methodologies, in measured soil properties, and scope. These differences are discussed in detail in chapter 2.2 below.

Some SFSI data is publicly available online[2]. The SFSI data analysed in this project were supplied by the Department of Soil and Environment at SLU. Data from the inventory years 2013-2022 representing one complete cycle were used.

The LUCAS data from the surveys in 2009, 2015 and 2018 were downloaded from the ESDAC database (Panagos et al. 2022), where it is available to the public after filling the relevant request forms. Presented results are based on LUCAS 2018 data, expect for soil texture which was not measured on revisited plots after 2009 and therefore taken from the 2009 and 2015 data sets.

2.2 Sampling methodologies

2.2.1 Sampling design

SFSI

The SFSI applies a systematic random sampling approach, which is based on a regular grid. At the vertices of this grid square-shaped tracts are placed and the sample plots are located at the edges of those squares. The grid is denser in the south (ca. 5 km x 5 km) than in the north (ca. 15 km x 15 km). This sampling design was originally created for Sweden's national forest inventory (NFI) and the SFSI uses a subset of about 10 000 monitoring plots (SLU 2023). At one to two circular plots per tract soil pits are dug and site descriptions are recorded, and the soil is sampled for chemical analysis. In addition, there are plots without pit-digging where only the humus layer is sampled. Soil sample plots are circular with a radius of nine metres. The exact location where the soil is sampled within the plot is determined by a systematic procedure that ensures that the pit is not dug where the soil has been disturbed in previous surveys (Ranneby et al. 1987; Olsson 1999; SLU 2024a).

[2] https://www.slu.se/institutioner/mark-miljo/miljoanalys/markinfo/

LUCAS

The LUCAS sampling points were selected using a multi-stage stratified random sampling approach. Elevation, slope, aspect, slope curvature and land use were used as stratification covariates at a resolution of 12 km. Initially the number of selected points per land cover type was proportional to the actual land cover distribution, but one third of the points in forest land has been removed and was replaced by an equal number of additional points in cropland and grassland. The stratification and the process of sample point selection was described by Tóth et al. (2013) for the LUCAS 2009 survey. In 2015, 90 % of those points were maintained while 10 % were substituted with new plots, in part some at > 1 000 m altitude, an elevation that was not covered previously (Jones et al. 2020). The same sample points as in 2015 were also targeted by the 2018 survey (Fernández-Ugalde et al. 2022). In Sweden, 2 255 points were sampled in 2009, 1 903 in 2015, and 1 906 in 2018.

In the 2022 LUCAS Soil Survey, several changes were made, including a higher number of sample plots, soil sampling to a depth of 30 cm (instead of 20 cm) and litter sampling in forest soils[3]. However, neither a detailed descriptions of the new methodology nor data has been published yet.

2.2.2 Sampling procedure in the field

How the soil samples are taken is different between the two programmes. The SFSI requires surveyors to take different samples from several depths of the soil profile. In contrast, for LUCAS only one composite sample from the top 20 cm of soil is taken.

SFSI

The sampling procedure depends on whether only the humus layer or also the mineral soil is sampled. In any case, a humus sample, named H30 or H10, is taken, and on mineral soil sampling plots, additional samples (MP5, M10, M20, M65) are collected. The mineral soil samples require a soil pit, dug to a depth of 1 m (SLU 2024a). In the data from the period 2013-2022 mineral soil sampling was planned on 4101 out of 8743 plots. Sometimes pit-digging may not be possible because of waterlogging or stoniness. This was the case on 497 out of the 4101 planned mineral sampling plots from 2013-2022. The sample types and the conditions for them to be collected are summarized in Table 1.

[3] A short description of changes in LUCAS 2022 is available on this website: https://esdac.jrc.ec.europa.eu/content/lucas-2022-topsoil-data [Accessed: 22 April 2025] Data and reports describing the methodology will also be published there.

Table 1. Summary of SFSI sample types (SLU 2024a)

Sample type	Description	Sampling conditions
H10	Always taken from the surface down to 10 cm depth, regardless of the humus thickness. It might also contain some "clean" mineral soil.	The humus sample to be taken when the humus type is mull.
H30	This sample covers the entire humus layer down to the top of the mineral soil or to a depth of 30 cm if the humus is deeper than that.	The humus sample to be taken when the humus type is mor or peat.
H50	A deep humus sample covering the depth 30-50 cm.	Taken only on sites with planned pit-digging (i.e. mineral soil sampling), where the humus type is mor or peat and the humus is deeper than 40 cm.
MP5	Taken from the top 5 cm of the B horizon.	An MP5 sample is only taken from podzols and if the humus layer and the eluviated soil combined are less than 55 cm deep.
M10	Taken from the upper 0-10 cm of the mineral soil, directly below the O horizon (if present) without gap.	Taken when the humus is of mor- or peat-type and less than 55 cm deep or when there is no humus present. In case of mull humus, the top of the mineral soil is included in the H10 sample.
M20	Taken from a depth of 10-20 cm from the upper edge of the mineral soil.	Always taken when mineral soil is sampled unless the humus layer is deeper than 45 cm.
M65	Taken from a depth of 55-65 cm from the upper edge of the mineral soil.	Always taken when mineral soil is sampled unless the humus layer is deeper than 30 cm.

LUCAS

Compared to the SFSI sampling protocol, the LUCAS Soil Survey uses a simpler methodology, which was described by Tóth, Jones and Montanarella (2013). For the surveys in 2009, 2015 and 2018, surveyors were instructed to sample the soil "to an approximate depth of 20 cm". Vegetation residues and litter needed to be removed. The description of the methodology does not specifically mention how humus layers should be treated. A problem that is discussed in detail in chapter 2.2.4.

At each sample site five sub-samples were taken: one in the central sampling location and four sub-samples, each 2 m from the centre and offset by 90 degrees from one another, so that the five sub-sample sites from a cross. The five sub-samples are mixed to form one composite sample and sent to the lab for analysis (Tóth et al. 2013).

In 2018, the sample collection was modified on 35 % of the plots, where cylinders were used to collect soil cores from 0-10 and 10-20 cm deep. This change made it possible to measure the bulk density. On a small subset of sample sites

DNA was extracted to measure biodiversity for the first time as well (Fernández-Ugalde et al. 2022).

2.2.3 Measured soil properties

Both soil survey programmes measure a wide range of physical and chemical soil properties. An overview is presented in Table 2. In addition, the SFSI and LUCAS measure soil biodiversity from DNA samples on at least some of the plots (Fernández-Ugalde et al. 2022; Karltun et al. 2022). For more details regarding the methods used to measure the soil properties see the references at the end of the table below.

Table 2. Physical and chemical soil properties measured in the SFSI and LUCAS.

Soil property	**SFSI**	**LUCAS (2009-2018)**
pH	Laboratory – 1:5 suspension in H_2O and 0.01 M $CaCl_2$	Laboratory – 1:5 suspension in H_2O and $CaCl_2$
Cation exchange capacity (CEC)	Laboratory	Laboratory
Organic carbon (OC)	Laboratory – dry combustion	Laboratory – dry combustion
Nitrogen (N)	Laboratory – dry combustion	Laboratory – modified Kjeldahl method
Phosphorus (P)	Not measured	Laboratory
Potassium (K)	Laboratory – ICP	Laboratory
$CaCO_3$	Not measured	Laboratory – volumetric method
Soil texture	Field (see chapter 2.4.1 "Soil texture")	Laboratory – 2009: sieving and sedimentation; 2015: laser diffraction (only on new sample sites)
Bulk density	Not measured	Laboratory (only in 2018 on 35 % of sample sites)
Humus form	Field	Not recorded
Organic soil depth	Field (including humus depth on non-peat soils)	Field (only in 2018)
Total soil depth	Field	Not measured
Others	Exchangeable aluminium	Electrical conductivity, hyperspectral soil spectroscopy, Clay mineralogy, heavy metal contents (in 2009 and partly 2018)
References	Karltun et al. 2022; SLU 2024	Tóth et al. 2013; Jones et al. 2020; Fernández-Ugalde et al. 2022

2.2.4 Treatment of litter and humus

Humus forms in the SFSI

In the SFSI, humus forms are grouped into the three main types mor, mull and peat (in Swedish mår, mull and torv). The classification is based on the relative thickness of the O, H and A horizons. Per definition, an organic horizon consists at least of 20 wt% organic carbon, i.e. 35 wt% organic matter. Depending on the reduction in the rate of decomposition, mainly determined by the water table and availability of oxygen, an organic horizon is considered an H-horizon (strongly reduced rate of decomposition) or an O-horizon (weakly or no reduced rate of decomposition). The O-horizon includes two subtypes: the Of-horizon (also decomposition layer) which consists of organic material in various stages of decomposition with more than 50 % of the volume being plant residues that still contain some of their original structure; the Oh-horizon (also humic layer) which is made up of at least 75 wt% organic matter and more than 50 % of the volume being fine crumbly humus with no discernible tissue-structure (SLU 2024a).

In soils where the organic layer is present in form of an O-horizon, the humus form is mor. It is further subdivided in mor type 1, mor type 2 or moder based on the thickness ratio between the Of and Oh layers (see Figure 1). Mor layers can often overlay a thin A-horizon, which – if present – is included in the sampling of the humus layer (collected as an H30 sample; see chapter 2.2.2).

The mull humus form is characterized by an aggregated A horizon. The upper part of the humus layer already contains mineral soil because of heavy mixing between the humus and mineral horizons. A thin Of-horizon may be present (< 2 cm). In case of the transitional humus type mull-like moder, a thin Of-horizon (< 2 cm) overlays an Oh-horizon with greater mineral soil content than in moder and an A-horizon, which is thicker than in moder but thinner than in mull.

Peat soils are variable in appearance and origin but are clearly identifiable by the presence of an H-horizon. The defining feature of peats is the inhibition of biological activity because of water-logging and subsequent accumulation of poorly decomposed organic material. In the SFSI, a soil is classified as peat if the H-horizon is at least 30.5 cm thick or, when overlaying bedrock, if the H-horizon is at least 10 cm thick. When the H-horizon is thinner the humus type is recorded as peat-like mor (SLU 2024a).

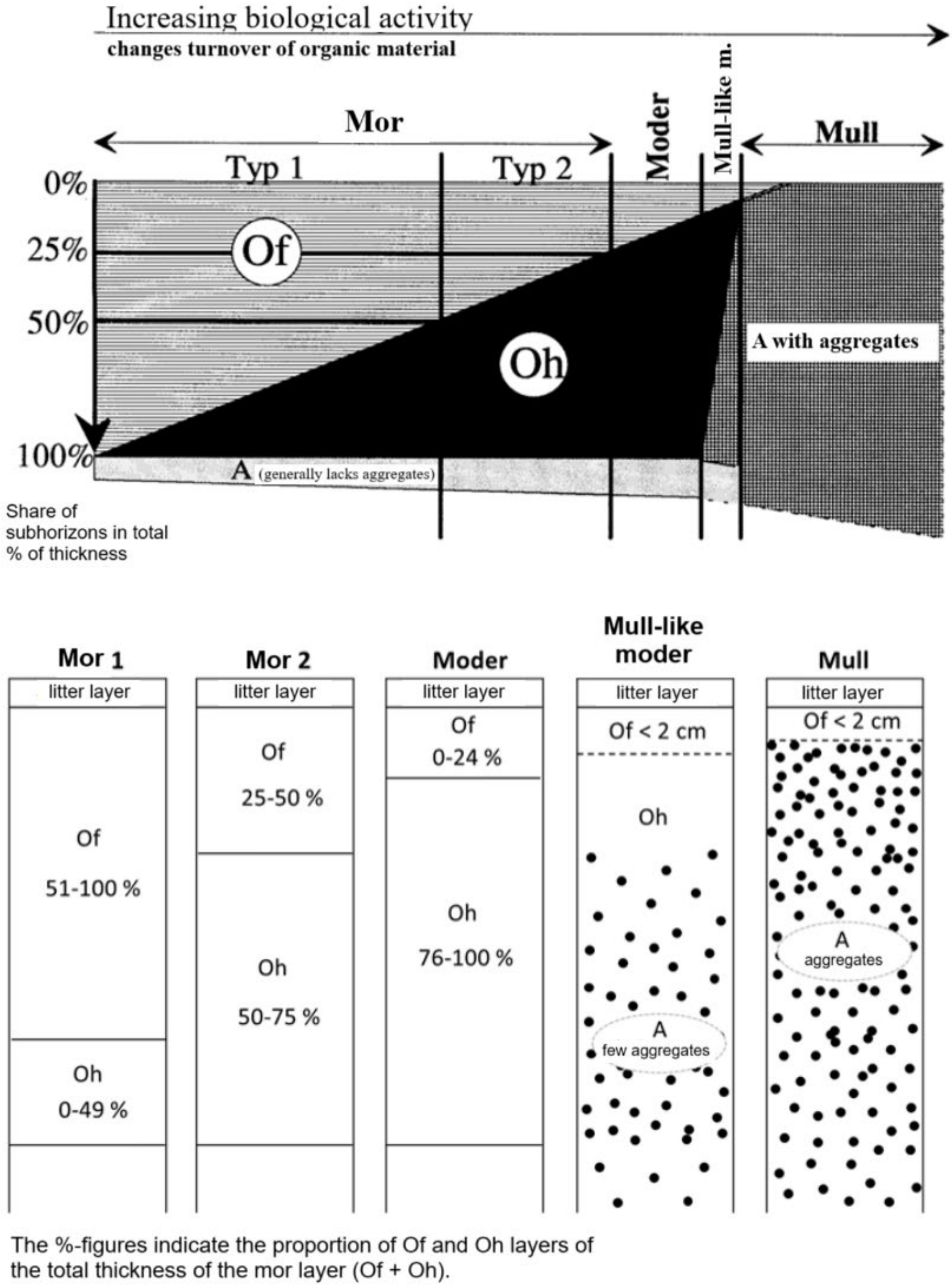


Figure 1. Visual representation of the definition of humus forms mor and mull, including sub- and transitional types, in the SFSI. Taken from the field manual and translated to English (Fältinstruktion, SLU, 2024).

Humus treatment in LUCAS

In the JRC reports describing the LUCAS surveys' methodologies (Tóth et al. 2013; Jones et al. 2020; Fernández-Ugalde et al. 2022) it is not clearly stated if and how the humus layer is sampled. Tóth et al. (2013), who are also referenced by the later reports, explain:

> In mineral soils, vegetation residues and litter were removed from the surface and the topsoil was sampled to an approximate depth of 20 cm. In case of peat, organic material was sampled. A 'mineral' topsoil may still contain fine roots, their parts and brownish homogeneous organic materials, which would have been removed by the central soil sampling laboratory through sieving in accordance with established procedures.

This paragraph does not clearly describe how an organic humus layer on top of the mineral soil is treated – except in case of peat. While it is safe to assume that mull humus, per definition well decomposed and heavily mixed with the topsoil, would be included in the soil sample, it remains unclear if mor would be removed or not since the sampling of different soil layers is also not described in more detail in any of the other JRC reports (Jones et al. 2020; Fernández-Ugalde et al. 2022).

The sampling instructions for surveyors of the LUCAS Soil Survey 2018[4] do not provide full clarity either, however they imply that organic material should be separated from the soil sample, where possible. The document states that surveyors should "[r]emove vegetation residues, grass and litter, if any, from the surface. Some fine roots and brownish organic material from the upper part of the soil can remain in the sample, as it is difficult to remove it completely" (Fernández-Ugalde et al. 2017). Surveyors are further instructed to remove "any remaining vegetation residues, stones, and litter" after taking each subsample with a spade and to again "remove any extra vegetation residues and litter" when mixing the five subsamples into one composite sample.

Because of the repeated instructions to remove organic material from the sample at several steps from the sampling to the laboratory analysis it was assumed that only humus in form of mull or peat is included in LUCAS soil samples (also see chapter 3.1). This assumption has implications for the calculation of average concentrations in the topsoil in the SFSI data (see chapter 2.4).

2.3 Statistical analysis

The statistical analysis of the two data sets consisted of the following steps which are elaborated on in the following chapters:

1. Descriptive comparison
    a. Assuring comparability between the SFSI and LUCAS
    b. Comparing distribution and statistics (mean, SD, CV) of selected soil properties (OC, N, pH, texture) at national and regional scale
2. Analysis of SML requirements and comparison with optimized sample distribution

[4] The sampling instructions for surveyors of LUCAS campaigns 2009/12 and 2015 are not available online anymore. But since the sampling procedure remained identical the version from 2018 is assumed to be valid for the earlier campaigns too.

a. Determination of optimal stratification and sample size using the R script and soil property maps provided by the JRC

b. Comparison of sampling densities between the optimized sample and the existing surveys

c. Investigating an alternative definition of soil units based on the presence of a peat layer in the SFSI data

2.4 Descriptive comparison

For the comparison between the SFSI and LUCAS, the LUCAS data from 2018 was chosen as the most suitable from the available three (2009, 2015, 2018). Following the example from Froger et al. (2024), it was deemed best to use the data set closest in time to the collection of the SFSI samples. The SFSI time period considered in this study is 2013-2022, which 2018 sits right in the middle of. Four soil properties were considered in the comparison: OC and N concentrations, as well as pH and soil texture.

2.4.1 Comparability between data sets

Before the data sets were ready to be compared some pre-processing steps were required. The biggest challenge lay in the distinctly different sampling procedures (see chapter 2.2.2) that the two monitoring programmes apply. While both use comparable methods to measure OC, N and pH, the SFSI measures each sample from different layers separately but in LUCAS there is only one value representing the composite sample from the upper 20 cm of (mineral) soil. Therefore, moulding the data into a comparable format required a transformation of the SFSI layer data to averages of the top 20 cm of soil.

Deriving top 20 cm averages from SFSI data

In order to calculate the average concentrations and pH in the upper 20 cm of the SFSI plots, the first step was to determine all possible combinations of samples taken and depths of the organic and mineral soil layers. For each possible case, the average needed to be computed differently, because the 20 cm of interest were represented by different samples of varying thickness.

This step also required a decision about whether the humus layer should be included in the upper 20 cm of soil or not. This is due to the ambiguity with which humus is treated in the LUCAS methodology (Tóth et al. 2013) – a problem that has been covered extensively in chapter 2.2.4. After closely examining the LUCAS data and experimenting with in- or excluding different humus types from the SFSI data (see chapter 3.1 in the results), a decision was made to disregard humus samples when the humus type is mor (including subtypes mor I, mor II, and moder). On plots where the humus type was mull, mull-like moder, peat, or peat-like mor, the humus was included in the topsoil averages. This approach results in more comparable distributions of organic carbon between the considered 20 cm in the SFSI and the soil sampled in LUCAS, while including as many SFSI sample plots as possible.

In total, 12 different cases with unique soil depth, humus layer and type, and taken sample combinations were recognized. They are presented in Table 3. For the cases where the total soil depth is < 20 cm, there could be between one and three samples taken at different depths. Therefore, when the soil depth was so shallow, *required* and *disallowed samples* were used to define the cases. For example, case 2 and 3 are differentiated by the presence of an M20 sample. The absence of an M20 sample means that the mineral part of the soil is no more than 10 cm deep, either because of a ≥ 10 cm organic layer or a very shallow soil (case 2). On the other hand, if an M20 sample is present, the organic layer must be thin enough to allow more than 10 cm of mineral soil (case 3), otherwise no M20 sample could be taken from a soil with a total depth < 20 cm. This approach is easier than relying on the relative depths of the organic and mineral layers and at the same time checks, whether all needed samples to correctly calculate the average are present.

The simplest cases are 1, 4, 7, and 8 (Table 3). Here the top 20 cm averages equal the concentration in just one sample. In 4 and 7 this is because the soil is less than 10 cm deep and only made up of one layer (organic or mineral), therefore only one sample is taken. The same is true for case 1, although here the soil could be up to 20 cm deep. Case 8 is special because this is the only case where the resulting value likely represents more than just the upper 20 cm. The condition for it is an organic layer depth of ≥ 20 cm and the H30 sample that is used must therefore extend to a depth of 20-30 cm. There is no unbiased way to correct for this, so the concentrations are assumed to be representative of the top 20 cm.

Table 3. Overview of the 12 cases used to calculate the top 20 cm averages in the SFSI data set. The conditions that needed to be met for each equation to be applied are shown in columns 2-6.

Case Nr.	Total soil depth (cm)	Humus type	Organic layer depth (cm)	Required samples	Disallowed samples	Equation
1	< 20	Peat	Any	H30	M10, M20	$= H30$
2	< 20	Peat	Any	H30, M10	M20	$= H30 * \frac{d_o}{d_{tot}} + M10 * \frac{d_{min}}{d_{tot}}$
3	< 20	Peat	Any	H30, M10, M20	-	$= H30 * \frac{d_o}{d_{tot}} + M10 * \frac{10}{\mathrm{d_{tot}}} + M20 * \frac{\mathrm{d_{min}} - 10}{\mathrm{d_{tot}}}$
4	< 20	Mull	Any	H10	M20	$= H10$
5	< 20	Mull	Any	H10, M20	-	$= H10 * \frac{10}{d_{tot}} + M20 * \frac{d_{min} - 10}{d_{tot}}$
6	< 20	Mor or none	Any	M10, M20	-	$= M10 * \frac{10}{d_{min}} + M20 * \frac{d_{min} - 10}{d_{min}}$
7	< 20	Mor or none	Any	M10	M20	$= M10$
8	≥ 20	Peat	≥ 20	H30	-	$= H30$
9	≥ 20	Peat	≥ 10	H30, M10	-	$= H30 * \frac{d_o}{20} + M10 * \frac{20 - d_o}{20}$
10	≥ 20	Peat	< 10	H30, M10, M20	-	$= H30 * \frac{d_o}{20} + M10 * \frac{1}{2} + M20 * \frac{10 - d_o}{20}$
11	≥ 20	Mull	Any	H10, M20	-	$= \frac{H10 + M20}{2}$
12	≥ 20	Mor or none	Any	M10, M20	-	$= \frac{M10 + M20}{2}$

d_o: Organic layer depth

d_{tot}: Total soil depth

d_{min}: Mineral soil depth (equals $d_{tot} - d_o$)

In cases 2, 3, 5, 6, 9, and 10 the averages are weighted means, and the weights are determined by a combination of the total soil depth and the depths of the mineral and organic layers. In the first three cases, the weights are required to adjust for the total soil depth being less than 20 cm and the lowest sample potentially being representative of less than 10 cm, while in cases 9 and 10 the weights depend on the depth of the organic layer, since the depth of the H30 sample is also variable. In cases 11 and 12 a regular unweighted mean can be used.

Estimating means with area weights

Since the SFSI uses a systematic sampling design based on a grid of varying size (see chapter 2.2.1) a weighted mean must be used. For this purpose, every plot is linked to the area that it represents in the sampling grid which is used as an areal weight. These weights are different for the mineral and the humus layer because of the different sampling density of the two layers. Essentially a humus sample represents a much smaller area than a mineral sample. Due to the difference in the weights for the mineral and the humus layer, a weighted tract level mean was used to account for the different densities of the grid.

First the average of the top 20 cm concentrations by tract was estimated by averaging all plot level values in a tract. The national or regional mean is then estimated by taking the weighted mean of all tracts in the area of interest, using the average area represented by the humus samples in the tracts. Note that the standard deviation of these national and regional means does not consider the variance on tract level but only represents the variance between tracts.

Soil texture

One area where the SFSI data is less precise than LUCAS is soil texture. In the SFSI soil texture is assessed by the surveyors in the field and categorized on two levels. First, the parent material type is described. It can either be bedrock (häll), tills (morän), poorly or well sorted sediments (sediment med låg/hög sorteringsgrad) or peat (torv). Based on this differentiation, there are two different tables available – one for mineral sediments (based on Lindén 2002) and one for tills (based on Atterberg's classification) – that determine the categorisation of the soil texture in one of nine groups, depending on tests of the shape and rollability of the mineral soil (SLU 2024a).

For comparison with LUCAS, where soil texture is measured in the laboratory (Jones et al. 2020) by sieving and sedimentation (2009) or laser diffraction (2015), the SFSI texture categories needed to be transformed into numerical values. For the conversion, a table provided by Ťupek et al. (2016) was used.

Table 4. Texture conversion table adapted from Ťupek et al. *(2016).*

	Sediments			**Tills**		
Texture category	**Sand %**	**Silt %**	**Clay %**	**Sand %**	**Silt %**	**Clay %**
Bedrock	0	0	0	0	0	0
Boulder	0	0	0	0	0	0
Gravel	10	0	0	10	0	0
Coarse Sand	40	5	0	40	5	0
Sand	80	10	0	45	10	0
Fine sand	70	25	5	55	15	0
Coarse silt	50	40	10	65	20	5
Fine silt	10	75	15	55	35	10
Clay	0	65	35	0	85	15

In the LUCAS data the particle size distribution is given as the sand, silt and sand share of the fine soil fraction (< 20 mm), ignoring the coarse fraction, which is reported separately. Therefore, the values shown in Table 4 had to be further adapted to be comparable. This was done by dividing the absolute share of a texture group by the absolute share of all texture groups. For instance, the final sand texture values have been calculated as the share of sand divided by the sum of the sand, silt and sand shares. The resulting sand share in soils with the texture category *gravel* subsequently becomes 100 %.

Limiting the LUCAS data to forest land

Contrary to the SFSI, the LUCAS Soil Survey is carried out not only on forest land but also on cropland and grassland. Hence, for a comparison those other land uses were omitted. The LUCAS data sets come with a Corine Land Cover (CLC) class included which was replaced by a more accurate national land cover data source, the National Land Cover Database (NMD) (Swedish Environmental Protection Agency 2024). The NMD provides a raster file with 10 m x 10 m resolution in the SWEREF99 coordinate reference system (CRS), mapping the land cover classes forest, open wetland, cropland, other open land, developed (built-up) land, and water, each with several detailed subcategories. For the comparison of the soil data sets only those LUCAS plots that lie on forest land have been selected, using a forest mask that was created from the NMD map from 2018.

2.4.2 Comparison of selected soil properties

The comparison of the selected soil properties C, N, pH and texture is drawing from the work of Froger et al. (2024) and many of the figures and statistics presented here were created using adapted code from the R script that was developed by their working group to compare LUCAS to national soil inventory systems.

In addition to the presented graphs, detailed descriptive statistics (mean, standard deviation, coefficient of variation, median, quartiles, min, max) were calculated for all selected soil properties on regional (NUTS[5] 2) and national level. For the SFSI, means and standard deviations (SD) were calculated using the areal weights described in chapter 2.4.1. The formulas used were

$$\bar{x} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}, \tag{1}$$

and

$$\sigma^2 = \frac{\sum_{i=1}^{n} w_i (x_i - \bar{x})^2}{\sum_{i=1}^{n} w_i}, \tag{2}$$

where $\bar{x}$ is the weighted mean, of the tract averages x_i with weights w_i, and σ^2 is the squared SD (variance).

For LUCAS, the means and SDs were calculated as they would be for simple random sampling. To avoid bias introduced by the stratification, a weighted mean (and variance) of the strata means (and variance) should be used (de Gruijter et al. 2006). However, no information about the strata used to determine the LUCAS sample points is available. Hence, this biased cannot be corrected and the equations

$$\bar{x} = \frac{1}{\mathrm{n}} \sum_{i=1}^{n} x_i\,, \tag{3}$$

and

$$\sigma^2 = \frac{1}{\mathrm{n} - 1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \tag{4}$$

were used instead.

The coefficients of variation (CV) were calculated as the ratio between the SD and the (weighted) mean:

$$CV = \frac{\sigma}{\bar{x}} \tag{5}$$

2.4.3 Complementary data sources and used software

In addition to the two soil data sets, the following data sources were used:

- A shapefile of EU NUTS regions from the Eurostat website[6] at the highest available resolution (1:1 000 000)

[5] NUTS stands for Nomenclature of Territorial Units for Statistics, which is a classification of EU regions for statistical purposes with three different levels (NUTS 1, 2, and 3).
[6] https://ec.europa.eu/eurostat/web/gisco/geodata/statistical-units/territorial-units-statistics

- FAO and WRB soil maps from the ESDAC European Soil Database (Panagos et al. 2022)

The NUTS regions were used to calculate regional statistics from the SFSI and LUCAS sampling plots. The FAO and WRB soil maps (resolution 1 km x 1 km) were used to assess how well the soil data sets cover the distribution of soil types in Sweden.

All data preparation, calculations and plotting was done using R (R Core Team 2024) and the *tidyverse* environment (Wickham et al. 2019). The packages *terra* (Hijmans 2025) and *sf* (Pebesma 2018) with its *units* extension (Pebesma et al. 2016) were used for working with spatial data. Figures were created with *ggplot* (Wickham 2016) and *tmap* (Tennekes 2018) with additions from *patchwork* (Pedersen 2024), and *rnaturalearth* (Massicotte & South 2023).

2.5 Assessment of SML criteria

The JRC script provided to Member States determines an optimized stratified sample that complies with the sampling requirements of the SML. The process to generate this ideal sample is illustrated in chapter 2.5.1 and in chapter 3.3 the solution is compared to the existing surveys of Swedish forest soils to evaluate their compliance with the SML requirements.

2.5.1 Determining an optimized stratified sample in line with the SML

In this project, a modified version of the script provided by JRC was used. It was created over the course of a project at the French research institute for agriculture, nutrition, and environment (INRAE). The modifications were done by Nicolas Saby to simplify the script and adapt it to the French circumstances. Various versions of the script are available on Saby's GitLab page[7]. The script that was used for this project was created by adapting the file *2-SMLFilltableJRCTests.R*. Changes to the original file included adapting the input data to Swedish conditions, necessary adaptations to run the tests explained at the end of this chapter, and a general restructuring of the script for more ease of use. All relevant steps in the code are explained below.

Raster maps

As a starting point, information about the distribution of soil properties over the area of interest (Sweden) was required. In the JRC approach, this takes the form of Europe-wide soil maps of organic carbon, nitrogen, phosphorus, pH, cation exchange capacity (CEC), sand, silt, clay, and bulk density, which were provided

[7] https://forgemia.inra.fr/nicolassaby/smlsamplingeu

by the JRC. The maps have a 100 m x 100 m resolution in WGS 84 / Pseudo-Mercator (EPSG:3857) and are based on soil maps created by Ballabio et al. (2016, 2019) from LUCAS soil data. The raster files were combined with rasters displaying the NUTS 1 and 2 regions and the CLC land cover to form one stacked raster. For the application to Sweden, CLC was replaced with the NMD land cover map and the stack was reprojected to SWEREF99 (EPSG:3006).

Creating a sampling frame

A random sample was drawn from the stacked raster because of the large amount of data (the raster maps for Sweden have 113 million pixels). The sample size used for Sweden is 1 440 000. Initially the script was designed with Belgium as an example with 80 000 sampled pixels. This number was increased proportionally to the area of Sweden compared to Belgium. The size of the sample was further reduced by applying a conditioned latin hypercube sampling[8] (CLHS) algorithm with the R package *clhs* (Roudier 2011) to reduce the sample size to 90 000 pixels (again upscaled from 5 000 for the area of Belgium), which was then used as input (sampling frame) to the algorithm that determines the optimal stratification.

Optimizing the strata and sample size

In stratified simple random sampling, the area of interest is divided into smaller areas, which are called strata. The aim of this stratification is to reduce the variation of the target variables within the strata, which allows to save sampling cost (smaller sample size) or reduce overall variance in the data (de Gruijter et al. 2006). When designing an optimal sampling scheme in line with the soil monitoring law, there are multiple areas of interests to consider – the soil units, introduced in chapter 1.1.1. In the JRC script, these areas of interest are referred to as "domains" and within each domain separate strata are created for an optimal distribution of sampling sites while achieving a CV below 5 %.

To choose the stratification that is optimal for the data in question specialized algorithms can be used. In the script, the R package *SamplingStrata* (Barcaroli 2014) was used for the optimization process. It applies a genetic algorithm to optimize the stratification and determine the minimum sample size per strata while also using the Bethel algorithm (Bethel 1989) for calculating the sampling cost (Ballin & Barcaroli 2013).

The functions included in the package require the following inputs:

- A data frame containing the 90 000 pixels selected by the CLHS algorithm and all variables from the raster files

[8] Conditioned latin hypercube sampling can be used to efficiently sample variables from their multivariate distributions. It is an effective method to replicate the distribution of the variables in the sample (Minasny & McBratney 2006).

- Stratification variables (X): the variables that the stratification should be based on
- Target variables (Y): the variables of interest
- Areas of interest (domains)
- Maximum coefficients of variation (CVs) per domain and target variable

All available variables from the raster files (OC, N, P, pH, CEC, texture, bulk density) were used as both stratification and target variables since all those properties are also required to be sampled by the SML. In addition, using all available data to define the strata ensures that they are as homogeneous as possible (de Gruijter et al. 2006).

The coefficients of variation were set to 0.05 for all variables and all domains to match the criteria of the SML (European Commission 2023).

The *SamplingStrata* package defines “domains” as equivalent to the soil units introduced by the SML. EU Member States may decide how to define the soil units within their territory. Therefore, it is not clear yet, how this will be implemented in Sweden. Soil units should reflect areas with relatively uniform soil characteristics and Member States may use the soil type, land use and administrative regions such as NUTS to delineate them. In this project, three different combinations of these categories are used to define the domains.

Variations of domain definitions used to determine the optimized sample

Following the recommendation in the SML, the map of “Soil Regions of the European Union and Adjacent Countries” (BGR, 2005; see Figure 2) is used as data source for defining the dominant soil type. In combination with the NMD land cover map and the NUTS regions (for maps of both, see Appendix 2), it forms the basis of defining the domains used in the script.

To test how the resulting sample size responds to the domain definitions, three different tests were run. In the first, the domains were defined as units of the same land use type and the same soil and NUTS region. For the administrative regions, a combination of NUTS 1 and 2 regions was used. In the south, the NUTS 1 regions (SE1, SE2) were used, and in the north the NUTS 2 regions (SE31, SE32, SE33). The NUTS regions were designed to represent areas of similar population size but with Sweden’s north being much more scarcely populated than the south, the NUTS regions there are geographically much larger. Therefore, combining NUTS 1 with NUTS 2 regions results in administrative regions of more comparable size, which is a more reasonable delineation for a spatial sampling application. The second test ignores the soil region and uses only the combined NUTS regions and the land use type to design the optimized sample, while test 3 uses only NUTS 1 regions but is otherwise identical to test 1.

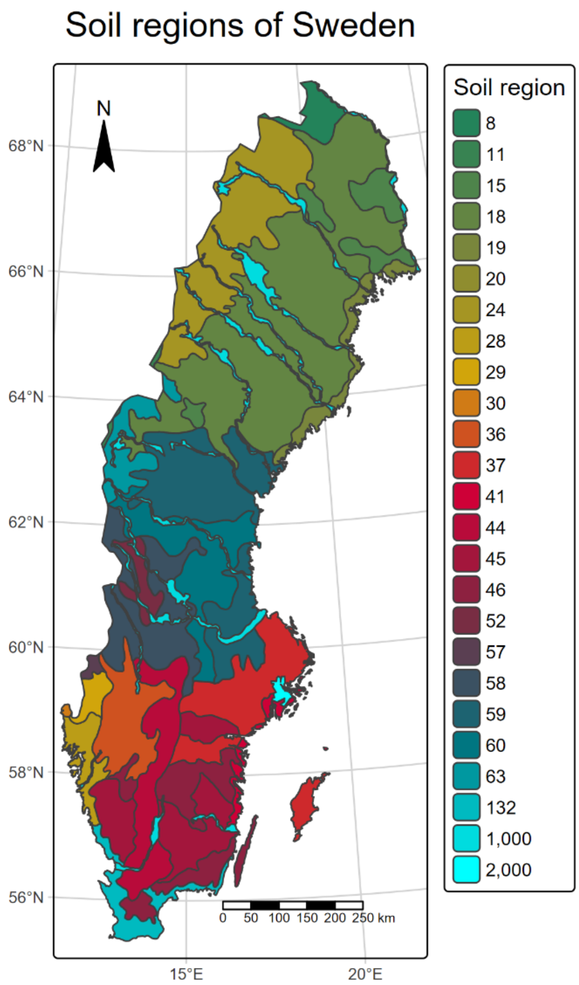


Figure 2. Soil Regions of Sweden based on the Soil Regions of the European Union and Adjacent Countries map (BGR 2005). See Appendix 2 for an explanation of the soil region numbers.

3 Results

The first section of the results covers the comparison between the two soil surveys. The spatial coverage of the SFSI and LUCAS are presented, together with detailed descriptive statistics of the analysed soil properties. Possible correlations between sample plots of different surveys in close proximity are also explored. The second section deals with the optimized sample created based on the requirements of the soil monitoring law and how it compares to the existing surveys, as well as exploring an SFSI data-based soil unit definition based on the presence of a peat layer.

3.1 Organic carbon distribution depending on humus type inclusion

As described in chapter 2.4.1, the data SFSI had to be made comparable to LUCAS by calculating top 20 cm averages. In addition to combining the different sample types taken by the SFSI into representative average values (see Table 3) a decision about which humus types to incorporate had to be made.

The distribution of the organic carbon concentration differs considerably, depending on which humus types are included (Figure 3). When all samples collected by the SFSI are included (Figure 3a), the data shows a clear bimodal distribution, with the left peak around 50 g kg^{-1} OC representing mineral soils with a shallow humus layer and the right peak peat soils and mineral soils with a thick carbon-rich humus layer with more than 400 g kg^{-1} OC. Comparatively to this first treatment of the SFSI, the distributions in LUCAS (Figure 3e and 3f) are distinctly different with a sharp peak at very low OC contents and a long thin tail and a minor second peak at 450-500 g kg^{-1} OC.

When matching the SFSI data with LUCAS, selected humus types were excluded from the calculation of the top 20 cm in the SFSI data. Two versions were tested: exclusion of the humus layer when the humus type is mor (Figure 3b), and when the humus type is mor or peat-like mor (Figure 3c). In both cases the distributions are quite similar, showing a much smaller peak in the carbon-rich soils because only true peat (and peat-like mor) is included.

However, an even more comparable distribution to LUCAS was produced by including only those plots where pit-digging and mineral soil sampling occurred (Figure 3d). Thereby almost all peat soils (where pit-digging is difficult because of waterlogging) are excluded. Mor humus is also excluded from the top 20 cm average.

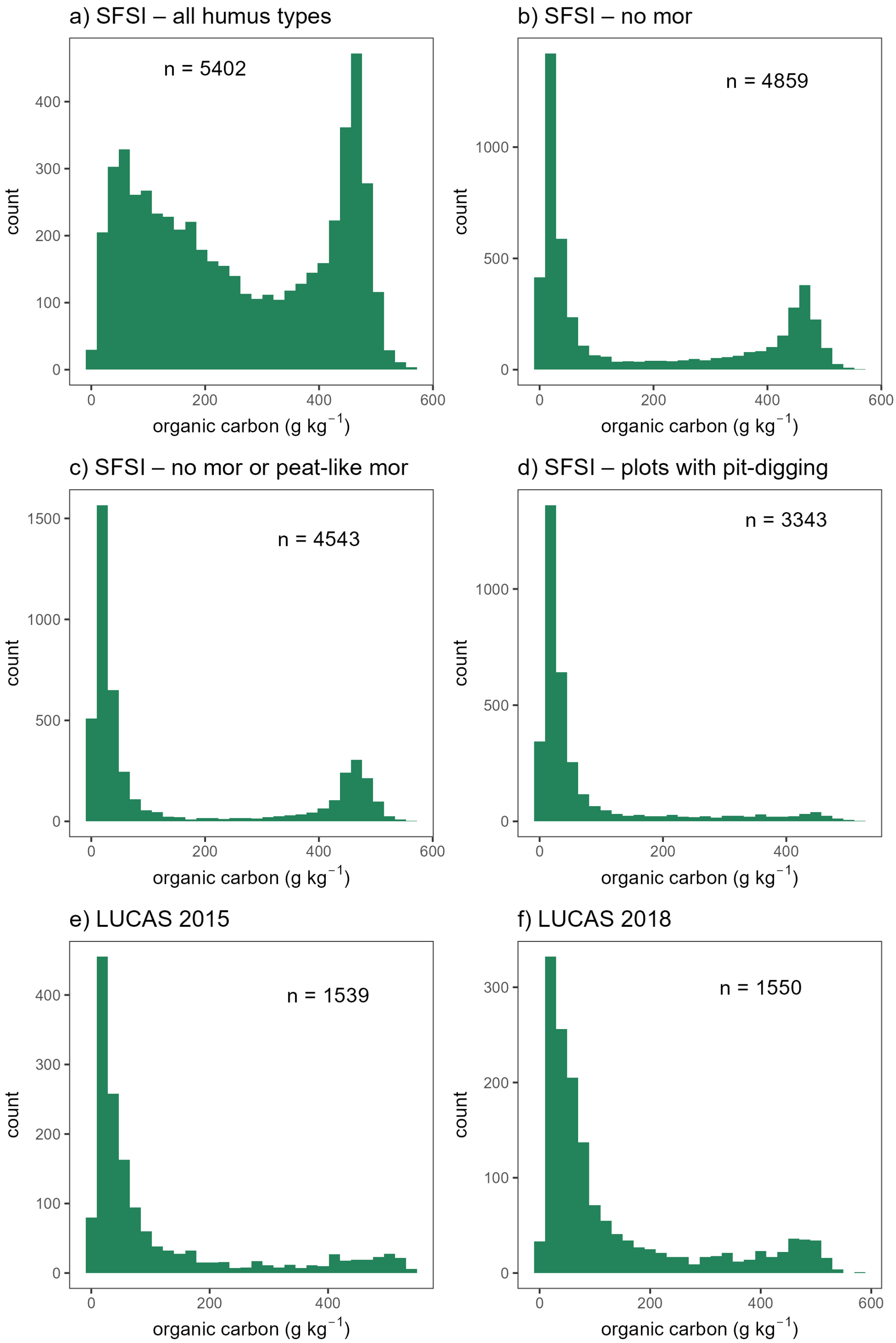


Figure 3. Statistical distribution of organic carbon in the top 20 cm of soil based on SFSI data 2013-2022, including all humus types (a), excluding mor samples (b), excluding mor and peat-like mor (c), and only including plots with pit-digging (d), compared to the LUCAS surveys 2015 (e) and 2018 (f). All subplots include the number of observations (n).

3.2 Comparison between the SFSI and LUCAS

The following chapters compare the SFSI to LUCAS, showing the spatial distribution of sample points, national and regional statistics of selected variables (C and N concentration, pH, soil texture) and the relationship between spatially close sample points.

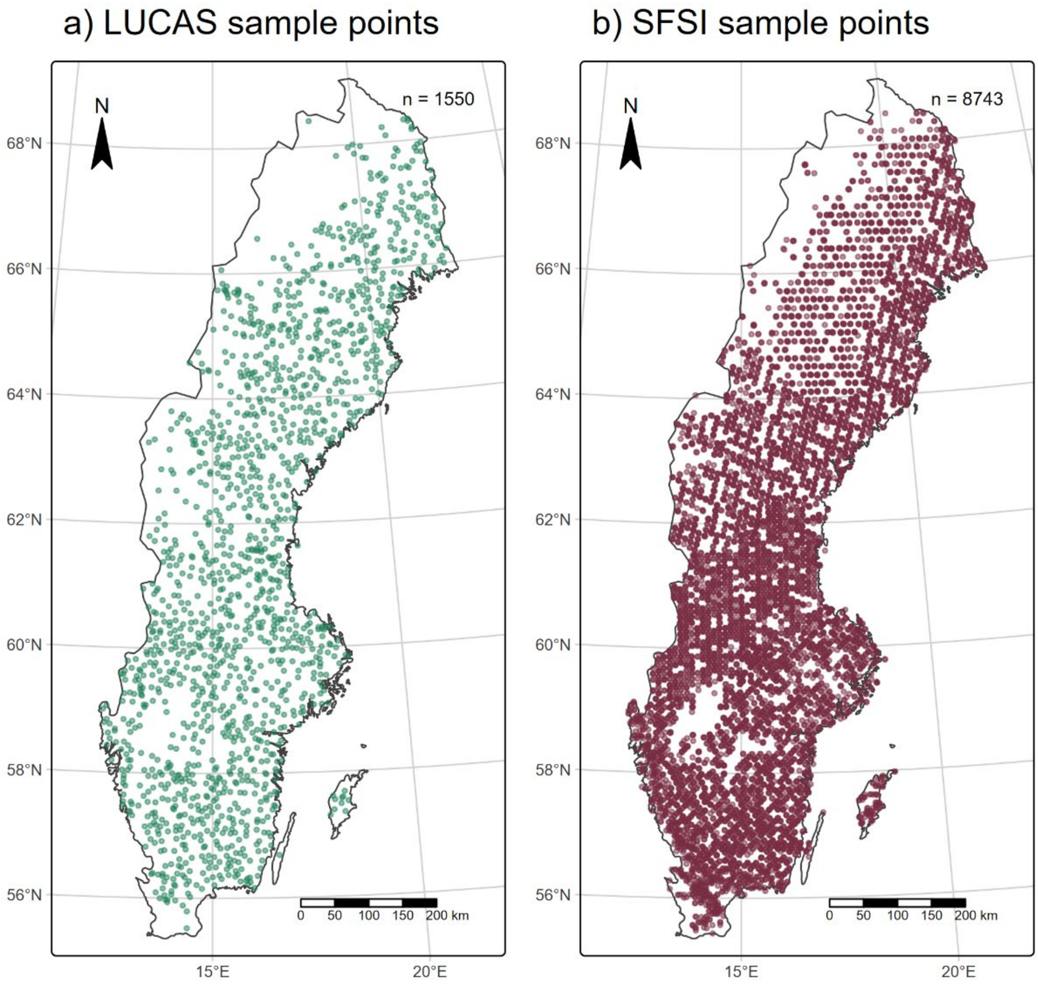


Figure 4. Mapped Swedish sample points of LUCAS 2018 (left) in forest land and SFSI (right) plots sampled in 2013-2022.

3.2.1 Spatial distribution

Figure 4 shows the spatial distribution of LUCAS and SFSI sample points over Sweden. Only LUCAS points in forest land are shown. The SFSI map includes all sample points of the inventory cycle, including those with only humus sampling and where no top 20 cm average could be calculated (see chapter 2.4.1). It is clear to see that the SFSI is based on a regular grid (of varying density) by the systematic placement of the sample plots at regular intervals, while the LUCAS plots were determined using stratified random sampling.

The difference in the spatial coverage of the two surveys is also illustrated in Figure 5, where the point density is shown. The density is much higher in the SFSI with up to 40 plots per pixel, while in LUCAS the biggest count per pixel is 8. The SFSI's grid makes the density also appear more uniform than LUCAS, where also more pixels than in the SFSI contain no sample plots at all. Both inventory programmes show higher sample density in the southern half of the country, with the north and especially the mountainous region in the north-west being scarcely covered. The empty pixels around 58-59 °N are due to the big lakes in the area.

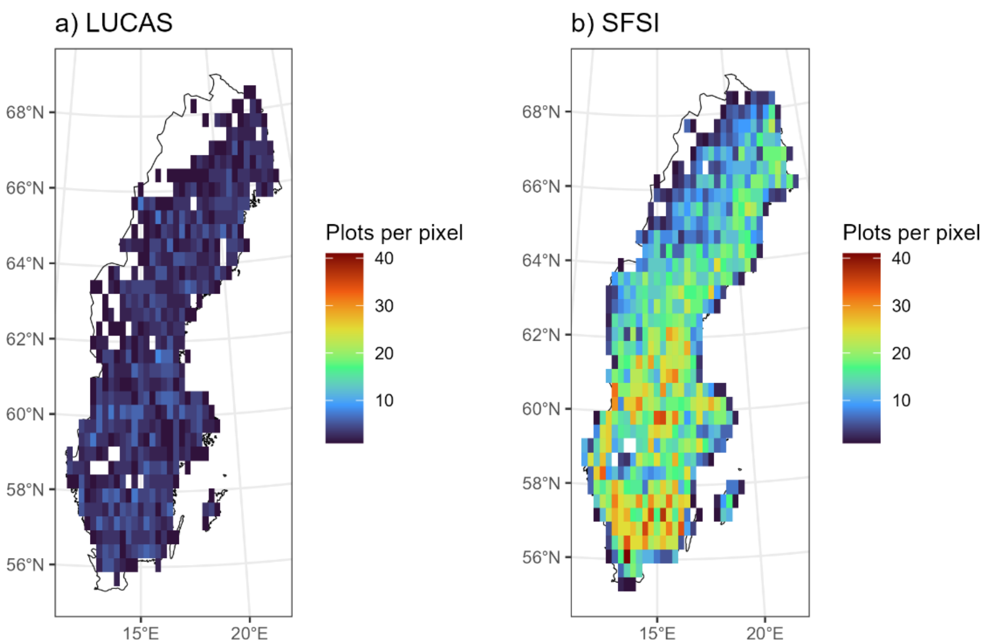


Figure 5. Point density of LUCAS 2018 (left) and SFSI 2013-2022 (right) sample plots in forest land over the area of Sweden. The colour scale indicates the number of plots per pixel. Empty (white) pixels contain no sample plots.

Coverage of different soil types

Figure 6 and Figure 7 show the monitoring coverage minus the reference coverage (in %) of soil types based on FAO and WRB soil maps – in other words, the difference between the percentage of sample plots per (dominant) soil type and the share of each soil type in Swedish forests. A negative difference indicates an underrepresentation and a positive difference an overrepresentation of the soil type.

For the SFSI, all sample plots (including those without top 20 cm averages) are considered and the plots are weighted using the areal weights depending on the local grid size. This results in a very good coverage of soil types with differences staying below 1 % for all soil types in both maps. The soil coverage in LUCAS is less uniform and Podzols are underrepresented by more than 4 %. In turn, Regosols, Cambisols, Leptosols (WRB), and Lithosols (FAO) are overrepresented by 1-2 %.

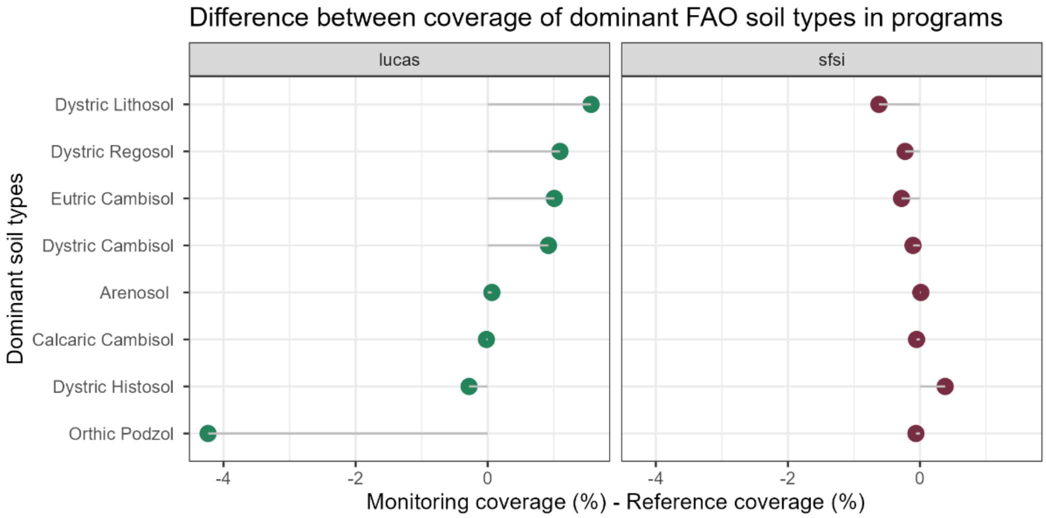


Figure 6. Difference between monitoring and reference coverage (in %) of dominant FAO soil types based on the FAO soil type map included in the European Soil Database v2.0 Raster Library (Van Liedekerke et al. 2006). The monitoring coverage was assessed by extracting the soil type from the map at the coordinates of the sample plots (SFSI 2013-2022, LUCAS 2018).

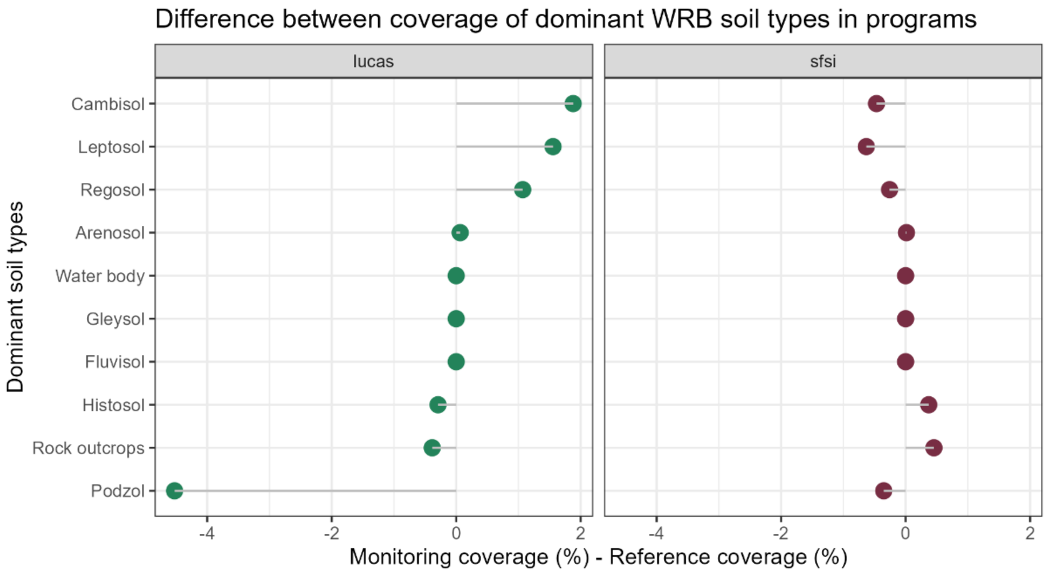


Figure 7. Difference between monitoring and reference coverage (in %) of dominant WRB soil types based on the WRB soil type map included in the European Soil Database v2.0 Raster Library (Van Liedekerke et al. 2006). The monitoring coverage was assessed by extracting the soil type from the map at the coordinates of the sample plots (SFSI 2013-2022, LUCAS 2018).

3.2.2 Comparison of selected soil properties

The soil properties selected for a closer analysis were organic carbon, nitrogen, pH and soil texture. A detailed breakdown of descriptive statistics (mean, SD, median, quartiles, min, max, n) for all these variables per NUTS 2 region is available in Appendix 1. All values shown represent the top 20 cm of the soil, calculated for the SFSI as described in chapter 2.4.1. A short overview of the national means and standard deviations is included in Table 5.

Table 5. Overview table showing national mean and standard deviation and coefficient of variation of the analysed soil properties for the SFSI 2013-2022 (n = 2765; number of tracts) and LUCAS 2018 (n = 1550) in the top 20 cm of soil.

Property [unit]	**SFSI**			**LUCAS**		
	Mean	SD	CV	Mean	SD	CV
OC [g kg^{-1}]	157.96	153.04	0.97	132.47	148.48	1.12
N [g kg^{-1}]	5.72	5.62	0.98	5.53	5.90	1.07
pH [-]	4.37	0.49	0.11	4.53	0.51	0.11
Sand [%]	71.98	18.57	0.26	56.45	18.86	0.33
Silt [%]	24.32	13.02	0.54	34.83	14.61	0.42
Clay [%]	3.70	6.24	1.69	8.70	8.40	0.97

The mean values are similar between the SFSI and LUCAS for N and pH, while OC is higher in the SFSI. The texture values are less comparable. The variation in the data is very high for OC, N and clay with CVs around 1 or even higher. In the OC and N concentrations the SFSI achieves a slightly smaller CV than LUCAS. The measured pH is less variable with a CV of 0.11 in both surveys.

Figure 8 shows the distributions of the selected properties. Sand, silt and clay are not continuously distributed in the SFSI but show multiple peaks. This is a result of the conversion of the categories assessed in the field to numerical values (see chapter 2.4.1). Silt and clay appear to be underestimated when compared to the LUCAS data determined in the laboratory, while sand is overestimated. The distribution of N, OC, and pH are more comparable between the two inventories. Notably all three properties have a (stronger) second peak in the SFSI.

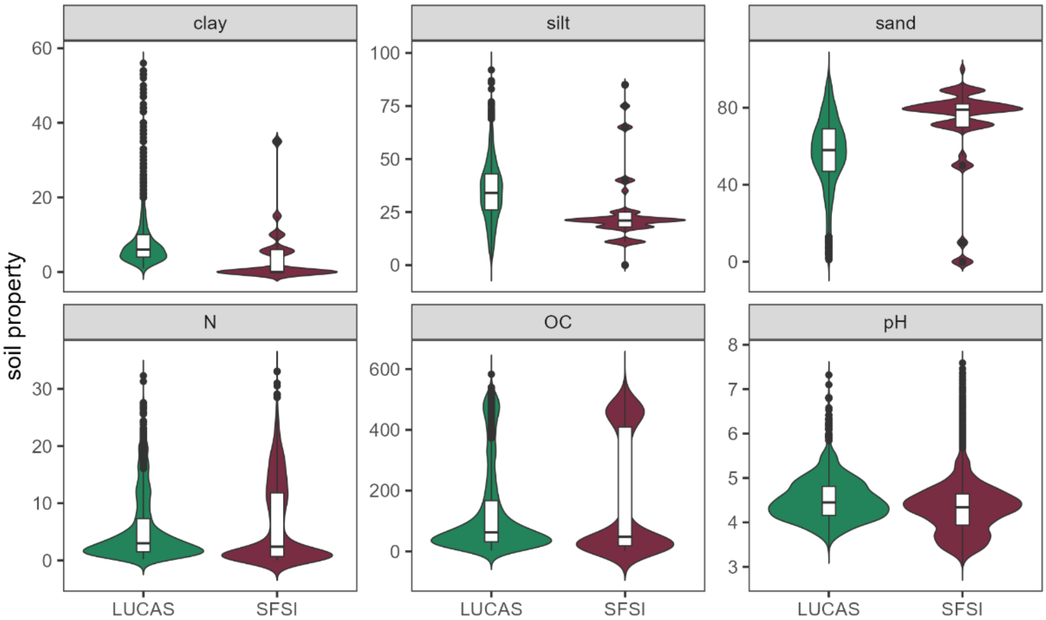


Figure 8. Violin plots overlayed with box plots, showing the distribution of clay, silt, sand [%], N [g kg^{-1}], OC [g kg^{-1}], and pH in the SFSI 2013-2022 and LUCAS 2018 in the top 20 cm of soil.

3.2.3 Examining the relationship between sample points in close proximity

To assess whether points from different soil surveys that are relatively close to each other are correlated in any way, all LUCAS points in 5 km around an SFSI point were identified using a buffer and linked to that SFSI plot. The OC, N and pH values for those point pairs are plotted against each other in Figure 9.

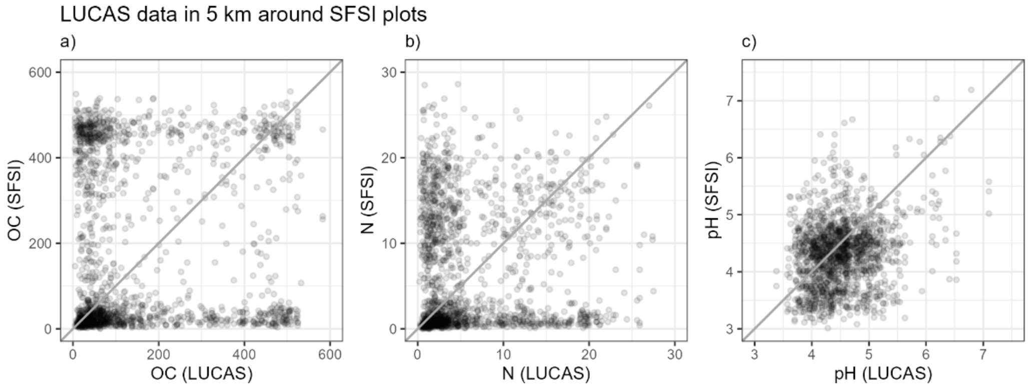


Figure 9. Comparing the OC [g kg^{-1}], N [g kg^{-1}] and pH values of SFSI plots (y-axis) from the campaigns 2013-2022 to LUCAS 2018 plots within a distance of 5 km (x-axis).

Generally, there is no detectable correlation between spatially close sample plots for any of the three soil properties. The points in Figure 9a are laid out in a square with clusters in each of the corners. This is due to the sampled soils broadly falling into the categories mineral soils or peat soils with either very low or very high organic carbon content. When the SFSI point and the associated LUCAS point are both mineral soils they lie in the lower left corner of the graph and when they are both peat soils they lie in the upper right corner. If they have different soil types they lie in one of the other two corners.

For N (Figure 9b), the graph looks similar to OC, however the upper right cluster is missing and generally the points are scattered a bit more diffusely. The pH (Figure 9c) values appear to be more or less randomly distributed with most of them around pH 4-5.

3.3 Optimized stratified sample based on JRC-script

As described in chapter 2.5.1, there were three different tests carried out with the JRC-designed script to create an optimized sample. The three tests used different domain definitions. Domains are analogous to the soil units, prescribed by the SML. The number of domains per test and the resulting sample sizes and strata are shown in Table 6 (see Appendix 1, Figure 14-16, for a more detailed comparison of the result of the three tests).

Table 6. Different domain definitions in the tests run with the JRC-script and resulting number of domains, sample sizes and number of strata for Sweden (all land use classes).

Test	Domain definition	Number of domains	Sample size	Number of strata
1	NMD, NUTS 1 and 2, soil regions	134	3990	336
2	NMD, NUTS 1 and 2	20	941	72
3	NMD, NUTS 1, soil regions	112	3376	291

The resulting sample strongly depends on the number of domains. The sample size was highest in test 1 with 3990 samples from 336 optimized strata, slightly smaller in test 3 at 3376 samples from 291 strata and much smaller when not using the soil regions for defining the domains in test 3. In that case, the total sample size is only 941 because of the much smaller number of domains (20 compared to 134 in test 1). The optimized sample size in forest land for test 1 is 1390. That is slightly smaller than the number of LUCAS 2018 sample plots in forest land (1550) but far below the sample size of the SFSI cycle 2013-2022 (8743).

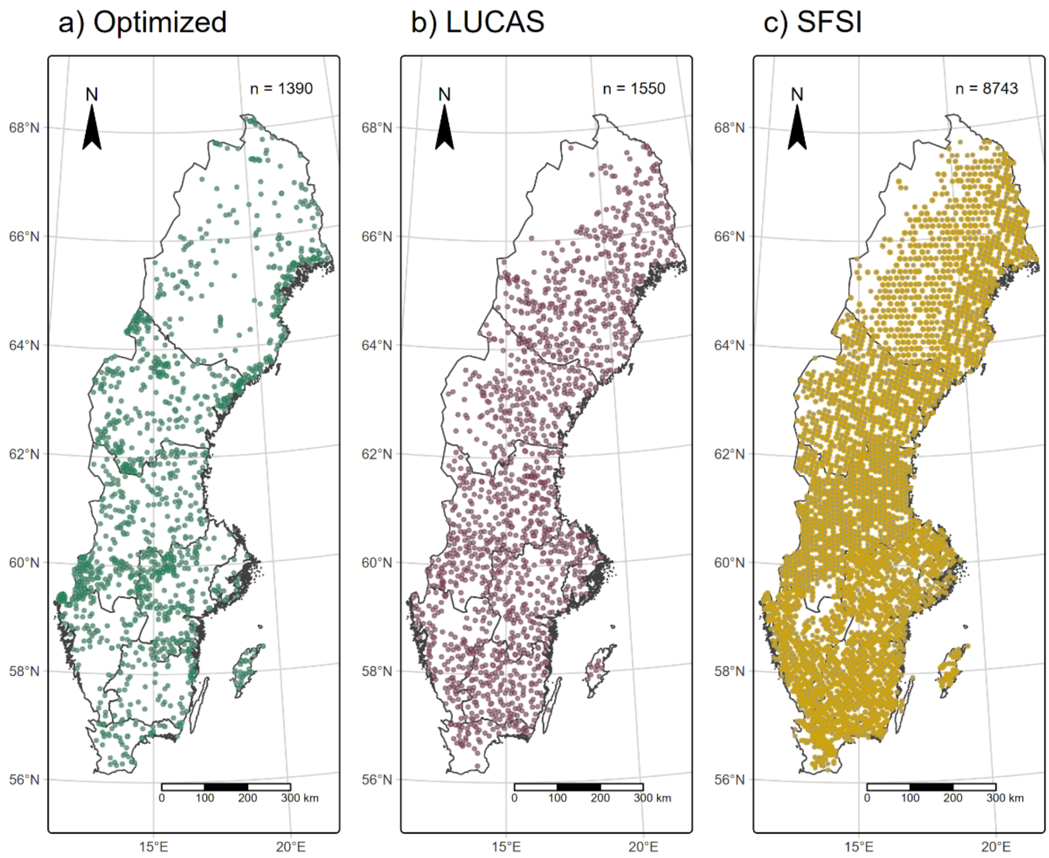


Figure 10. Mapped sample points suggested by the optimization script (test 1; a) compared to the LUCAS 2018 (b) and SFSI 2013-2022 (c) sample points. The borders of Sweden's NUTS 2 regions are shown as well.

The sample suggested by test 1 is shown in Figure 10, compared to the LUCAS and SFSI sample points. The density of sample points varies a lot from small areas with a very high number of samples to large regions which are barely sampled at all (e.g. Halland in the south-west, and most of Upper Norrland). This unequal clustering of the points is caused by the incorporation of the soil regions in the domain definition. The sample density is especially high in areas with smaller soil regions (see Figure 2 and Figure 12).

3.3.1 Region-wise comparison of suggested sample size

When comparing the sample density by NUTS region (looking at the same combination of NUTS 1 and 2 regions that was used to define the domains in tests 1 and 2), the optimized sample result shows a point density comparable to LUCAS but much smaller than the SFSI in all five regions (Figure 11). All three sample sets clearly show the lowest density in Upper Norrland. The region with the densest sampling network, however, is different between all of them. For the SFSI, it is

Southern Sweden, for LUCAS it is Middle Norrland and for the optimized sample it is Eastern Sweden.

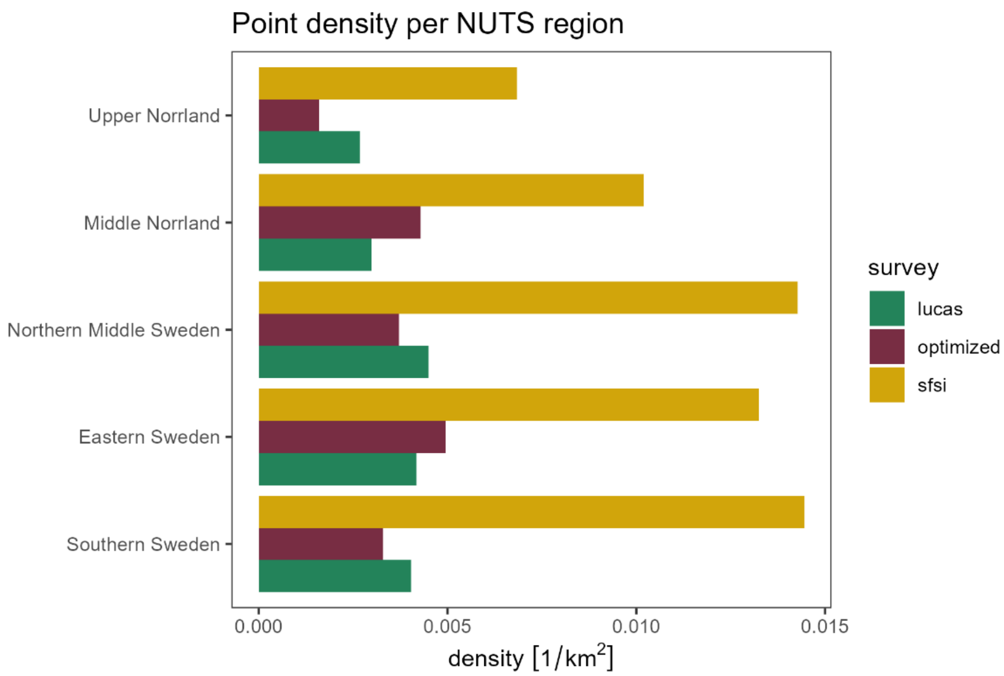


Figure 11. Point density per NUTS region (combination of NUTS 1 and 2) of the sample plots from the SFSI 2013-2022, LUCAS 2018 and the optimized sample proposed by the JRC script (test 1) The regions are in order from north (top) to south (bottom).

In Figure 12 the point density per soil region (see Figure 2) is shown. It varies quite a lot between soil regions for all three sets of plots but is clearly the most variable for the optimized sample. It has the highest sampling density in the smallest soil regions (nr. 30 with 301 km^2, nr. 11 with 677 km^2, and nr. 57 with 983 km^2) and the lowest in the biggest soil region (nr. 18 with 110 291 km^2). The point density in the SFSI and LUCAS seems not to be correlated as much with the soil regions size.

In soil region 8, the optimized sample density is much higher than in LUCAS and the SFSI. This region covers the mountainous region in northern Norway but also the northernmost tip of Sweden. There is very little forest land up there, hence there are no LUCAS sample points, and only one SFSI plot. However, the optimization script assigns 27 sample point to that soil region.

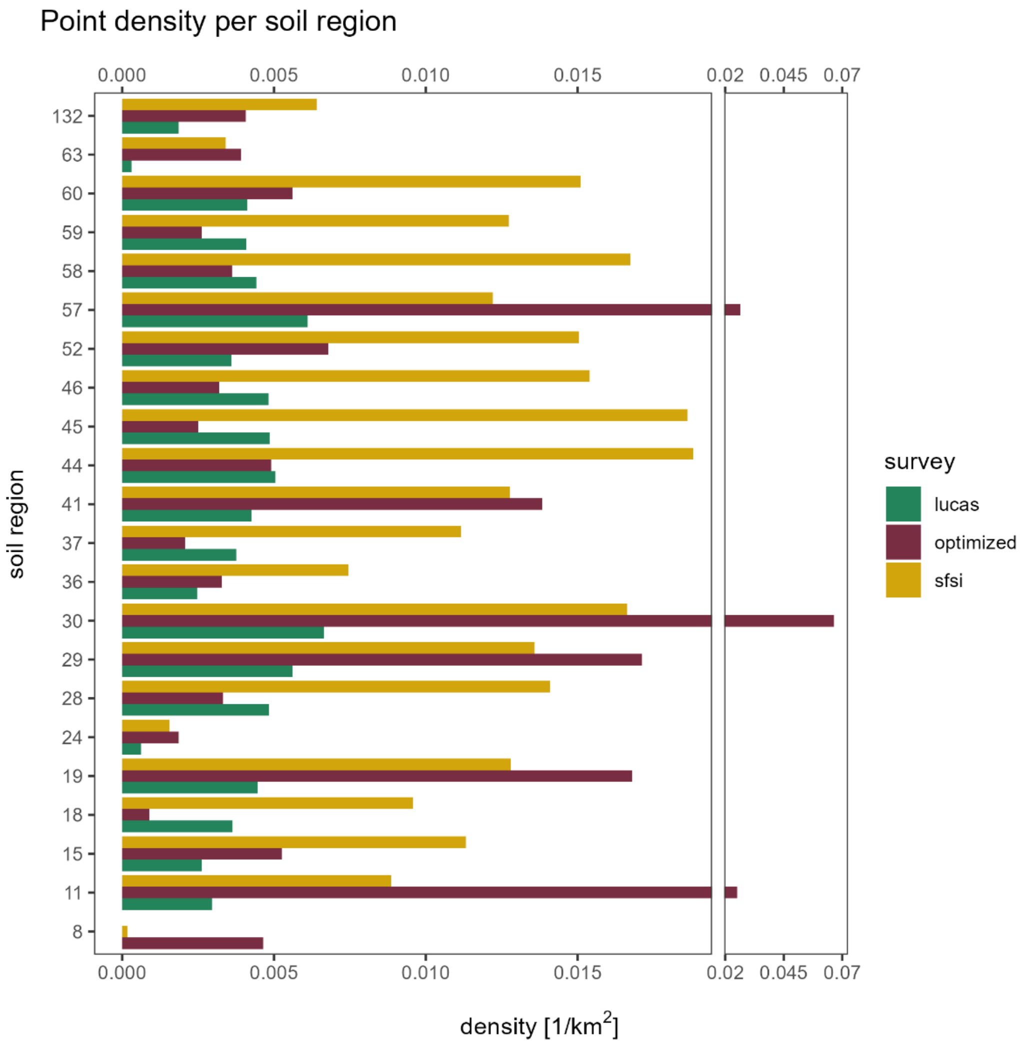


Figure 12. Point density per soil region of the sample plots from the SFSI 2013-2022, LUCAS 2018 and the optimized sample proposed by the JRC script (test 1).

3.4 Coefficient of variation in the SFSI by region and humus type

The selection of soil districts and units that the Swedish government chooses to report its soil data in does not need to match one of the domain definitions shown here. The SML grants some flexibilities, including using higher quality national data instead of the raster maps supplied by the JRC and the map of soil regions. One example of a simple definition of soil units in forest land, would be to divide the forest land in NUTS regions (the soil districts) with two soil units each, soils with peat layer and soils without. Soils with a peat layer may also include soils not classified as peat soils (Histosol), because the soil type classification is based on the mineral soil underneath when the peat layer is less than 40 cm thick.

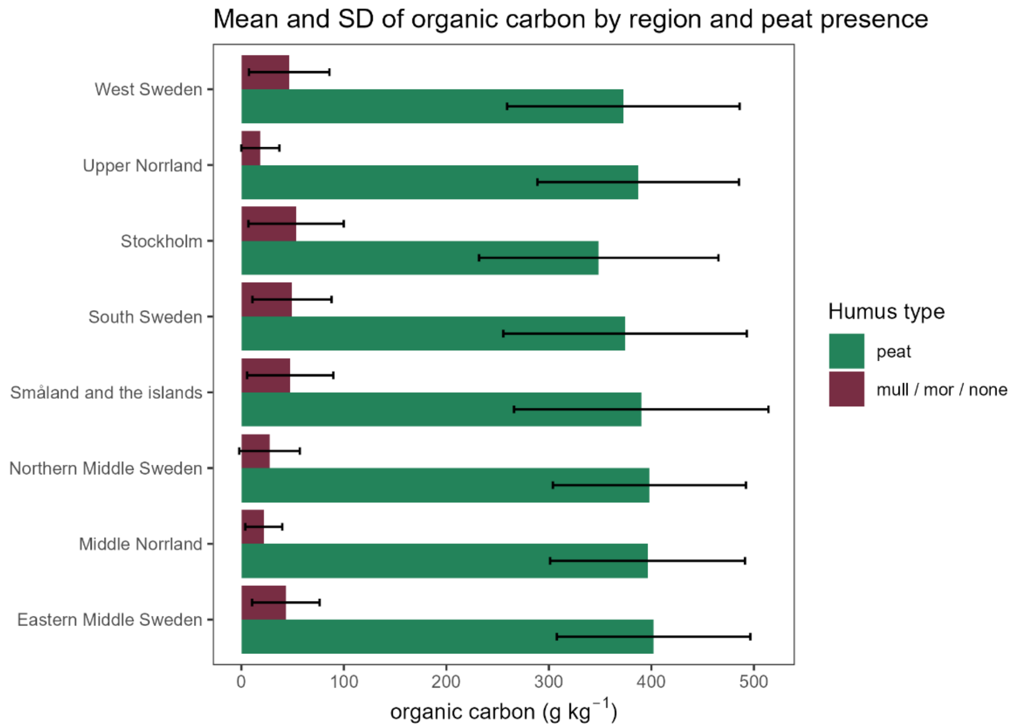


Figure 13. Mean organic carbon values incl. error bars representing the standard deviation (SD) of SFSI 2013-2022 plots per NUTS 2 region and separated into soils with and without peat layer.

The presence of a peat layer is an important predictor of the OC (and also N) concentration, as shown in Figure 13. Soil with a peat layer average 350-400 g kg^{-1} of OC in the upper 20 cm, whereas soils with other humus types have mean OC concentrations of around 50 g kg^{-1} or less. The standard deviations are higher in soils with a peat layer but in combination with the much higher mean values results in lower CVs compared to soils with no peat layer (Table 1). In most cases, the variation in the data is also smaller than without the division by peat layer (see Appendix 1), except for soils without peat layer in Northern Middle Sweden and West Sweden.

Table 7. Coefficients of variation for organic carbon and nitrogen concentrations, and pH (top 20 cm of the soil), divided by NUTS 2 region and soils with and without peat layer.

Region	**OC**		**N**		**pH**	
	peat	other	peat	other	peat	other
Eastern Middle Sweden	0.23	0.76	0.34	0.79	0.17	0.11
Middle Norrland	0.24	0.83	0.36	0.78	0.16	0.08
Northern Middle Sweden	0.24	1.07	0.34	1.23	0.10	0.08
Småland and the islands	0.32	0.88	0.32	0.92	0.12	0.16
South Sweden	0.32	0.78	0.27	0.87	0.16	0.12
Stockholm	0.34	0.87	0.33	1.00	0.24	0.10
Upper Norrland	0.25	1.00	0.38	1.18	0.12	0.05
West Sweden	0.30	0.84	0.33	0.94	0.10	0.08

4 Discussion

4.1 Comparability of SFSI and LUCAS

Comparing the two data sets which apply different methodologies in sample design, sampling procedure and measurement of soil properties required certain adaptations and assumptions. These assumptions – e.g. which humus layers to include in the SFSI data – result in limitations which are important to consider when interpreting the data. In this project, the SFSI data had to be heavily adapted to be comparable to the LUCAS data. This was only possible because the SFSI uses a sophisticated, clearly explained methodology which allows working with the data in a flexible way. Sampling the soil profile layer by layer down to a mineral soil depth of 65 cm makes it possible to examine organic and mineral soil separately, to study changes along the soil profile, or – as done here – to combine individual layers of interest, while excluding others.

The LUCAS methodology on the other hand is rather simple, just sampling the upper 20 cm of mineral soil (or peat) in form of a mixed composite sample. This saves sampling cost and time but also delivers less detailed information about the sampled soils.

Loss of information by excluding mor humus

An important assumption to achieve comparability between the data sets was that mor humus is excluded from the sampling in LUCAS. This assumption was based on the limited information provided in the reports describing the methodology (Tóth et al. 2013; Jones et al. 2020; Fernández-Ugalde et al., 2022; see chapter 2.2.4) and the comparison of the distribution of OC in different versions of the SFSI data to the LUCAS surveys 2015 and 2018 (chapter 3.1).

Therefore, all mor humus layers sampled in the SFSI were excluded from the analysis and only the mineral soil was considered as part of the top 20 cm. When comparing the distribution in Figure 3a), where all humus layers are included in the SFSI data, to the other SFSI distributions (Figure 3b-d) it is clear that mor humus is a very significant carbon pool. Examining the data more closely reveals that mor is the most common humus form (5211 out of 8743 plots) and can grow very thick with a mean depth of 8.2 cm (range: 1-63 cm), with high concentrations of OC (mean: 377 g kg^{-1}) and N (mean: 11.2 g kg^{-1}), and low pH (mean pH: 3.84). By removing this humus layer, the LUCAS Soil Survey ignores an important pool of organic matter, which is not only a very common feature of boreal forest soils, but also relevant to greenhouse gas inventory reporting under the Paris Agreement as the litter pool (Swedish Environmental Protection Agency 2025). This crucial omission indicates that the LUCAS methodology has been designed with primarily agricultural soil in mind, which are also sampled at a higher density than forest soils (Tóth et al. 2013).

This issue and resulting inconsistencies in the sampling of woodland soils, especially in Scandinavia, have been acknowledged by Jones et al. (2021), who proposed changes to the field sampling protocol for LUCAS 2022 to address this. Since neither data nor exact methodologies for the LUCAS 2022 survey have been published yet, it is unclear at the time of writing how these changes were implemented.

Consequences of top 20 cm averaging for the analysis of SFSI data

Different layers sampled separately in the SFSI had to be combined and averaged in order to generate OC, N and pH values comparable to LUCAS (chapter 2.4.1). This approach has a few drawbacks which are important to be aware of, when drawing conclusions from the presented results. While it allows a relatively accurate comparison to the LUCAS data, a significant amount of data that the SFSI collects is disregarded here. The SFSI has not been designed to be used in this way, where samples from the humus and the mineral soil are combined on plot level. Instead, the two soil layers are intended to be investigated separately. This design also allows to compute variables required for reporting at national scale, such as carbon stocks in the litter and soil pools for the greenhouse gas inventory (Swedish Environmental Protection Agency 2025). This is not possible with the LUCAS Soil Survey[9], because it ignores the litter pool by omitting mor humus and covers only the upper 20 cm of the soil and does not measure the carbon stored further down the soil profile. For the greenhouse gas inventory, carbon stocks representing at least 30 cm of the soil are required (IPCC 2006).

Despite the complex system of cases to derive top 20 cm averages (see Table 3), on many plots the collected SFSI data was not sufficient to calculate values that are representative of the part of the soil that would have been sampled by LUCAS at the same location. Often this was the case when only humus sampling and no pit digging took place at a plot. These plots could only be considered when the organic layer depth was ≥ 20 cm. Otherwise C and N concentrations would be over- and pH underestimated, because information from the mineral soil below is not available. This problem is further amplified by the omission of mor type humus. When the humus form was mor and only humus sampling took place, a plot automatically had to be discarded. In addition, there was a small number of plots where a sample required for the calculation of the averages was missing for an unknown reason. Likely, the sample has been lost somewhere on the way from the field to the analysis in the lab. On plots where this was the case, a top 20 cm average is missing as well. In total, these issues lead to 3884 plots being excluded from the analysis, the combined area weight of which is 121 674 km^2 or 45 % of all the area covered

[9] Bulk density, which is required for stock calculation, has only been recorded on a subset of plots in 2018.

by the SFSI (273 108 km^2). An average of the top 20 cm could only be calculated on 4859 out of 8743 plots visited from 2013-2022. This might bias the SFSI results towards higher OC and N concentrations, since the majority of the excluded plots are those where only humus sampling took place and the humus layer was shallower than 20 cm, while plots with a thick (non-mor) humus layer remain in the data set. However, through the averaging of the plots at tract level and the exclusion of mor humus this effect is reduced.

4.2 Strengths and weaknesses of individual data sets

Sampling densities and designs

The sampling density of the SFSI is much higher than of LUCAS with in total 8743 sampling points (inventory cycle 2013-2022) compared to 1550 (LUCAS 2018) in forest land. On about 60 % of the SFSI's sampling points (5139) only the humus and not the mineral soil has been sampled. The number of plots with mineral soil sampling was 3604 (out of 4101 plots with planned mineral soil sampling), which is still 133 % higher than the LUCAS 2018 sample size. The number of sites was higher in LUCAS 2009 with 2255 samples taken. Why the sample size was reduced is unclear but might be due to the EU-wide redistribution of sample sites from 2015 onwards to include areas above 1000 m elevation (Jones et al. 2020), which only make up a small proportion of Sweden's territory.

In context of the sampling densities, the sampling designs are also important to consider. The SFSI is based on systematic random sampling, using five different grid sizes and orientations in different regions of Sweden. Systematic random sampling ensures optimal spatial coverage and often delivers more accurate results than other random sampling designs according to de Gruijter et al. (2006).

Stratified random sampling, as used in the design of the LUCAS Soil Survey, is generally very efficient and can achieve a smaller sampling variance at equal cost than simple random sampling. However, this improvement in efficiency depends on the quality of the stratification, i.e. how homogeneous the individual strata are (de Gruijter et al. 2006), which in turn depends on the quality of the data used to define the strata. Inappropriate stratification may even lead to a loss in efficiency (de Gruijter et al. 2006). Systematic random sampling also generally results in a more precise estimate of the mean than stratified random sampling (Brus 2022), calling the application of the latter further into question – especially when measuring a wide array of variables, which may not be correlated.

Coverage of different soil types

Another effect of the stratified random sampling used by LUCAS is that the soil type coverage is not as uniform over the country as in the SFSI. Podzols are underrepresented by LUCAS by more than 4 % and several other dominant soil types are slightly overrepresented (see Figure 6 and Figure 7). In comparison, the difference between monitoring and reference coverage of dominant soil types is less than 1 % for the SFSI regarding all FAO and WRB soil types.

Since Podzols are classified as the dominant soil type on 80 % of Swedish forest land in the WRB and FAO soil type maps, it is to be expected that the sample allocation in stratified random sampling is biased towards soil types with less areal coverage. It is important to consider though, that these maps only show the dominant soil types in a region, therefore the allocation of fewer sample plots in regions with Podzols might lead to underrepresenting smaller scale non-dominant soil types which are often found in Podzol-dominated areas, e.g. Histosols. This might be the cause of the underestimation of OC and N concentrations by LUCAS compared to the SFSI (see Table 5, Figure 8, and Appendix 1).

Regional means and standard deviations of OC, N and pH

Generally, the mean values of OC, N and pH calculated from SFSI and LUCAS data are similar on regional and national level (see Appendix 1). There are a few exceptions, for example OC in the Stockholm region (SFSI: OC = 125.07 ± 126.94 g kg^{-1}; LUCAS: OC = 78.97 ± 44.06 g kg^{-1}), and Upper Norrland (SFSI: OC = 148.03 ± 154.60 g kg^{-1}; LUCAS: OC = 88.62 ± 131.88 g kg^{-1}) where LUCAS drastically underestimates the OC concentration compared to the SFSI.

Standard deviations are high for OC and N, with coefficients of variation around 1.0 in most regions. The pH is significantly less variable with coefficients of variation at 0.11 at national scale for both surveys.

It has to be noted here that there exists no unbiased estimator of the sampling variance for a systematic random sampling design (Brus & Saby 2016). A simple random sampling variance estimator can be used (as done here) but Brus & Saby (2016) showed that it overestimates the variance, using the French soil monitoring network as an example. At the same time, the shown standard deviations for the SFSI are based on the tract-level averages, ignoring the variance within the tract, potentially cancelling out the effect of using the simple random sampling estimator.

In theory, calculating the regional means and variances of the LUCAS Soil Survey also requires a different approach, taking into account the stratification as described in chapter 2.4.2. However, since the LUCAS data sets do not provide information about the used stratification, the means and variances are not directly comparable between the two programmes, and the results from LUCAS might deviate from the true values.

More precise soil texture measurement in LUCAS

The presented soil texture values from LUCAS and the SFSI are less comparable than OC, N and pH. This is because in the SFSI soil texture is not analysed in the lab by soil sieving or laser diffraction but only categorized in the field based on the shape and rollability of the mineral soil (SLU 2024a). These categories had to be translated into numerical values as described in chapter 2.4.1. Despite this conversion, in Figure 8 it is still clear that the sand, silt, and clay fractions in the SFSI are based on categorical data, i.e. they are not smoothly distributed like the LUCAS data but show various spikes, which represent the initial categories.

Overall, the SFSI shows much higher mean sand (72.0 %) and lower silt (24.3 %) and clay (3.7 %) fractions compared to LUCAS (56.5 %, 34.8 %, 8.7 %). The most probable explanation for this is the imprecise conversion of the categorical SFSI field data. But also the underestimation of Podzols by LUCAS could be an explanation for the lower mean sand fraction compared to the SFSI (see Figure 8 and Table 5) as the topsoil in Podzols is characterized by a sandy single grain structure (Blume et al. 2016).

There is currently a project underway at the Swedish University of Agricultural Sciences (SLU) to measure the particle size distribution on a subset of SFSI samples to improve the translation of the categorical data to numerical values. Since soil texture change happens very slowly without external disturbances, soil texture measurements would not need to be repeated with every inventory cycle. Soil texture measurement is also not repeated on revisited plots in LUCAS (Jones et al. 2020; Fernández-Ugalde et al. 2022). Nonetheless, for the time being LUCAS delivers more precise soil texture data for Swedish forest soils.

4.3 Assessment of SML requirements

Assessing whether LUCAS or the SFSI fulfil the requirements of the SML for the monitoring of Swedish forest soils proved difficult. The main challenge is that at the time of writing the final text of the directive is not known yet. This leads to uncertainty over the phrasing of some key passages, where the wording differs between the original proposal of the Commission (European Commission 2023) and the positions of the Parliament (European Parliament 2024) and the Council (Council of the European Union 2024). In addition, the output of the R script supplied to Member States by the JRC to design an optimized stratified sample, is not guaranteed to produce a sample set compliant with the SML. The comparison between this suggested optimal sample and the existing surveys therefore also cannot provide a clear answer. These issues are elaborated on in the following chapters.

4.3.1 Interpretation of the SML requirements

As presented in the introduction (chapter 1.1.1), the most important passages defining the requirements for Member States' soil monitoring systems are the ones relating to the establishment of soil districts and soil units, and the precision constraint of "a maximum percent error (or Coefficient of Variation) of 5 % for the estimation of the area having healthy soils" (European Commission 2023) or to “represent the variability of the chosen soil descriptors within the soil units” (Council of the European Union 2024).

What makes both of these wordings difficult to interpret, is that the terms “coefficient of variation” and “maximum percent error” are not clearly defined in the text. A common definition of the CV which may apply can be found on the Eurostat website:

> The coefficient of variation is generally defined as the standard deviation of a random variable divided by the mean. (Eurostat n.d.)

The (maximum) percent error most likely refers to the relative difference between an observed and a true value (more commonly known as the relative error), calculated using the formula:

$$Percent\ error = \frac{|v_{true} - v_{observed}|}{v_{true}} * 100 \qquad (6)$$

This is problematic, because these two statistical terms do not describe the same metric. The CV, as defined by Eurostat, can be used as an easily comparable description of the variance in the data. Thereby, a plausible interpretation of the Council’s text would be that the monitoring systems are required to define soil units and allocate the sampling size in such a way that the coefficient of variation in the data stays below 5 % for all measured soil properties. This would be a tall ask, however, considering that the coefficients of variation for OC and N are around 100 % on regional level in both, SFSI and LUCAS data. These properties are highly variable by nature (Brejda et al. 2000) and even a drastic increase in sampling size is unlikely to reduce them to 5 %. One option might be to look at organic and mineral layers separately and/or apply very narrowly defined soil units to achieve such low variation in the data.

These strict requirements regarding the variability in the data would bring about very high sampling costs though. Therefore, the more likely interpretation might be that the means of the measured soil properties shall not deviate from the actual mean in the soil unit by more than 5 %. This should be a more easily reachable goal but of course, the actual mean per soil property in a soil unit is not known. That means, estimating a relative error of the SFSI and LUCAS poses a statistical challenge that is beyond the scope of this project.

The main take-away is, that neither of the two phrasings from the Commission and the Council are statistically unequivocal and an accurate assessment of the sampling requirements will only be possible after the final text will be agreed and published.

4.3.2 Challenges in optimizing a spatial stratified random sampling scheme

The true distribution of soil properties is not known

To ensure appropriate stratification, the variability and distribution of the target variables or of correlated ancillary variables need to be known (de Gruijter et al. 2006). In the approach that is proposed in the R script, which was supplied to the Member States by the JRC, Europe-wide raster maps at 100 x 100 m resolution of seven soil properties were used as input data (see chapter 2.5.1). The maps are based on LUCAS data and were created using hybrid approaches like regression kriging (Ballabio et al. 2016, 2019). Therefore, the maps cannot represent the variability of Swedish forest soils more precisely than the LUCAS Soil Survey they are based on and are likely far less precise than maps using the SFSI data as a basis. However, these maps (or rather a 90 000 pixel-sized sample of them) are treated as the "true distribution" of soil properties in Sweden in the R script provided by the JRC and the accuracy of the prediction by the optimized sample is evaluated against those maps. In this step, the variance introduced by the LUCAS Soil Survey and the models used to create maps is not considered, meaning that it is unlikely that the resulting optimal sample truly meets the 5 %-criterium.

In addition, the maximum percent error constraints in the *SamplingStrata* package (Barcaroli 2014) are implemented in such a way that they test the difference between the predicted mean and the actual mean value per soil property in each soil unit. If the final SML text deviates from that definition of the "5 % maximum percent error", this approach will need to be reevaluated.

The best data source about the distribution of various soil properties in Swedish forests that is currently available is the SFSI. Therefore, SFSI data should be used as a basis in future efforts to either design a new sampling programme or adapt the SFSI according to the requirements of the SML.

A statistical analysis of the medium- to small-scale variation of soil properties in Swedish forests could aid in estimating the error of regional and national mean values predicted by the SFSI. A suitable data set could be soil data from the Krycklan catchment study (Larson et al. 2023, 2024), where the soil in a 68 km^2 boreal catchment has been sampled on 430 plots using SFSI methodology. By performing an analysis of how many of those plots are needed to achieve mean values within 5 % of the mean of all 430 plots, one could estimate the required sample density to achieve a 5 % error on national scale.

The soil unit definition is critical for the sample size and distribution

The biggest effect on the sample size that the optimization script suggested had the exclusion of the soil regions from the soil unit (or domain) definition (see Table 6). The resulting sample size strongly depends on the number and size of the domains, with bigger domains receiving much smaller sampling densities than smaller areas (see Figure 10, Figure 11, and Figure 12). Using soil units with such varying sizes is therefore probably not sensible when using the JRC approach and an optimized stratified sampling regime because it results in a very unequal coverage of samples across the country. Possible solutions could be to aggregate smaller soil regions with similar characteristics or use national soil type maps based on SFSI data. It is also important to ensure that the data that is considered for the delineation of soil units is actually correlated with the variables of interest. One example would be to treat soils with and without peat layers as separate soil units, as shown in chapter 3.4. The existence of a peat layer results in high OC and N concentrations and low pH values in the top 20 cm of soil. That way the variance in the data can be reduced, and lower sample sizes may be possible.

4.3.3 Comparing the SFSI and LUCAS to the optimized sample

One way of assessing the fulfilment of the SML requirements by the SFSI and LUCAS is to compare it to the optimized sample, assuming that the approach used by the JRC is consistent with how the 5 % requirement should be interpreted. The SFSI has a higher sampling density in all NUTS regions and almost all soil regions than the optimized sample (see Figure 11 and Figure 12). The sampling density of the optimized sample differs much more strongly between soil regions than the two existing surveys and is inversely related to the size of the soil region. The algorithm in the script distributes similar numbers of samples to soil regions of very different sizes. Since the required sample size is directly tied to the variation of soil properties in the soil units, this indicates that the variation is not reduced significantly by introducing smaller soil units based on the soil regions. It also leads to a suboptimal coverage of the country and clustering of sample points in small soil units (see Figure 10).

LUCAS and the optimized sample have similar sample sizes (1550 and 1390) but the sampling densities also differ strongly by soil region, for the reasons listed above. With the definition of the soil units the optimization script used, the LUCAS data set might struggle to fulfil the maximum percent error requirement of the SML. But this cannot be said with high confidence as the assessment of the error in the script is based on raster maps, without considering their variance (see chapter 2.5.1).

4.4 Added value of the Soil Monitoring Law

Despite remaining uncertainty about the exact requirements Member States will have to fulfil under the SML and criticism about the insufficient inclusion of forest soils (Wellbrock et al. 2024), proposed soil health indicators (Mäkipää et al. 2024), and a lack of protective measures (Kotschik et al. 2024) the implementation of the SML has the potential to drastically improve our knowledge of the state of soil across Europe. Not all Member States do have an operating national soil monitoring system in place and those that do, differ in design, sample size, and methodology (Froger et al. 2024). Harmonizing national soil monitoring systems across Europe including the LUCAS Soil Survey will be one of the biggest challenges in the implementation of the SML. But its success will determine the effectiveness of future policies to halt soil degradation and improve soil health (Panagos et al. 2024). In addition, a well-designed EU-wide soil monitoring system will support EU policies on a wider range of environmental problems like climate change mitigation, by enabling countries to observe the carbon balance in soils across all land use classes (Bellassen et al. 2022).

5 Conclusion

5.1 Suitability of SFSI and LUCAS for SML sampling

The issues discussed in the previous chapters make it difficult to come to a clear conclusion about the suitability of the SFSI and LUCAS for sampling under the SML. Whether the SFSI's sampling network is dense enough to fulfil the requirements strongly depends on how the passage introducing the 5 % error condition is worded in the final text. In case the Council's wording remains in the published directive, achieving coefficients of variation < 5 % would likely be impossible – even with much more sampling sites – because of the naturally high variability of soil properties like OC and N. Meeting a maximum percent error (of the predicted mean) of 5 % per soil unit would likely be easier to achieve, but depends on which reference is used as the true mean.

Currently there exits not better data source about the distribution of various soil properties in Swedish forest soils than the SFSI. Therefore, the SFSI's error is difficult to assess. LUCAS is less precise than the SFSI, especially because the strata used to distribute the sampling plots are not known, which means that statistically unbiased averages and variances cannot be calculated. This is a major shortcoming of the LUCAS Soil Survey when regionally aggregated values are required for reporting purposes.

However, the SFSI will fall short of the SML requirements when it comes to certain soil properties, most notably soil texture. Here, LUCAS data may be used to supplement the SFSI for the reporting, as long as no lab-measured texture values from the SFSI are available.

5.2 Recommendations for future soil monitoring

Details about what changes to the forest soil monitoring might be required in the future will only become clear once the Soil Monitoring Law has been enacted. However, based on the information available at the time of writing, the SFSI appears to be a good foundation to build on and fulfilling requirements might only require changes in the lab analysis, i.e. to include quantitative measurements of soil texture and other soil properties that are stated in the law but currently not measured at the appropriate level.

The SFSI's soil measurement network is much denser than LUCAS or the optimal sample generated by the JRC script. If the optimized sample is an accurate implementation of the SML's requirements, the SFSI should be compliant with the demands while at the same time providing a much better spatial coverage over the country.

LUCAS data can be used to supplement the SFSI, for example for soil texture, which is measured in more detail by LUCAS. With the improvements implemented in the 2022 survey, especially the introduction of humus sampling, LUCAS might become more usable to reporting and validating the SFSI data in the future.

The definition of the soil units will be a critical step which must be based on the best available data. This is currently the SFSI. It was shown that a soil unit definition based on soil regions and administrative borders (NUTS regions) does not necessarily result in a good coverage of the country when using the approach proposed by the JRC. The small-scale variation in forest soils is too high for the division in soil regions to have a meaningful effect on the variability of the soil properties. One way to reduce the variance that was identified, was separating the data in soils with and soils without a peat-layer.

References

Ågren, A.M., Hasselquist, E.M., Stendahl, J., Nilsson, M.B. & Paul, S.S. (2022). Delineating the distribution of mineral and peat soils at the landscape scale in northern boreal regions. *SOIL*, 8 (2), 733–749. https://doi.org/10.5194/soil-8-733-2022

Ballabio, C., Lugato, E., Fernández-Ugalde, O., Orgiazzi, A., Jones, A., Borrelli, P., Montanarella, L. & Panagos, P. (2019). Mapping LUCAS topsoil chemical properties at European scale using Gaussian process regression. *Geoderma*, 355, 113912. https://doi.org/10.1016/j.geoderma.2019.113912

Ballabio, C., Panagos, P. & Monatanarella, L. (2016). Mapping topsoil physical properties at European scale using the LUCAS database. *Geoderma*, 261, 110–123. https://doi.org/10.1016/j.geoderma.2015.07.006

Ballin, M. & Barcaroli, G. (2013). Joint determination of optimal stratification and sample allocation using genetic algorithm. *Survey Methodology*, 39 (2), 369–393

Barcaroli, G. (2014). SamplingStrata: An R package for the optimization of stratified sampling. *Journal of Statistical Software*, 61 (4), 1–24. https://doi.org/10.18637/jss.v061.i04

Bellassen, V., Angers, D., Kowalczewski, T. & Olesen, A. (2022). Soil carbon is the blind spot of European national GHG inventories. *Nature Climate Change*, 12 (4), 324–331. https://doi.org/10.1038/s41558-022-01321-9

Bethel, J. (1989). Sample Allocation in Multivariate Surveys. *Survey Methodology*, 15 (1), 47–57

BGR (2005). Soil Regions of the European Union and Adjacent Countries 1:5,000,000 (Version 2.0) . http://data.europa.eu/88u/dataset/ae71ffee-1ae9-4624-ae3f-f49513fe9dcb [2025-05-02]

Blume, H.-P., Brümmer, G.W., Fleige, H., Horn, R., Kandeler, E., Kögel-Knabner, I., Kretzschmar, R., Stahr, K. & Wilke, B.-M. (2016). *Scheffer/Schachtschabel Soil Science*. Springer. https://doi.org/10.1007/978-3-642-30942-7

Bradshaw, C.J.A. & Warkentin, I.G. (2015). Global estimates of boreal forest carbon stocks and flux. *Global and Planetary Change*, 128, 24–30. https://doi.org/10.1016/j.gloplacha.2015.02.004

Brejda, J.J., Moorman, T.B., Smith, J.L., Karlen, D.L., Allan, D.L. & Dao, T.H. (2000). Distribution and Variability of Surface Soil Properties at a Regional Scale. *Soil Science Society of America Journal*, 64 (3), 974–982. https://doi.org/10.2136/sssaj2000.643974x

Brus, D.J. (2022). *Spatial sampling with R*. First edition. CRC Press, Taylor & Francis Group. https://dickbrus.github.io/SpatialSamplingwithR/

Brus, D.J. & Saby, N.P.A. (2016). Approximating the variance of estimated means for systematic random sampling, illustrated with data of the French Soil Monitoring Network. *Geoderma*, 279, 77–86. https://doi.org/10.1016/j.geoderma.2016.05.016

Council of the European Union (2024). General approach on the soil monitoring law. https://data.consilium.europa.eu/doc/document/ST-11299-2024-INIT/en/pdf

Deluca, T.H. & Boisvenue, C. (2012). Boreal forest soil carbon: distribution, function and modelling. *Forestry: An International Journal of Forest Research*, 85 (2), 161–184. https://doi.org/10.1093/forestry/cps003

Dettmann, U., Frank, S., Wittnebel, M., Piayda, A. & Tiemeyer, B. (2022). How to take volume-based peat samples down to mineral soil? *Geoderma*, 427, 116132. https://doi.org/10.1016/j.geoderma.2022.116132

European Commission (2023). Proposal for a Directive of the European Parliament and of the Council on Soil Monitoring and Resilience. https://data.europa.eu/doi/10.2777/821504 [2025-05-02]

European Parliament (2024). European Parliament legislative resolution of 10 April 2024 on the proposal for a directive of the European Parliament and of the Council on Soil Monitoring and Resilience (Soil Monitoring Law). https://www.europarl.europa.eu/doceo/document/TA-9-2024-0204_EN.html

Eurostat (2024). *Statistics Explained. Forests, forestry and logging*. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Forests,_forestry_and_logging [2025-12-05]

Eurostat (n.d.). *Glossary: Coefficient of variation*. *Eurostat. Statistics Explained*. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Coefficient_of_variation [2025-05-21]

Fernández-Ugalde, O., Orgiazzi, A., Jones, A., Lugato, E. & Panagos, P. (2017). LUCAS 2018: Soil Component: Sampling Instructions for Surveyors. Publications Office of the European Union. https://data.europa.eu/doi/10.2760/023673 [2025-02-25]

Fernández-Ugalde, O., Scarpa, S., Orgiazzi, A., Panagos, P., Van Liedekerke, M., Marechal, A. & Jones, A. (2022). *LUCAS 2018 Soil Module: Presentation of dataset and results.* Publications Office of the European Union. https://data.europa.eu/doi/10.2760/215013 [2025-01-22]

Froger, C., Tondini, E., Arrouays, D., Oorts, K., Poeplau, C., Wetterlind, J., Putku, E., Saby, N.P.A., Fantappiè, M., Styc, Q., Chenu, C., Salomez, J., Callewaert, S., Vanwindekens, F.M., Huyghebaert, B., Herinckx, J., Heilek, S., Sofie Harbo, L., De Carvalho Gomes, L., Lázaro-López, A., Antonio Rodriguez, J., Pindral, S., Smreczak, B., Benő, A., Bakacsi, Z., Teuling, K., van Egmond, F., Hutár, V., Pálka, B., Abrahám, D. & Bispo, A. (2024). Comparing LUCAS Soil and national systems: Towards a harmonized European Soil monitoring network. *Geoderma*, 449, 117027. https://doi.org/10.1016/j.geoderma.2024.117027

Gazol, A., Camarero, J.J., Anderegg, W.R.L. & Vicente-Serrano, S.M. (2017). Impacts of droughts on the growth resilience of Northern Hemisphere forests. *Global Ecology and Biogeography*, 26 (2), 166–176. https://doi.org/10.1111/geb.12526

de Gruijter, J., Brus, D., Bierkens, M. & Knotters, M. (2006). *Sampling for Natural Resource Monitoring*. Springer.

Hijmans, R.J. (2025). *terra: Spatial data analysis*. https://CRAN.R-project.org/package=terra

IPCC (2006). *2006 IPCC guidelines for national greenhouse gas inventories*. Eggleston, H.S., Buendia, L., Miwa, K., Ngara, T., & Tanabe, K. (eds) (Eggleston, H. S., Buendia, L., Miwa, K., Ngara, T., & Tanabe, K., eds). Institute for Global Environmental Strategies.

Jones, A., Fernández-Ugalde, O. & Scarpa, S. (2020). *LUCAS 2015 Topsoil Survey: Presentation of dataset and results.* Publications Office of the European Union. https://data.europa.eu/doi/10.2760/616084 [2025-01-22]

Jones, A., Fernández-Ugalde, O., Scarpa, S. & Eiselt, B. (2021). *LUCAS soil 2022: ISSG planning document.* Publications Office of the European Union. https://data.europa.eu/doi/10.2760/74624 [2025-01-20]

Karltun, E., Bandh, C., Spångberg, M., Fiedler, J., Sundgren, M. & Nilsson, T. (2022). *Metodbeskrivningar av provberedning och markkemiska analyser inom Markinventeringen*. Institutionen för mark och miljö. Sveriges lantbruksuniversitet.

Kotschik, P., Princz, J., Silva, C. de L. e, Renaud, M., Marti-Roura, M., Brooks, B., Pieper, S., Rijk, I., Simini, M., Andres, S., Scholz-Starke, B. & Grenni, P. (2024). The upcoming European Soil Monitoring Law: An effective instrument for the protection of terrestrial ecosystems? *Integrated Environmental Assessment and Management*, 20 (2), 316–321. https://doi.org/10.1002/ieam.4834

Larson, J., Kuglerová, L., Högberg, P. & Laudon, H. (2024). Variability in soil carbon-to-nitrogen ratios explained by environmental conditions in a boreal catchment. *Forest Ecology and Management*, 568, 122108. https://doi.org/10.1016/j.foreco.2024.122108

Larson, J., Wallerman, J., Peichl, M. & Laudon, H. (2023). Soil moisture controls the partitioning of carbon stocks across a managed boreal forest landscape. *Scientific Reports*, 13 (1), 14909. https://doi.org/10.1038/s41598-023-42091-4

Laudon, H., Mensah, A.A., Fridman, J., Näsholm, T. & Jämtgård, S. (2024). Perspectives: Swedish forest growth decline: A consequence of climate warming? *Forest Ecology and Management*, 565, 122052. https://doi.org/10.1016/j.foreco.2024.122052

Lindén, A.G. (2002). *Beskrivning till jordartskartan 9F Finspång NO*. Sveriges geologiska undersökning (SGU.

Mäkipää, R., Menichetti, L., Martínez-García, E., Törmänen, T. & Lehtonen, A. (2024). Is the organic carbon-to-clay ratio a reliable indicator of soil health? *Geoderma*, 444, 116862. https://doi.org/10.1016/j.geoderma.2024.116862

Massicotte, P. & South, A. (2023). *rnaturalearth: World map data from natural earth*. https://CRAN.R-project.org/package=rnaturalearth

Minasny, B. & McBratney, A.B. (2006). A conditioned Latin hypercube method for sampling in the presence of ancillary information. *Computers & Geosciences*, 32 (9), 1378–1388. https://doi.org/10.1016/j.cageo.2005.12.009

Olsson, M. (1999). Soil Survey in Sweden. In: *Soil resources of Europe*. (Research Report No. 6). European Soil Bureau. Joint Research Center. 145–151.

Olsson, M., Erlandsson, M., Lundin, L., Nilsson, T., Nilsson, Å. & Stendahl, J. (2009). Organic carbon stocks in Swedish Podzol soils in relation to soil hydrology and other site characteristics. *Silva Fennica*, 43 (2). https://doi.org/10.14214/sf.207

Orgiazzi, A., Ballabio, C., Panagos, P., Jones, A. & Fernández-Ugalde, O. (2018). LUCAS Soil, the largest expandable soil dataset for Europe: a review. *European Journal of Soil Science*, 69 (1), 140–153. https://doi.org/10.1111/ejss.12499

Panagos, P., Borrelli, P., Jones, A. & Robinson, D.A. (2024). A 1 billion euro mission: A Soil Deal for Europe. *European Journal of Soil Science*, 75 (1), e13466. https://doi.org/10.1111/ejss.13466

Panagos, P., Van Liedekerke, M., Borrelli, P., Köninger, J., Ballabio, C., Orgiazzi, A., Lugato, E., Liakos, L., Hervas, J., Jones, A. & Montanarella, L. (2022). European Soil Data Centre 2.0: Soil data and knowledge in support of the EU policies. *European Journal of Soil Science*, 73 (6), e13315. https://doi.org/10.1111/ejss.13315

Paré, D., Bognounou, F., Emilson, E.J.S., Laganière, J., Leach, J., Mansuy, N., Martineau, C., Norris, C., Venier, L. & Webster, K. (2024). Connecting forest soil properties with ecosystem services: Toward a better use of digital soil maps—A review. *Soil Science Society of America Journal*, 88 (4), 981–999. https://doi.org/10.1002/saj2.20705

Pebesma, E. (2018). Simple features for R: Standardized support for spatial vector data. *The R Journal*, 10 (1), 439–446. https://doi.org/10.32614/RJ-2018-009

Pebesma, E., Mailund, T. & Hiebert, J. (2016). Measurement units in R. *R Journal*, 8 (2), 486–494. https://doi.org/10.32614/RJ-2016-061

Pedersen, T.L. (2024). *patchwork: The composer of plots*. https://CRAN.R-project.org/package=patchwork

R Core Team (2024). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Ranneby, B., Cruse, T., Hagglund, B., Jonasson, H. & Sward, J. (1987). Designing a new national forest survey for Sweden. *Studia Forestalia Suecica*, 177

Regulation (EU) 2018/841 of the European Parliament and of the Council of 30 May 2018 on the inclusion of greenhouse gas emissions and removals from land use, land use change and forestry in the 2030 climate and energy framework, and amending Regulation (EU) No 525/2013 and Decision No 529/2013/EU (2018). . https://eur-lex.europa.eu/eli/reg/2018/841/oj/eng

Romero, F., Labouyrie, M., Orgiazzi, A., Ballabio, C., Panagos, P., Jones, A., Tedersoo, L., Bahram, M., Guerra, C.A., Eisenhauer, N., Tao, D., Delgado-Baquerizo, M., García-Palacios, P. & van der Heijden, M.G.A. (2024). Soil health is associated with higher primary productivity across Europe. *Nature Ecology & Evolution*, 8 (10), 1847–1855. https://doi.org/10.1038/s41559-024-02511-8

Roudier, P. (2011). *clhs: a R package for conditioned Latin hypercube sampling.*

SLU (2023). *Markinventeringen*. *SLU.SE*. https://www.slu.se/centrumbildningar-och-projekt/markinventeringen/ [2025-05-15]

SLU (2024a). *Fältinstruktion 2024. Riksinventeringen av skog*. SLU institutionen för skoglig resurshushållning och institutionen för mark och miljö. http://www.slu.se/riksskogstaxeringen

SLU (2024b). *Markinfo*. *SLU.SE*. https://www.slu.se/institutioner/mark-miljo/miljoanalys/markinfo/ [2025-05-19]

Statistics Sweden (2023). Land Use in Sweden 2020.

Swedish Environmental Protection Agency (2024). National Land Cover Database (NMD). Data set. https://www.naturvardsverket.se/en/services-and-permits/maps-and-map-services/national-land-cover-database/ [2025-03-11]

Swedish Environmental Protection Agency (2025). *National Inventory Document Sweden 2025*. Naturvårdsverket.

Tennekes, M. (2018). tmap: Thematic maps in R. *Journal of Statistical Software*, 84 (6), 1–39. https://doi.org/10.18637/jss.v084.i06

Tóth, G., Jones, A. & Montanarella, L. (2013). *LUCAS topsoil survey: methodology, data and results*. Publications Office of the European Union. https://data.europa.eu/doi/10.2788/97922 [2025-01-20]

Ťupek, B., Ortiz, C.A., Hashimoto, S., Stendahl, J., Dahlgren, J., Karltun, E. & Lehtonen, A. (2016). Underestimation of boreal soil carbon stocks by mathematical soil carbon models linked to soil nutrient status. *Biogeosciences*, 13 (15), 4439–4459. https://doi.org/10.5194/bg-13-4439-2016

Van Liedekerke, M., Jones, A. & Panagos, P. (2006). ESDBv2 Raster Library - a set of rasters derived from the European Soil Database distribution v2.0 (published by the European Commission and the European Soil Bureau Network, EUR 19945 EN). https://esdac.jrc.ec.europa.eu/content/european-soil-database-v2-raster-library-1kmx1km [2025-03-27]

Veerman, C., Pinto Correia, T., Bastioli, C., Biro, B., Bouma, J., Cienciala, E., Emmett, B., Frison, E.A., Grand, A., Hristov, L., Kriaučiūnienė, Z., Pogrzeba, M., Soussana, J.-F., Vela, C. & Wittkowski, R. (2020). *Caring for soil is caring for life: ensure 75% of soils are healthy by 2030 for healthy food, people, nature and climate : interim report of the mission board for soil health and food*. Publications Office of the European Union. https://data.europa.eu/doi/10.2777/918775 [2025-05-12]

Wellbrock, N., Cools, N., de Vos, B., Jandl, R., Lehtonen, A., Leitgeb, E., Mäkipää, R., Pavlenda, P., Schwärtzel, K. & Šrámek, V. (2024). There is a need to better take into account forest soils in the planned soil monitoring law of the European Union. *Annals of Forest Science*, 81 (1), 22. https://doi.org/10.1186/s13595-024-01238-7

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T.L., Miller, E., Bache, S.M., Müller, K., Ooms, J., Robinson, D., Seidel, D.P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K. & Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4 (43), 1686. https://doi.org/10.21105/joss.01686

Yu, Z., Loisel, J., Brosseau, D.P., Beilman, D.W. & Hunt, S.J. (2010). Global peatland dynamics since the Last Glacial Maximum. *Geophysical Research Letters*, 37 (13). https://doi.org/10.1029/2010GL043584

Popular science summary

With the introduction of the Soil Monitoring Law the European Union aims to tackle the worsening condition of soils across the continent. This includes the establishment of a soil monitoring system and reporting requirements for the EU's Member States. This means, countries will need to regularly measure certain properties in their soils, for example how much organic carbon they contain or how acidic they are, and report that information to the European Commission.

Many countries already have a national soil monitoring system. In Sweden, forest soils are monitored by the Swedish Forest Soil Inventory, or SFSI. In addition, the EU-wide LUCAS Soil Survey, which is carried out by EU institutions, also repeatedly samples the soil in Swedish forests. In this project, I compared the methods that are used by the SFSI and the LUCAS Soil Survey and analysed how and why the measurements are different. In a second step, I explored the required sampling density, meaning at how many sites in the country soil samples will need to be collected.

There are very big differences in how soil samples are taken between the two soil monitoring systems. The SFSI samples soils by digging up to 1 m deep pits and takes samples from several layers, starting from the humus at the very top and going all the way down to the deep mineral subsoil. In comparison, during field work for LUCAS only 20 cm deep slices of the soil are taken for lab analysis. Crucially, humus – the decomposed plant residues that pile up on the forest floor – is removed from the samples in most cases. Because humus has very different chemical properties from regular mineral soil, whether it is included in the sampling or not makes a big difference on the results of the lab analysis. This is an important reason for why the results of the LUCAS Soil Survey show lower concentrations of organic carbon and nitrogen in forest soils than the SFSI.

One important factor when determining the required sample size was the definition of so-called "soil units". These are areas that have similar soil conditions and represent the domains of interest, which means that we want to know the soil properties in these regions with some degree of certainty. For example, we want to be sure that the mean carbon concentration of all our sample sites in a soil unit – let's say all forest soils in the county of Dalarna with the soil type Podzols – differs less than 5 % from the value that we would get when measuring *all* of the soil in that soil unit. Ideally, we would sample at as many sites as possible but because that is very expensive and time consuming, we apply statistical methods to determine how many samples we need to get *close enough* to the true value.

This is quite a challenging task, partly because of course no one has ever measured soil properties in *all* of Sweden, therefore we do not know the true value. In this project, a map of soil properties was used, which was made using measurements from the LUCAS Soil Survey and models that try and "guess" the soil properties in between the sampling sites, using other known variables such as the topography. Based on this map the required sample size to measure soil properties with the necessary accuracy in Swedish forests is 1390. The sample sizes of both, the SFSI (8743) and LUCAS (1550) are higher than that. However, it is not as simple as saying a higher sample size means that the requirements are fulfilled. The location of the samples is equally important because the more variable a region is, the more samples we need to accurately capture the soil conditions there. Furthermore, the responsible authorities in Sweden still have to decide on how to define their soil units. This will also influence the required sample size.

Further analyses will be needed once the Soil Monitoring Law is finally agreed upon and enacted. It certainly will be a challenge for many countries to assess how they need to sample their soils to be able to measure soil properties with the required accuracy.

Appendix 1

Organic Carbon

Table 8. Descriptive statistics of organic carbon in top 20 cm of soil per NUTS 2 region. Values are given in g kg^{-1}.

Region	Survey	Mean	SD	CV	Median	Q1	Q3	Min	Max	n
Eastern Middle Sweden	sfsi	147.60	144.42	0.98	65.50	33.03	235.70	2.00	502.00	354
	lucas	134.43	131.62	0.98	79.85	47.23	164.55	15.10	539.90	188
Middle Norrland	sfsi	154.14	152.45	0.99	108.38	18.79	252.79	3.10	498.00	420
	lucas	130.02	152.34	1.17	49.90	28.10	163.25	8.50	520.40	230
Northern Middle Sweden	sfsi	168.51	148.66	0.88	158.10	24.50	266.42	2.60	520.00	539
	lucas	150.77	155.87	1.03	74.35	42.05	223.90	6.80	582.90	324
Småland and the islands	sfsi	167.21	155.73	0.93	87.25	35.79	281.08	1.70	555.00	322
	lucas	166.70	158.19	0.95	87.90	53.00	267.90	4.70	524.50	169
South Sweden	sfsi	160.42	161.63	1.01	68.00	33.85	265.35	5.80	524.00	167
	lucas	152.61	147.38	0.97	85.10	53.55	184.70	32.50	519.70	47
Stockholm	sfsi	125.07	126.94	1.01	53.90	33.38	190.25	7.00	470.00	52
	lucas	78.97	44.06	0.56	74.50	44.75	95.80	27.30	179.40	23
Upper Norrland	sfsi	148.03	154.60	1.04	137.75	15.55	255.25	3.20	522.00	611
	lucas	88.62	131.88	1.49	30.40	17.75	75.00	3.30	539.50	443
West Sweden	sfsi	196.14	157.20	0.80	153.03	45.44	311.68	6.60	498.00	300
	lucas	197.48	155.86	0.79	153.20	62.45	319.93	19.00	523.70	126
All	sfsi	157.96	153.04	0.97	112.40	25.30	261.80	1.70	555.00	2765
	lucas	132.47	148.48	1.12	63.00	31.70	167.10	3.30	582.90	1550

Nitrogen

Table 9. Descriptive statistics of nitrogen in top 20 cm of soil per NUTS 2 region. Values are given in g kg^{-1}.

Region	Survey	Mean	SD	CV	Median	Q1	Q3	Min	Max	n
Eastern Middle Sweden	sfsi	5.83	5.53	0.95	3.40	1.60	8.63	0.10	28.20	354
	lucas	6.09	5.31	0.87	3.95	2.68	7.33	0.70	26.70	188
Middle Norrland	sfsi	5.65	5.75	1.02	3.89	0.79	9.53	0.10	26.60	420
	lucas	5.40	6.31	1.17	2.30	1.30	6.78	0.50	27.00	230
Northern Middle Sweden	sfsi	5.41	4.76	0.88	4.63	0.85	8.69	0.10	23.80	539
	lucas	5.84	5.90	1.01	3.15	1.90	7.98	0.40	27.50	324
Småland and the islands	sfsi	6.36	5.61	0.88	4.40	1.65	10.06	0.00	25.90	322
	lucas	6.92	5.82	0.84	4.30	2.60	10.80	0.30	25.60	169
South Sweden	sfsi	7.02	6.54	0.93	3.40	1.55	11.25	0.40	23.00	167
	lucas	6.85	5.42	0.79	4.50	3.20	9.10	1.80	20.40	47
Stockholm	sfsi	5.58	5.13	0.92	3.25	1.44	9.04	0.30	21.20	52
	lucas	4.32	2.45	0.57	3.20	2.35	5.60	1.90	10.30	23
Upper Norrland	sfsi	5.26	5.79	1.10	3.95	0.60	9.04	0.10	28.90	611
	lucas	3.63	5.48	1.51	1.40	0.90	3.10	0.20	32.30	443
West Sweden	sfsi	7.33	5.64	0.77	5.88	2.20	11.31	0.10	28.90	300
	lucas	8.63	5.96	0.69	7.10	3.50	12.88	1.00	22.80	126
All	sfsi	5.72	5.62	0.98	4.30	1.05	9.50	0.00	28.90	2765
	lucas	5.53	5.90	1.07	3.00	1.50	7.30	0.20	32.30	1550

pH

Table 10. Descriptive statistics of pH in top 20 cm of soil per NUTS 2 region.

Region	Survey	Mean	SD	CV	Median	Q1	Q3	Min	Max	n
Eastern Middle Sweden	sfsi	4.52	0.62	0.14	4.48	4.17	4.82	3.21	7.12	354
	lucas	4.69	0.65	0.14	4.55	4.19	5.04	3.72	7.10	188
Middle Norrland	sfsi	4.50	0.47	0.10	4.47	4.20	4.76	3.34	6.98	420
	lucas	4.67	0.54	0.12	4.59	4.30	4.97	3.59	6.81	230
Northern Middle Sweden	sfsi	4.25	0.42	0.10	4.27	3.98	4.50	3.28	6.67	539
	lucas	4.41	0.40	0.09	4.36	4.13	4.64	3.60	6.33	324
Småland and the islands	sfsi	4.37	0.78	0.18	4.24	3.90	4.57	3.11	7.46	322
	lucas	4.48	0.67	0.15	4.34	4.02	4.71	3.38	7.32	169
South Sweden	sfsi	4.32	0.58	0.13	4.25	3.95	4.51	3.29	7.59	167
	lucas	4.31	0.50	0.12	4.16	3.94	4.47	3.57	5.63	47
Stockholm	sfsi	4.78	0.68	0.14	4.72	4.44	5.13	3.36	6.78	52
	lucas	4.81	0.55	0.11	4.87	4.36	5.13	3.74	5.93	23
Upper Norrland	sfsi	4.36	0.36	0.08	4.37	4.12	4.60	3.25	5.53	611
	lucas	4.54	0.40	0.09	4.51	4.25	4.84	3.49	5.93	443
West Sweden	sfsi	4.14	0.43	0.10	4.12	3.85	4.42	3.29	6.18	300
	lucas	4.36	0.39	0.09	4.29	4.10	4.54	3.62	5.61	126
All	sfsi	4.37	0.49	0.11	4.34	4.03	4.61	3.11	7.59	2765
	lucas	4.53	0.51	0.11	4.45	4.16	4.81	3.38	7.32	1550

Sand

Table 11. Descriptive statistics of sand in top 20 cm of soil per NUTS 2 region. Values are given in %.

Region	Survey	Mean	SD	CV	Median	Q1	Q3	Min	Max	n
Eastern Middle Sweden	sfsi	57.90	29.41	0.51	72	40	81	0	100	354
	lucas	47.16	20.64	0.44	51	34	62	2	94	188
Middle Norrland	sfsi	72.52	15.88	0.22	79	72	82	0	100	420
	lucas	50.04	20.52	0.41	51	39	63	1	94	230
Northern Middle Sweden	sfsi	73.28	18.01	0.25	79	72	82	0	100	539
	lucas	59.66	18.91	0.32	63	51	71	4	94	324
Småland and the islands	sfsi	71.74	16.07	0.22	77	70	80	0	100	322
	lucas	60.55	15.01	0.25	59	51	71	7	91	169
South Sweden	sfsi	73.31	14.36	0.20	79	70	82	0	89	167
	lucas	62.69	10.28	0.16	63	58	70	35	79	47
Stockholm	sfsi	56.01	30.82	0.55	72	25	79	0	89	52
	lucas	55.37	18.89	0.34	54	44	67	23	89	23
Upper Norrland	sfsi	76.40	12.99	0.17	79	75	82	0	100	611
	lucas	57.67	18.43	0.32	58	48	70	2	98	443
West Sweden	sfsi	65.66	21.63	0.33	75	65	79	0	100	300
	lucas	59.11	15.74	0.27	61	53	68	4	91	126
All	sfsi	71.98	18.57	0.26	79	70	82	0	100	2765
	lucas	56.45	18.86	0.33	58	47	69	1	98	1550

Silt

Table 12. Descriptive statistics of silt in top 20 cm of soil per NUTS 2 region. Values are given in %.

Region	Survey	Mean	SD	CV	Median	Q1	Q3	Min	Max	n
Eastern Middle Sweden	sfsi	33.33	19.91	0.60	22	18	43.5	0	85	354
	lucas	36.97	13.21	0.36	37	28.5	46	5	70	188
Middle Norrland	sfsi	24.03	11.69	0.49	21	18	23	0	85	420
	lucas	39.26	15.93	0.41	39	28	47	3	87	230
Northern Middle Sweden	sfsi	23.49	12.40	0.53	21	18	22	0	85	539
	lucas	33.63	14.49	0.43	32	25	41	4	76	324
Småland and the islands	sfsi	24.51	11.74	0.48	21	19.5	25	0	85	322
	lucas	30.81	11.65	0.38	33	22	40	6	67	169
South Sweden	sfsi	23.53	10.02	0.43	21	18	25	11	85	167
	lucas	29.06	7.64	0.26	30	23.5	34	15	45	47
Stockholm	sfsi	34.94	21.90	0.63	22	21	52.5	11	85	52
	lucas	26.32	11.80	0.45	28	18	32.5	7	54	23
Upper Norrland	sfsi	21.41	9.66	0.45	21	18	22	0	85	611
	lucas	36.47	16.05	0.44	37	26	45	1	92	443
West Sweden	sfsi	28.48	14.57	0.51	22	21	30.5	0	85	300
	lucas	32.16	12.46	0.39	31	25	37.5	7	76	126
All	sfsi	24.32	13.02	0.54	21	18	25	0	85	2765
	lucas	34.83	14.61	0.42	34	26	43	1	92	1550

Clay

Table 13. Descriptive statistics of silt in top 20 cm of soil per NUTS 2 region. Values are given in %.

Region	Survey	Mean	SD	CV	Median	Q1	Q3	Min	Max	n
Eastern Middle Sweden	sfsi	8.77	10.62	1.21	5	0	15	0	35	354
	lucas	15.86	13.61	0.86	10	7	20.5	1	54	188
Middle Norrland	sfsi	3.45	4.85	1.41	0	0	6	0	35	420
	lucas	10.70	9.38	0.88	7	5	12	2	56	230
Northern Middle Sweden	sfsi	3.23	6.24	1.93	0	0	5	0	35	539
	lucas	6.71	6.64	0.99	4	3	7	1	43	324
Småland and the islands	sfsi	3.75	5.02	1.34	2.5	0	6	0	35	322
	lucas	8.59	6.15	0.72	7	5	9	2	43	169
South Sweden	sfsi	3.16	4.90	1.55	0	0	5	0	35	167
	lucas	8.23	4.61	0.56	7	6	9	3	26	47
Stockholm	sfsi	9.05	10.49	1.16	6	0	15	0	35	52
	lucas	18.05	12.16	0.67	13	9.5	21.5	4	44	23
Upper Norrland	sfsi	2.19	3.82	1.75	0	0	3	0	35	611
	lucas	5.87	4.45	0.76	5	4	7	1	40	443
West Sweden	sfsi	5.86	7.82	1.33	3	0	6	0	35	300
	lucas	8.66	6.60	0.76	7	5	10	2	52	126
All	sfsi	3.70	6.24	1.69	0	0	6	0	35	2765
	lucas	8.70	8.40	0.97	6	4	10	1	56	1550

Comparison of optimized sample tests

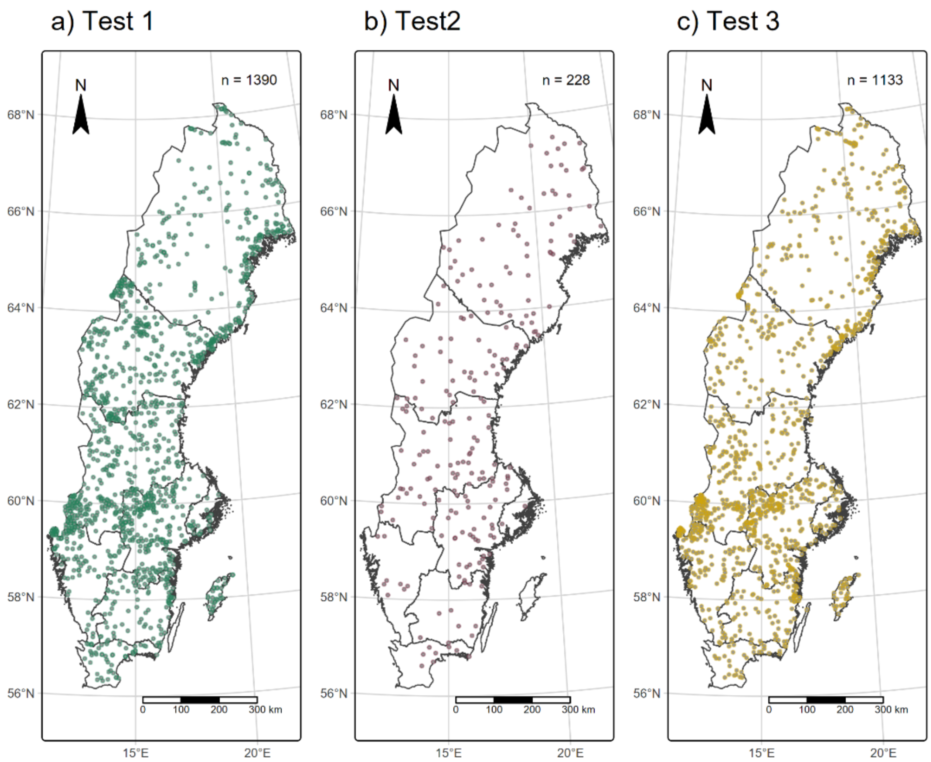


Figure 14. Mapped sample points in forest land from the three tests carried out with the JRC script to determine an optimal sample. The difference between the three tests is in the domain definition (see Table 6). The borders of Sweden's NUTS 2 regions are shown as well.

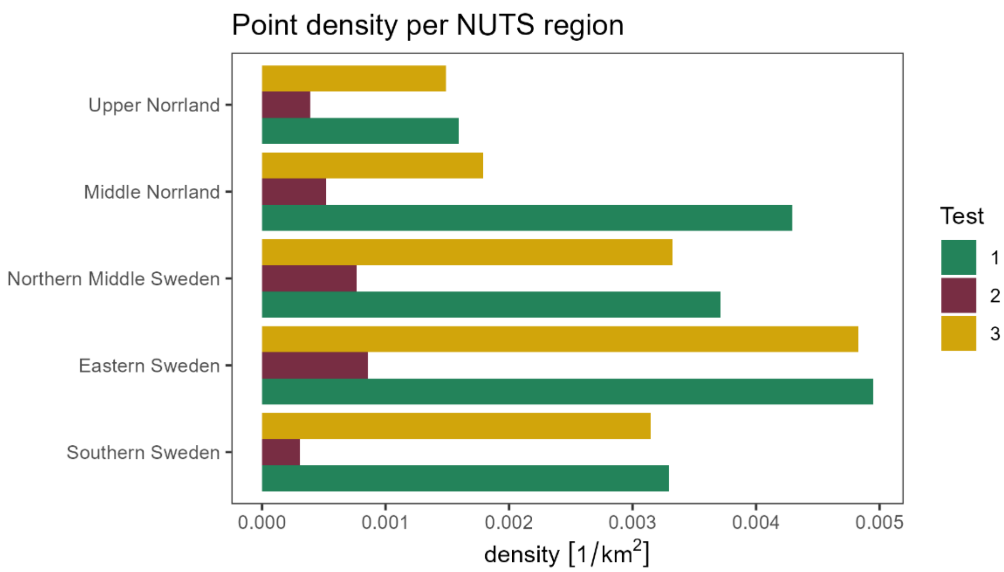


Figure 15. Compared point density in forest land per NUTS region of the three tests carried out with the JRC script to determine an optimal sample.

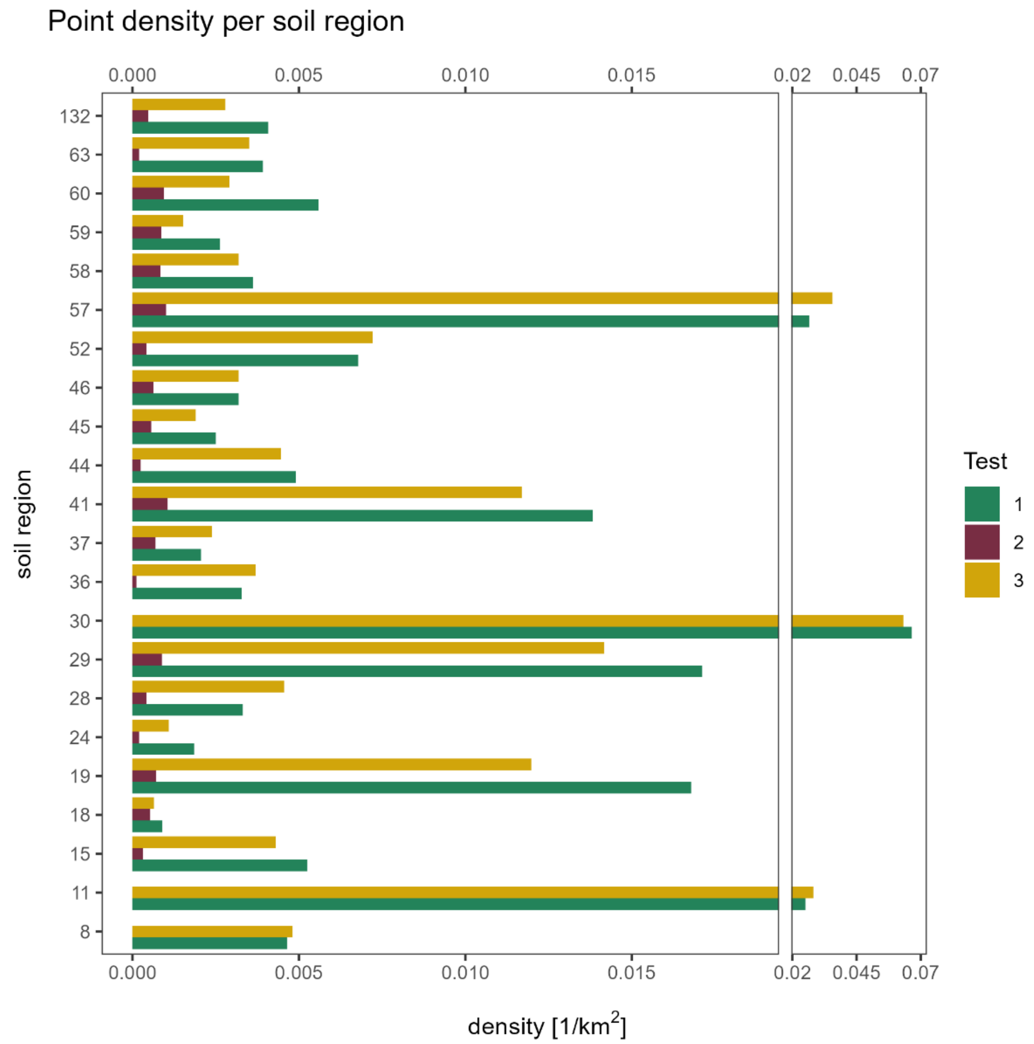


Figure 16. Compared point density in forest land per soil region of the three tests carried out with the JRC script to determine an optimal sample.

Appendix 2

Maps of NUTS 2 regions and NMD land cover

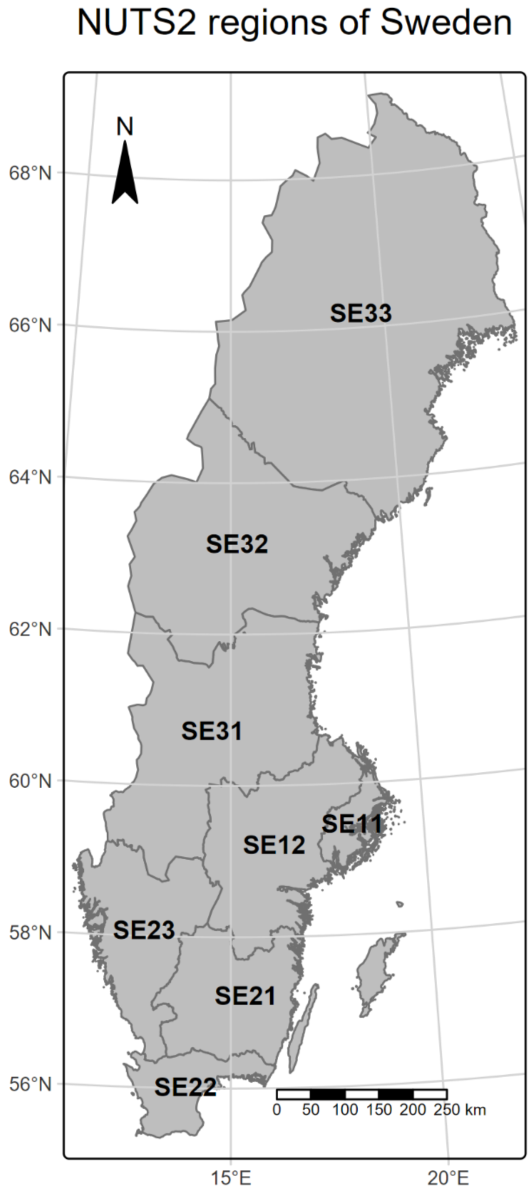


Figure 17. The map shows the NUTS 2 regions of Sweden. The first digit denotes the corresponding NUTS 1 region.

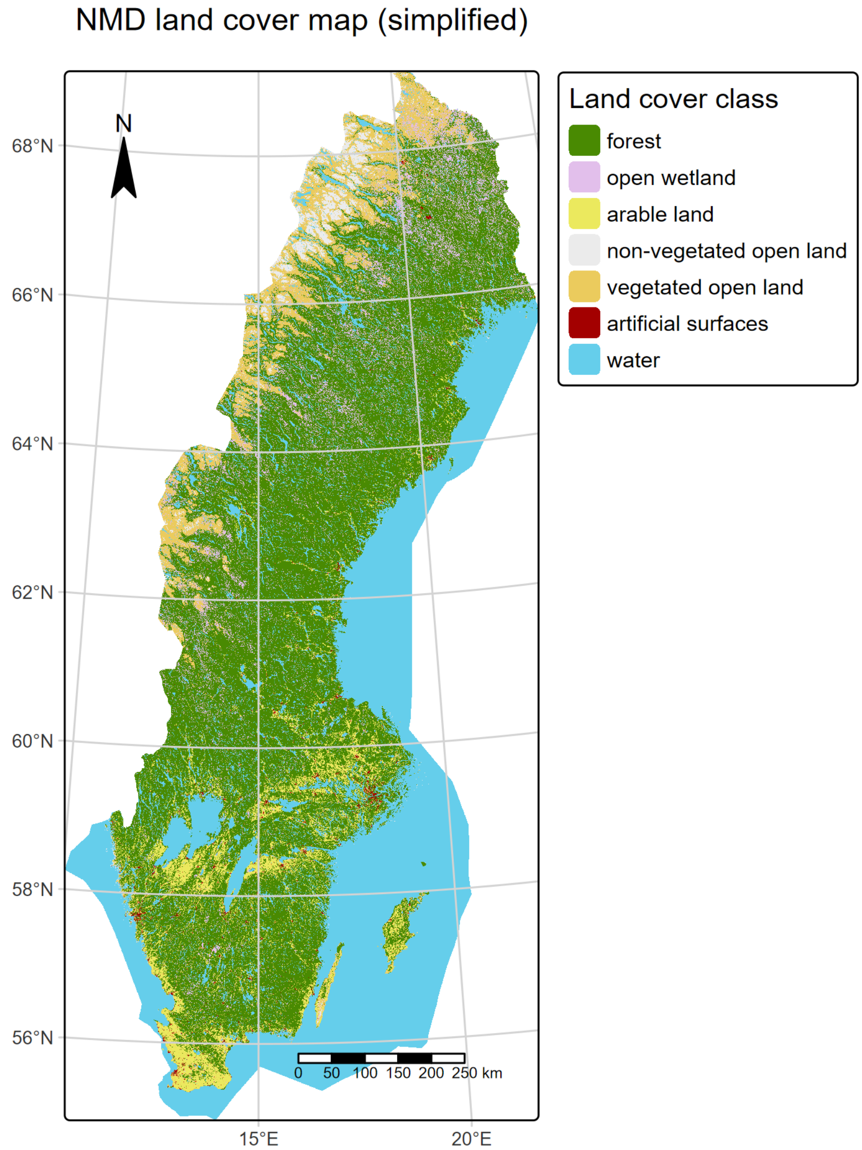


Figure 18. NMD land cover map 2018 aggregated to categories comparable to IPCC land use categories at 10 m x 10 m resolution.

Soil regions of Sweden

Table 14. Legend of all relevant soil regions in Sweden in the map of soil regions of the European Uniton and adjacent countries, available online at: *https://data.europa.eu/data/datasets/ae71ffee-1ae9-4624-ae3f-f49513fe9dcb?locale=en*

Number	Climate	Geographical position	Dominant soils	WRB-code	Associated soils	WRB-code	Dominant parent material	Dominant land use	Elevation class	Slope class
8	Subpolar oceanic to boreal oceanic	North Norway: *Finnmark and Troms*	Lithic and Umbric Leptosols Dystric Regosols	LPli, LPum RGdy	Haplic Podzols Dystric Histosols	PZha HSdy	Igneous and metamorphic rocks, partly covered by thin morainic deposits, peat	Semi-natural mixed pattern	Hill to low mountain	Level to sloping
11	Boreal oceanic	Norway: *Nord- and Soer-Troendelag*	Haplic Podzols Dystric Regosols	PZha RGdy	Lithic and Umbric Leptosols Dystric Histosols	LPli, LPum HSdy	Igneous and metamorphic rocks, patchy morainic deposits, peat	grassland	hill	Level to sloping
15	Boreal continental	Finland, Russia: *Kola, Lappland, Karelia*	Eutric and Dystric Histosols	HSeu, HSdy	Haplic Podzols	PZha	Peat, loamy morainic deposits	Semi-natural mixed pattern	plain	flat
18	Boreal continental	Sweden: *Lappland, Norrbotten and Västerbotten*	Haplic, Gleyic and Carbic Podzols	PZha, PZgl, PZcb	Eutric and Dystric Histosols Dystric Cambisols	HSeu, HSdy CMdy	Loamy to sandy morainic deposits, peat, igneous and metamorphic rocks	Coniferous forest	hill	level
19	Boreal continental	Sweden: *coastal area of Norrbotten and Västerbotten*	Haplic Podzols Dystric Cambisols	PZha CMdy	Gleyic and Carbic Podzols Haplic Arenosols	PZgl, PZcb ARha	Morainic and glacifluvial to fluvial deposits, partly igneous and metamorphic rocks	Coniferous forest	plain	flat
20	Boreal continental	Central Finnland: *Oulu district*	Haplic Podzols Dystric Histosols	PZha HSdy	Dystric Cambisols	CMdy	Loamy morainic deposits, peat, igneous and metamorphic rocks	Semi-natural mixed pattern	plain	flat
24	Boreal mountainous	Norway: *Nordland* Sweden: *Norrbotten*	Lithic Leptosols	LPli	Haplic Podzols	PZha	Igneous and metamorphic rocks, morainic deposits	Semi-natural mixed pattern	Low mountain to mountain	Undulating level to sloping
28	Boreal sub-oceanic to temperate sub-oceanic	Southwest Sweden: *Bohus and Halland*	Lithic Leptosols Eutric and Dystric Cambisols	LPli CMeu, CMdy	Haplic Luvisols	LVha	Igneous and metamorphic rocks, marine deposits	Coniferous forest	plain	flat

29	Boreal sub-oceanic to temperate sub-oceanic	Sweden: *Dalsland*	Haplic, Gleyic and Carbic Podzols Lithic and Umbric Leptosols	PZha, PZgl, PZcb LPli, LPum	Dystric Cambisols	CMdy	Igneous and metamorphic rocks, morainic deposits	Coniferous forests	plain	flat
30	Boreal sub-oceanic to temperate sub-oceanic	South Norway: *Oslofjord area*	Haplic and Rustic Podzols Gleyic Albeluvisols	PZha, PZrs ABgl	Gleyic Cambisols Lithic and Umbric Leptosols	CMgl LPli, LPum	Marine deposits, igneous and metamorphic rocks, patchy morainic deposits	Mixed cultivation pattern	plain	Flat to level
36	Boreal sub-continental to temperate sub-continental	Sweden: *Vänern lake area*	Dystric and Vertic Cambisols	CMdy, CMvr	Eutric Cambisols Haplic Podzols	CMeu PZha	Loamy to clayey glaciolacustrine and morainic deposits	Mixed cultivation pattern	plain	Flat to level
37	Boreal sub-continental to temperate sub-continental	Sweden: *Uppsala and Södermanland district*	Dystric, Eutric and Vertic Cambisols	CMdy, CMeu, CMvr	Haplic Luvisols Lithic Leptosols	LVha LPli	Loamy to sandy morainic deposits and clayey to silty glaciolacustrine sediments, outcrops of igneous and metamorphic rocks	Mixed cultivation pattern	plain	Flat to level
41	Boreal sub-continental to temperate sub-continental	Southeast Sweden, South-Finland: *Baltic Sea Foreland*	Lithic Leptosols	LPli	Dystric Cambisols Haplic Podzols Dystric Gleysols	CMdy PZha GLdy	Igneous and metamorphic rocks, patchy morainic deposits	Mixed cultivation pattern	plain	flat
44	Boreal sub-continental to temperate sub-continental	South Sweden	Haplic and Gleyic Podzols	PZha, PZgl	Dystric Gleysols Dystric and Eutric Histosols	GLdy HSdy, HSeu	Sandy to loamy morainic and glaciolacustrine deposits, peat	Coniferous forest	plain	Flat to level
45	Boreal sub-continental to temperate sub-continental	Sweden: *Östergotland, Småland*	Haplic Podzols	PZgl, PZha	Dystric Cambisols Gleyic and Carbic Podzols	CMdy PZcb	Loamy morainic deposits	Coniferous forest	plain	Flat to level
46	Boreal sub-continental to temperate sub-continental	Sweden: *Central Skåne*	Dystric and Eutric Regosols Dystric Cambisols	RGdy, RGeu CMdy	Eutric Cambisols Haplic Podzols Lithic Leptosols	CMeu PZha LPli	Morainic deposits, outcrops of igneous, metamorphic and palaeozoic sedimentary rocks	Coniferous forest	plain	Flat to level
52	Boreal continental to temperate continental	Middle Sweden: *Kopparberg district*	Eutric and Dystric Histosols	HSeu, HSdy	Haplic Podzols	PZha	Peat, loamy morainic deposits	Coniferous forests	Low mountain	Moderate level

57	Boreal continental to temperate continental	South Norway: *Akershus district*	Rustic Podzols Vertic and Gleyic Cambisols	PZrs CMvt, CMgl	Lithic Leptosols Dystric and Eutric Histosols	LPli HSdy, HSeu	Igneous and metamorphic rocks, palaeozoic sedimentary rocks, glacial deposits, peat	Coniferous forest	Plain to hill	Level to sloping
58	Boreal continental to temperate continental	Sweden: *Kopparberg distr.* South Norway: *Hedmark*	Haplic Podzols Eutric Histosols	PZha HSeu	Gleyic and Carbic Podzols Gleysols	PZgl, PZcb GL	Loamy morainic deposits, peat, outcrops of igneous and metamorphic rocks	Semi-natural mixed pattern	Low mountain	Moderate level
59	Boreal continental to temperate continental	Middle Sweden: *Västernorrland and Jämtland*	Haplic Podzols Dystric Cambisols	PZha CMdy	Gleyic and Carbic Podzols Eutric Cambisols	PZgl, PZcb CMeu	Loamy morainic deposits, igneous and metamorphic rocks	Coniferous forests	hill	level
60	Boreal continental to temperate continental	Middle Sweden: *Jämtland and Västernorrland*	Haplic Podzols	PZha	Gleyic and Carbic Podzols	PZgl, PZcb	Loamy morainic deposits, igneous metamorphic rocks	Coniferous forest	hill	level
63	Boreal mountainous to temperate mountainous climate	Norway: *Oppland* Sweden: *Jämtland*	Lithic and Umbric Leptosols Dystric Regosols (partly permanent snow cover)	LPli, LPum RGdy	Haplic Podzols Dystric and Eutric Histosols	PZha HSdy, HSeu	Igneous, metamorphic and palaeozoic sedimentary rocks, morainic deposits, peat	Semi-natural mixed pattern	mountain	sloping
132	Temperate sub-oceanic to temperate sub-continental	Sweden: *Skåne, Oeland, Gotland* Denmark: *Sjaelland, Lolland, Bornholm*	Eutric, Dystric and Calcaric Cambisols	CMeu, CMdy, CMca	Haplic Luvisols Dystric and Calcaric Regosols	LVha RGdy, RGca	Loamy morainic deposits, sandy partly clayey glaciolacustrine deposits	Rainfed arable land	plain	Flat to level
1000	No regard to climate	All over Europe	Fluvisols, undifferentiated	FL	Gleysols, undifferentiated	GL	Sandy to loamy fluvial deposits		No specification	level
2000	No regard to climate	All over Europe	Anthrosols and Urban Areas	AT					Plain to hill	Level to sloping

Publishing and archiving

Approved students' theses at SLU can be published online. As a student you own the copyright to your work and in such cases, you need to approve the publication. In connection with your approval of publication, SLU will process your personal data (name) to make the work searchable on the internet. You can revoke your consent at any time by contacting the library.

Even if you choose not to publish the work or if you revoke your approval, the thesis will be archived digitally according to archive legislation.

You will find links to SLU's publication agreement and SLU's processing of personal data and your rights on this page:

- https://libanswers.slu.se/en/faq/228318

☒ YES, I, Merlin Mayer, have read and agree to the agreement for publication and the personal data processing that takes place in connection with this

☐ NO, I/we do not give my/our permission to publish the full text of this work. However, the work will be uploaded for archiving and the metadata and summary will be visible and searchable.