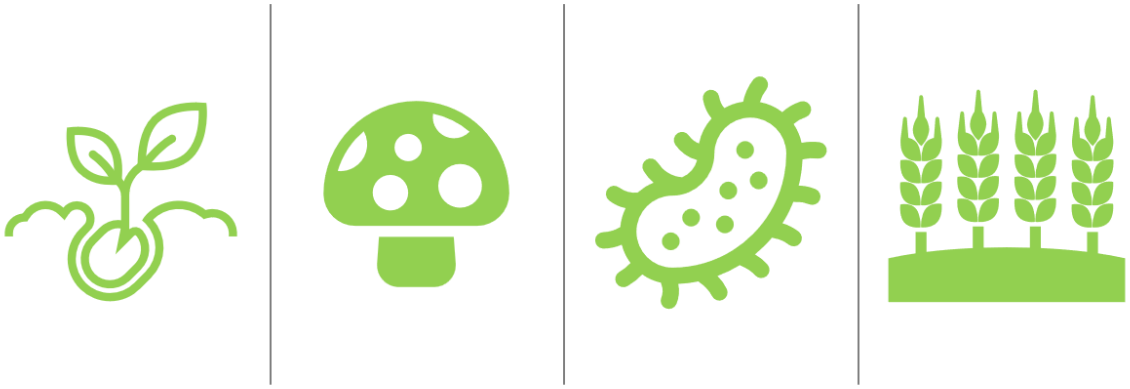

Investigation of linear correlations between soil indicators

Master's thesis in Industrial ecology

SOFIA WESTERGREN

DEPARTMENT OF TECHNOLOGY MANAGEMENT AND ECONOMICS
DIVISION OF ENVIRONMENTAL SYSTEMS ANALYSIS

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2025
www.chalmers.se

Investigation of linear correlations between soil indicators

SOFIA WESTERGREN

Department of Technology Management and Economics
Division of Environmental Systems Analysis
Chalmers University of Technology
Gothenburg, Sweden 2025

Investigation of linear correlations between soil indicators
SOFIA WESTERGREN

Supervisor: Mats Töpel, Swedish Environmental Research Institute
Examiner: Sverker Molander, Chalmers University of Technology

Department of Technology Management and Economics
Division of Environmental Systems Analysis
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: Created in Powerpoint by Sofia Westergren.

Gothenburg, Sweden 2025

Investigation of linear correlations between soil indicators
SOFIA WESTERGREN
Department of Technology Management and Economics
Chalmers University of Technology

Abstract

The soil is an important part of the world that contributes to ecosystem services such as food production, water purification, and protection from climate changes. Despite that, knowledge and understanding of the structure and functions of the soil system is deficient. The world faces challenges related to crop production and food consumption. These challenges are predicted to become even greater in the future. An increased understanding, will contribute to the work to assess the status of soils and preserve or regain the functions from soil. The aim of this study is to increase the understanding of how soil indicators can be used to describe soil and specifically to investigate the existence of linear correlations between soil indicators. The investigation includes calculation of biological soil indicators based on environmental DNA from *The Land Use/Cover Area frame statistical Survey* (LUCAS). The calculated biological soil indicators are the ratio of gram-positive and gram-negative bacteria, diversity index and presence of actinomycetes bacteria, mycorrhiza forming fungi, and saprotrophic fungi. The data analysis includes regression analysis and principal component analysis between the biological soil indicators and characteristic soil indicators (see Table 4.5). The results indicate that there may be no linear correlations between the chosen biological soil indicators and the characteristic soil indicators.

Keywords: Environmental DNA, Soil indicators, Soil health, LUCAS, Principal Component Analysis

Acknowledgements

I am grateful for the opportunity to carry out this project and the knowledge that I have gained from this master thesis. I would like to express my thanks for the support that I have received during the work. First, I want to express my gratitude to my supervisor Mats Töpel at the Swedish Environmental Research Institute for valuable support, input, and knowledge. In the same way, I want to thank the Swedish Environmental Research Institute for the opportunity to write the master thesis at the company. I also want to express my thankfulness to Jessica Johansson at Svensk Kolinlagring for additional guidance and feedback during the project. Furthermore, I want to express my gratitude to my examiner Sverker Molander at Chalmers University of Technology for feedback on the project. Finally, I would like to express my appreciation to my family and fellow students for valuable discussions and support during the thesis work.

Sofia Westergren, Gothenburg, December 2025

List of Acronyms

ASVs	Amplicon Sequence Variants
C	Carbon
CO_3^{2-}	Carbonate
DNA	Deoxyribonucleic acid
eDNA	Environmental DNA
EU	European Union
EUROSTAT	European Statistical office
GC-Ms	Gas Chromatography Mass Spectrometry
GDPR	General Data Protection Regulation
gram+ bacteria	gram-positive bacteria
gram- bacteria	gram-negative bacteria
JCR	Joint Research Centre
K	Potassium
LUCAS	Land Use/Cover Area frame statistical Survey
LPS	Lipopolysaccharid
N	Nitrogen
N_2	Nitrogen gas
NO_4^+	Ammonium
NO_2	Nitrite
NO_3^-	Nitrate
OTUs	Operational Taxonomic Units
P	Phosphorus
PCR	Polymerase Chain Reaction
PLFA	Phospholipid Fatty Acids
SDG	Sustainable Development Goals
°C	Degree Celsius

Nomenclature

Description of the indicators and variables that have been used in the report.

Soil health indicators

gram+ / gram-	Ratio of gram-positive and gram-negative bacteria [-]
actinomycetes	The proportion of ASVs of actinomycetes bacteria of all ASVs of bacteria in the data set [%]
saprotrophs	The proportion of ASVs of saprotrophic fungi of all ASVs of fungi in the data set [%]
mycorrhiza	The proportion of ASVs of mycorrhiza forming fungi of all ASVs of fungi in the data set [%]
H	Diversity index based on the Shannon-Wiener index [-]

Variables

$\bar{X}_i$	Calculated mean value of data set i
μ_i	Real mean value of data set i
s_i	Calculated standard deviation of data set i
s	Pooled variance
n_i	Number of data points in data set i
x	Variable describing the distribution function for a t-distribution with r degrees of freedom.
R^2	Coefficient of determination.
y_i	y value according to the data set.
$\hat{y}$	y value according to the function expression.
$\bar{y}$	Mean value of the y values in the data set.
λ	Eigenvalue of a matrix
$\mathbf{v}$	Eigenvector of a matrix
σ_i	Standard deviation along principal component i in PCA.

Contents

List of Figures

List of Tables

1

Introduction

The soil is a mainstay for a lot of functions at Earth [34]. For example, it plays a role in the food production due to its contribution of nutrients, water and important structure. Without well functioning soil, the challenge to secure food for the population will be even tougher than today. In addition to that, the forestry is dependent on the soil to make it possible to produce the amount of wood that is in demand. The soil is also a part of the water cycle. Depending on, for example, soil structure and the presence of organic matter, the capacity of water infiltration and storage changes. Another important aspect is that the work for well functioning soils is a part of the mitigation of the climate changes due to that soil is working as a large reservoir for carbon storage. The atmospheric carbon reaches the soil by photosynthesis by plats and then through the roots or decomposition of plant litter. The last example is that the soil is home for a lot of plants and animals. The life in the soil contribute to many of the ecosystem services that a related to soil such as decomposition of organic material, plant growth, nutrient cycling and water purification.

The way the humanity uses the soil contributes to a reduction of the soil functions [34]. Examples of the how the humanity inhibits the soil are coverage with for example buildings or infrastructure, soil compaction as a consequence of heavy vehicles, soil contamination and reduction of organic material in the soil. These kind of usage result in a soil with lower fertility, reduced capacity to hold water and reduced capacity to be a storage of carbon.

Due to an increasing global population, the capacity of soil need to be used in an efficient way. Because of the important role of soil and the way its functions are reduced as a consequence of human use, it is important to increase the understanding and knowledge about soil and how soil can be described. An increased understanding, will contribute to the work to assess the status of soils and preserve or regain the functions from soil. Therefore, this report will investigate methods to describe the soil.

There are a lot of different ways to describe the soil. One way is to use soil indicators. Examples of soil indicators can be pH, coarse fragment or the amount of different substances in the soil. Since 2009, measurements of soil indicators have been done in the *European Union* as a result of a decision of the *European Parliament* [24] (see more in Section 3.1.1). The project is called the *Land Use/Cover Area frame statistical Survey (LUCAS)* and is a collaboration between the *European*

Statistical office (EUROSTAT), the *Directorate General responsible for Agriculture* and the *Joint Research Centre (JCR)*. The main indicators in the survey are coarse fragments, clay content, sand content silt content, pH, electrical conductivity, bulk density and content of organic C, CO_3^{2-}, P, N and K. In addition to these indicators even country and land use for each specific sample site are available. In this report these indicators are called *characteristic soil indicators*.

It is also possible to describe soil with the organisms that are represented in the soil, and there are different methods to do that. In this report the indicators representing the set of organisms are called *biological soil indicators*. One method that can be used to for instance measure the biomass of different organism groups is called Phospholipid fatty acids (PLFA) analysis which identify the organisms according to the measured PLFA in the cell membrane (see more in Section 2.3.2.1). The company *Eurofins* uses PLFA analysis to measure soil indicators including the set of organisms [22]. The biological soil indicators that are analysed in this report are based on the work Eurofins does. The project of LUCAS has used another method to describe which organisms that are represented in the soil, Environmental DNA (eDNA) [24] (see more in Section 2.3.2.2). This method identify organisms (in this case bacteria and fungi) by comparing identified DNA sequences with an existing database. The survey started to include these type of soil description in 2018. The biological soil indicators that have been analysed in this report are determined by a combination of the data from LUCAS and the work that Eurofins does. Some of the biological soil indicators analysed by Eurofins and measured with PLFA analysis are not possible to measure with eDNA (due to the difficulties to measure biomass with eDNA). The biological soil indicators that are analysed in this report are based on the Eurofins' indicators related to bacteria and fungi and not based on biomass. The reason for this is that only data related to bacteria and fungi is available from LUCAS. The five indicators used are the ratio of gram+ and gram− bacteria, presence of actinomycetes bacteria, presence of saprotrophic fungi, presence of mycorrhiza forming fungi, and diversity index (motivation and a detailed description of the these indicators are found in Section 3.2).

Table 1.1 shows an overview of the soil indicators that are analysed in this report. To improve the understanding of how soil can be described, investigations about possible correlations between these soil indicators will be done.

Table 1.1: Overview of the parameters used in this report (see more in Section 3).

Characteristic soil indicators	Biological soil indicators
Country	Ratio of gram+/gram− bacteria
Land use	Presence of actinomycetes bacteria
Coarse fragments	Presence of saprotrophic fungi
Clay content	Presence of mycorrhiza fungi
Sand content	Diversity index
Silt content	
pH	
Electrical conductivity	
Organic C	
CO_3^{2-} content	
P content	
N content	
K content	
Bulk density	

1.1 Aim and research questions

The aim of this study is to increase the understanding of how soil indicators can be used to describe soil and specifically to investigate the existence of linear correlations between characteristic soil indicators and biological soil indicators with data available in LUCAS.

The following research question will be investigated:

- Are there correlations between the chosen characteristic soil indicators and the biological soil indicators measured in LUCAS?
- Are there correlations between the chosen biological soil indicators?

1.2 Boundaries

There are many factors that influence the soil in different ways. This report includes analyses based on some factors. Other important organisms that are not included in the statistical analyses are, for example, protozoa which are predators that live in the water of the soil and eat bacteria, and nematodes which also are predators that eat bacteria, fungi, and protozoa [22]. Even the macrofauna including organisms larger than 2 mm, such as earthworms, play an important role in maintaining the structure of the soil. The soil changes porosity when earthworms dig passages, which also influences water infiltration [56][20]. Organisms such as microbivores and detritivores are responsible for decomposition and nutrient cycling. This means both that they transform substances such as carbon, nitrogen, and phosphorus into the soil and decompose pesticides in contaminated soils. Predators and hyperparasites are invertebrates that eat e.g. bacteria and fungi. Because of that, they regulate the population size of the microbes and remove pathogenic microbes from the soil.

Due to the complexity of the soil system, it is necessary to have boundaries. The report focuses on indicators available in LUCAS. The statistical analyses will be limited to biological soil indicators related to bacteria and fungi. Other factors, such as vegetation and other sets of organisms in the soil, and their impact on soil are only included in the theory. The available eDNA data from LUCAS include data for bacteria and fungi. Therefore, only these indicators are used in the statistical analyses.

The spatial boundary is the European Union. Data are collected in countries that were members of the union in 2018. This means that the analysis may not be a good representation of soils in other geographical regions and climate zones. The temporal boundary for this data set is that the sampling process was carried out during 2018. Therefore, there are potential differences in the data depending on seasonal changes.

In chapter 2.5, the indicators that will be analysed in this report are described. Other biological soil indicators that are measured by *Eurofins* are, for example, total microbial biomass, bacterial biomass, fungi biomass, the ratio between fungi and bacteria, and the amount of protozoa. The reason why these indicators are not included in this report is that biomass is not available in LUCAS.

2
Background

The soil is an important part of the Earth that often works in silence. In many eyes, soil is something to walk on and use for cultivation, but there is much more than only stones and dirt under our feet. The soil system is a complex system that includes, for example, animals, plants, and minerals. The interplay between the components of the soil system results in ecosystem services from the soil and also how the soil system changes with time.

In this chapter, information that is considered necessary to understand the report is described. The areas that are described are the components of soil, the human use of soil, and the theory of principal component analysis. Finally, the concept of soil health is described.

2.1 The components of soil

The soil is made up of air, water, and solid material [65]. The solid material contains 10-15% organic material (containing carbon atoms) and the rest is inorganic material, see Figure 2.1.

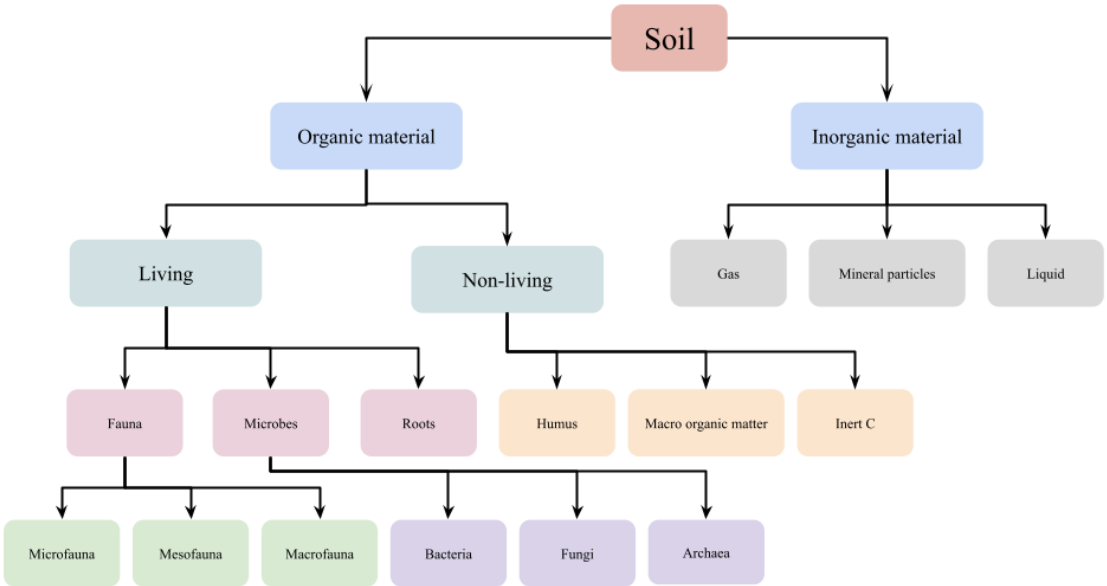

Figure 2.1: Hierarchy of the components in the soil. Inspiration from [57].

2.1.1 Inorganic material

Inorganic material is material that does not containing carbon atoms. It can for example be, small stones and minerals called mineral particles [69]. These particles contribute to 45-50% of the soil. The soil is also made up of cavities that are filled with liquid or gas. The most common gases in the soil are carbon dioxide and oxygen. These gases are used by animals and plants that live in the soil. The cavities represent 40-60% of the soil. It may differ between soils depending on the soil compaction. The most common liquid is water, and the water can behave in different ways [44]. Gravity water is pulled further into the ground as a result of gravity. Some of the gravity water may stick to smaller soil pores as a consequence of the surface tension of the water. The water molecules are affected by both the cohesive forces between the molecules and the adhesion forces from the soil particles. This leads to the fact that water is available for plant roots even when the gravity water has passed. There is also water called hygroscopic water that has a strong bound to the soil particles and is therefore not available to the plant roots. This is an important aspect for microorganisms because this water remains in the soil even in the case of dry conditions.

2.1.2 Organic material

Organic material is material that contains carbon atoms. As seen in Figure 2.1, the organic material can be divided into living and non-living [69].

2.1.2.1 Non-living organic material

Non-living material is, for example, macro-organic material, such as, dead leaf, twigs, and animals that have ended up in the soil. Some parts that are decomposed by terrestrial organisms become carbon dioxide, water, and minerals. The parts that are decomposed slow become humus instead, which is stored in the soil. Humus is important because it contributes with nutrients to plants and helps the soil to store water. To maintain the humus content, biomass needs to be added continuously. In the soil, there is even inert carbon that does not react with the surrounding environment.

2.1.2.2 Living organic material

Living organisms constitute the smallest volume fraction of the soil. To see some of this type of life, a microscope is needed. Even if this life is so small, it has an important impact on the life above the ground. More than 25 % of the biodiversity of Earth is found in the soil and a hand with soil includes up to one kilometre fungi and one billion bacteria [23]. Some part of the living in soil is plants roots. The roots contribute to stabilise the soil and make it possible for the plant to absorb nutrients [65]. The area around the roots is called the rhizosphere [44]. The rhizosphere is important because the exchange of nutrients between plants and soil organisms occurs there. The plants get energy through photosynthesis. This energy is used to produce chemicals that are secreted through the roots to the rhizosphere. The se-

creted chemicals can, for example, be carbohydrates and proteins. The combination of secreted chemicals and the availability of organic carbon makes the rhizosphere to be a good place for soil organisms such as bacteria, fungi, nematodes, and protozoa (these organisms are described more detailed below in this section). The rhizosphere acts as a foundation for many food chains in the soil. Microbes like bacteria and fungi need nutrients that the plant roots produce. The bacteria and fungi then become food for other larger organisms, such as, nematodes and protozoa. They in turn secrete nutrients and carbon to the soil. In the next step nematodes and protozoa become food for arthropods like spiders and insects which become food for larger animals such as snakes and birds. These food chains are even longer and they confirm one of the important role of soil.

Fauna is a group in the living part of the soil. The fauna can be divided into macrofauna, mesofauna, and microfauna [56][20]. In Table 2.1, some examples of organisms and animals for each group are shown.

Table 2.1: Examples of organisms and animals categorized into the groups microfauna, mesofauna, and macrofauna [56][20].

Microfauna *(< 0.1 mm)*	Mesofauna *(0.1-2 mm)*	Macrofauna *(> 2 mm)*
Nematodes	Mites	Earthworms
Protozoa	Springtails	Millipedes
	Potworms	Moles

The first subgroup of the soil fauna is the macrofauna. Here animals larger than 2 mm are found. Examples are earthworms, millipedes, and moles. Earthworms play an important role in maintaining the structure of the soil. The soil changes porosity when earthworms dig passages, which also influences water infiltration [56][20]. As mentioned before, a less compacted soil leads to larger root system for plants. The feces of the earthworms have a high level of nutrients which contributes to other animals and plants [37]. Earthworms eat mostly plants but can also eat animalistic bacteria, fungi, and feathers [18]. Their digestion starts with that the food is sucked into the body through a suppression. The worm does not chews the food, the food is pulled down from the mouth through the gullet where calcium is added to stabilise the calcium content in the blood. The food is broken down in the muscle belly and in the intestinal pepsin is added to break down the food even more. An earthworm needs to eat several times more than their body weight to get sufficient nutrients. Due to that, they naturally turn over and move the soil they live in [37]. There are several subgroups of earthworms. Three are endogeic, epigeic, and anecic. The subgroups live in different part of the soil and contribute in different ways. Epigeic earthworms eats organic material at the ground surface. Endogeic earthworms instead live in the upper soil layer. They digging horizontal tunnels and eat both organic and inorganic material. The last subgroup, anecic earthworms, live deeper into the ground and digging vertical tunnels to make it possible to reach food in the upper soil layer. Since earthworms are sensitive to soil pollutants, they are

a common bioindicator for contaminated sites [20]. Another animals that belong to detritivores that decompose organic carbon such as plant litter are millipedes [48]. Since millipedes moves above the ground and need a humid environment, is the species distribution low. This means that a specific species only lives in a small area. A consequence of this genetic isolation is that there are many different types of millipedes that are adapted to different environments. The body shape of millipedes is a series of similar segments of exoskeleton. The reason for why they need to live in humid environments is because they cannot close the openings between the segments and therefore are at greater risk of water loss. Different groups of millipedes use different methods for digging. Some groups penetrates the ground with the head first and use their collum to push away the soil. Other go into cracks in the soil and expand the crack by press their back against the ceiling. Some species are wedge shaped and can therefore slowly expand the cracks gradually as they move forward. Millipedes decompose larger organic material and make it possible for other organisms to continue the decomposition. Millipedes can act as a compliment in soils that do not have populations of earthworms. Finally, larger animals such as moles are included in the macrofauna too [49]. Moles are mammals and commonly live in meadow land, woodland, garden land, or cropland. They can grow between 11-16 cm long and they dig tunnels at a depth of 15-25 cm. These animals belong to the group of omnivores which means that they eat from the plant and animal kingdom. Their main food is, for example, earthworms, and insects.

The mesofauna includes organisms that are between 0.1-2 mm and they are commonly found 10-20 cm down in the soil [36]. These organisms contribute to the decomposition of organic material and the circulation of nutrients. Mites is an example of an arachnid included in the mesofauna [39]. It is a large group with more than 65 000 different species and fossils from the time period Devon are found. Even springtails live in the soil and are included in the mesofauna. They are also called Collembola and belong to the group of arthropods [33]. Like mites, even springtails have a long history with fossils from Devon. There are 415 different species of springtails in Sweden and there can be up to 750 millions of springtails per hectare in moist soils. Different individuals can have different appearance, the largest individuals can be 6 mm. Some species have no eye while other can have eight point eyes at each side of the head. Some groups bites while other sucks in the food. The group of springtails can be divided depending on the body shape. Symphypleona are rounder and eat living plants in the upper part of the soil. Arthropleona have instead a longer body and eat dead organic material such as plant litter. The last example from the mesofauna is potworms. It is a group of animals that live in the first 5 cm of the soil [32]. Potworms are worm-like animals that are structure in body segments. They are hermaphrodites and there are populations globally. Potworms are sensitive to climate changes and season based weather. Large difference result in reduced populations. These animals contribute to the soil health because they in the same way as nematodes and protozoa regulate populations of bacteria and fungi och decompose organic material.

The last subgroup of the fauna is the microfauna. This subgroup includes organisms

and animals that are smaller than 0.1 mm. Even these organisms contribute to the decomposition and circulation of nutrients in the soil [47]. Due to their influence of the nutrient level, they also regulate populations of other organisms. An organism included in the microfauna is nematodes, which is a worm-like organism. Nematodes is the group of multicellular organisms that is most represented on Earth [63]. Between 700-1050 species have been found in Sweden. They often have a symmetric body shape with narrower ends [54]. They have no respiratory system or organs for circulation, they change skin when they grow, and common movement is sinus-like pattern [63]. Nematodes live in many different environments as in water course, bottom sediments, and soil. They often live as parasites and can cause diseases in animals and plants. Even protozoa is a part of the microfauna [47]. Protozoa is a group of unicellular organisms which can be divided into three subgroups. They are ciliates, flagellates, and amoeba. They can live on bacteria, fungi, dead organic material, or as parasites. Their bodies are structured with an inner cell body with cytoplasm and organelles. This is inclosed with an outer cell membrane and an outer skeleton for some species. In the soil, they contribute to the dissolution of organic material and release nutrients such as nitrogen and phosphorus [47]. They also regulate the population of bacteria and thus the decomposition of organic material because they graze on soil bacteria [60].

The last group in Figure 2.1 is microbes, which has the subgroups archaea, bacteria, and fungi. Archaea was before a subgroup to bacteria but now it is an own group [8]. The first archaea were found in the 1970s by the American researcher Carl Woese and his research team [44]. First, they thought they had found a special type of bacteria that lived in extreme environments. The reason why archaea started to classified as their own domain was the differences in the molecular structure compared to bacteria and eukaryotes [8][44]. The theory is that the domain of archaea is the oldest of the three domains. The reason is that archaea today live in environments that are similar to the conditions on Earth when life was originated. There are many similarities between bacteria and archaea [44]. Examples are that they are similar in size, they move with flagelles, they form colonies, they reproduce by cell division, and they have similar body shapes. One thing that is different between bacteria and archaea is that their cell membranes have different structures. The cell membrane for archaea consists of lipids, amino acids, and sugar that cannot be found in bacteria. The cell membranes are also different compared to eukaryotes because the cells of eukaryotes can consist of chitin and cellulose, which does not happen for archaea. Even if archaea are prokaryotes like bacteria, their DNA is more similar to DNA from eukaryotes. Archaea contribute to the nitrogen cycle because they can fix nitrogen from the atmosphere. They are also decomposers that contribute to the circulation of organic material in the soil. Archaea collaborate with, for example, protozoa as a part of the digestion in the intestinal in ruminants.

The second type of microbes is bacteria. Bacteria is the type of organism with most individuals on Earth and they have lived here for more than three billion years [44]. Bacteria are unicellular prokaryotes. This means that they are one cell there their DNA is inclosed in a cell nucleus. Bacteria reproduce with cell division where

one cell divides into two cells. In favourable laboratory conditions with food and protection from other organisms, a bacteria can reproduce to five billion bacteria in one day. They have much lower reproduction rate in the soil because they, for example, can be eaten by other organisms, such as protozoa or that the conditions in the soil get too dry. Humidity is one important factor for bacteria because they need water to absorb nutrition and get rid of waste. In the case of too dry conditions, bacteria can hibernate and wait for more humid conditions. They also have the capacity to produce a biofilm of sugar, proteins, and DNA. This biofilm stabilise the pH in the soil and protect the bacteria from dehydration and antibiotics. Bacteria have an important role in many food chains that start in the soil. By decomposition of organic material they get the necessary energy and nutrients as nitrogen and phosphorus. This also leads to that the nutrients remain in the soil and are not washed away by water. The nutrients into the bacteria are also passed on to other organisms that eat the bacteria. Depending on the availability of organic material, the bacteria can adapt what they decompose, but the most common is that they decompose still fresh plants. The reason is that fresh plants consist of more sugar that are easier for bacteria to digest. One mechanism for how the digestion works is that nutrients diffuse from a high concentration outside the bacteria to the lower concentration inside [46]. This is called a passive transport and is thermodynamically spontaneous. Transport against the concentration gradient is also possible using different proteins in the cell wall. This is called active transport and requires energy. In the soil, most bacteria are located in the rhizosphere due to the readily available nutrients [44]. There are also some bacteria located outside the rhizosphere. These bacteria use enzymes to decompose longer carbon chains to shorter chains of sugar, fat, and amino acids. Bacteria have a role in the carbon and nitrogen cycles too. Aerobic bacteria have carbon dioxide as a by-product during the decomposition of organic material. The carbon dioxide is then used by plants through photosynthesis to produce new organic material. In the nitrogen cycle, nitrogen-fixing bacteria have en important role because they make nitrogen available for plants. Living organisms need nitrogen to produce amino and nucleic acids. The most nitrogen is available as gas (N_2) in the atmosphere. The strong triple bond in the nitrogen molecule makes the nitrogen unavailable for plants. Nitrogen-fixing bacteria fix the nitrogen and make it possible for plants to absorb and produce molecules that are a combination of nitrogen and oxygen. There are also bacteria that transform ammonium (NH_4^+) to nitrite (NO_2) and nitrate (NO_3^-). Bacteria are often associated with something negative and there are some negative affects with bacteria in soil too. There are, for example, bacteria that produce toxic alcohol that affects the soil and the living organisms in the soil negative. Other bacteria are pathogens to plants which affects the yield for farmers and leads to economical losses. Even if there are some negative consequences of bacteria in soil, the positive consequences are stronger. In summery, positive affects of soil bacteria are, for example, decomposition of organic material, they contribute to nutrition cycles and they protect the soil from toxic substances by decomposition of pollutants.

A specific type of bacteria is actinomycetes bacteria [44]. They live in the soil and are adapted for a wide range of pH. At first sign, they look like fungi due to their

filament and mycelium structure. But they are much smaller than fungi. There are theories that they produce mycelium to protect themselves against predators such as protozoa. Actinomycetes bacteria are also special because they have the ability to decompose cellulose and chitin even if they are long and complex carbon chains present in, for example, cell walls and the shells of insects [2][44]. Another difference from other bacteria is the genetic structure. DNA consists of pair of adenine-thymine and guanine-cytosine. Actinomycetes bacteria have a higher share of pair of guanine-cytosine [2]. This type of bacteria is also a common source of antibiotics. More than half of all antibiotics that are used by humans come from actinomycetes bacteria in the soil.

Bacteria can also be grouped into gram-positive (gram+) and gram-negative (gram−) bacteria, see Figure 2.2. Gram+ bacteria are larger and are characterised by their thick cell wall [22]. The reason for the name gram+ is that they give a positive result in a gram stain test [31]. Gram+ bacteria have the substance called peptidoglycal in their cell membrane. This substance gets a crystal violet colour after fixation with iodine. This marks out the bacteria if they are observed in an optical microscope. Due to their ability to produce spores and their cell wall, they are more resistant to environmental stressors, such as droughts or floods. With the thicker cell wall, they can resist physical stress such as osmotic pressure [62]. Gram− bacteria have a thinner cell wall and an outer cell membrane containing lipopolysaccharide (LPS) instead [22]. Gram− bacteria have less peptidoglycal in their cell membrane and do not get a crystal violet colour during a gram stain test due to the addition of ethanol. They get pink or red colour instead [30]. Due to the outer cell membrane, gram− bacteria are more resistant to stressors such as tilling and use of pesticides [7].

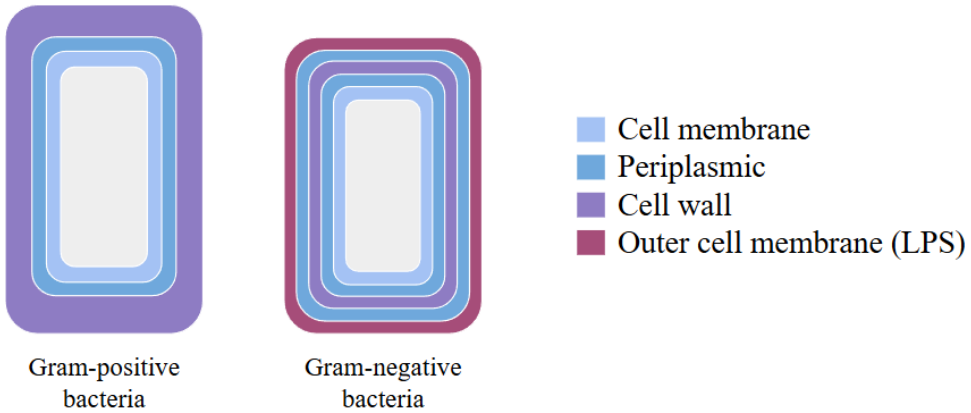

Figure 2.2: Visualisation of the differences in structure between gram+ and gram− bacteria [22].

The last subgroup of microbes is the fungi. Maybe the first thought when the word fungi is said is common edible mushrooms, such as chanterelle. But it hides even smaller and different fungi in the soil. Fungi are eukaryotes like plants and animals [44]. This means that they have cells with closed cell nucleus containing the

DNA. Fungi produce spores and hyphae that become a thread-like system called mycelium. As a consequence of the spread of the mycelium, some parts of the fungi can exist in soil with lower amount of nutrients and water [22]. A fungal hyphae can have a length of 2-15 μm and a diameter of 0.2-0.35 μm [44]. Even if that is much larger than bacteria, several hundred thousand fungal hyphae are needed to be visible to the human eye without a microscope. Fungi have a high growth rate for being such a small organism. They can grow up to 40 μm in one minute. If there are unfavourable conditions, the fungal spores can hibernate. Fungi decompose similar nutrient material as bacteria do. Since fungi are slower than bacteria, they get energy and nutrition from other sources too. They have the ability to break down rougher and more complex nutrient material, such as cellulose from plants. This is possible because fungi produce the enzyme phenol oxidase, which dissolves the lignin in the cellulose. The dissolved nutrients enter the fungi by osmosis in a similar way as for the bacteria. The nutrients are released from the fungi and become available to other organisms when the fungi die. Due to the fungal mycelium, small tunnels and corridors are created when the fungi die. This contributes to the transport of water and oxygen in the soil. Even bacteria can use these tunnels to protect themselves from predators like protozoa. The growth of the fungal hyphae makes it possible for the fungi to reach areas with higher nutrient content. This has proven to be a great advantage compared to bacteria. There is research showing that fungi can catch organisms, such as protozoa and nematodes, and decompose them [44]. This behaviour could be compared with hunting. Depending on the ratio of fungi and bacteria, the amount of available nitrogen changes. The nitrogen that becomes available when the fungi die is in ammonium, which is important for some plants. Other plants need instead nitrogen in nitrate and then nitrifying bacteria that transform ammonium to nitrate are necessary. A soil with more fungi often has a lower pH as a consequence of the enzymes produced by the fungi. Because nitrifying bacteria need a basic environment, they do not like soil with much fungi. Therefore, less nitrifying bacteria lead to more ammonium in comparison with nitrate in the soil.

The two types of fungi that are analysed in this report are saprotrophic fungi and mycorrhiza. Saprotrophic fungi decompose organic material in the soil. In contrast, mycorrhiza fungi work in symbiosis with plants. These fungi get energy from plants in exchange for nutrients and water. Thus, mycorrhiza forming fungi increase the plants ability to absorb nutrients from the soil [50]. The reason is that the fungi increase the available soil area due to a larger root system. Plants in symbiosis with mycorrhiza also have better resistance to stress, such as salinisation and pathogens. Agriculture has an influence on mycorrhiza fungi [13]. Aspects such as cultivation methods (organic or not), types and order of plants during the cultivation season, use of pesticides, soil tillage, and fertilization also affect mycorrhiza. One variant of mycorrhiza is arbuscular mycorrhiza [50]. This type of mycorrhiza penetrates the plant cells inside the plant. In the plant, the fungi create arbuscles and vesicles. Arbuscles are branches, and vesicles are storage structures. The penetration of the cell membrane increases the contact surface between the fungi and the plant and makes it possible for the fungi to give the plant nutrients. Arbuscular mycorrhiza

is totally dependent on the plant because it cannot decompose organic material on their own. Therefore, they are called obligate biotrophs. A large fraction of plant species have symbiosis with arbuscular mycorrhiza, around 78%. It contributes to the soil because they produce the protein glomalin, which is an important part in the storage of carbon in soil. The other common type of mycorrhiza forming fungi is ectomycorrhiza. They do not penetrate the cells in the plants. Instead, they penetrate between the cells. There is fewer different plants that have symbiosis with ectomycorrhiza, only 2%. Ectomycorrhiza often have symbiosis with trees. Therefore, they affect large ecosystems even if only 2% of the plants are included in the symbiosis.

The soil system is a complex part of Earth and there are still a lot of uncertainties about how different functions and components of the soil are structured and how they relate and affect each other. This is a difference compared to a complicated system, that can be built up of many components as well, but where the connection between the components are known [17]. It is possible to restore the system if a component breaks down, changes or is removed. If the component is replaced with a new similar one, the system is restored and works as before. A complicated system can, for example, be a car or a computer. A complex system is more difficult to understand and predict due to their emergent properties [41]. If a component is changed or removed, the whole system can change properties or collapse. It can also be difficult to restore a complex system, it can be irreversible. Examples of irreversible complex systems are some ecosystems.

2.2 Human use of soil

The soil is used by humans in many ways. One ecosystem service related to soil that is generally known is crop production by agriculture. Other ecosystem services from soil are, for example, water purification, mitigation of climate change, and human health [14]. Because of the complexity of the soil system, an analysis of a single ecosystem service does not result in a valid conclusion. Instead, many components of the system and how they are related to each other need to be taken into account.

Depending on the purpose for which the land will be used, different soil indicators play a role in the contribution to a healthy soil. For example, a healthy soil for forestry may have different meanings depending on the type of tree that will be cultivated. The purpose in woodlands is to have trees that are growing, but it is the opposite in cropland. In a meadow it is a mixture of different types of plants and flowers, while an owner of a lawn often wants to have grass only and not many other plants, such as moss, in the garden.

As an example of how different usage of land leads to different views on soil indicators, a forester that wants to grow pine, prefer a coarse-grained and dry soil [38]. If the purpose is to grow fir instead, the soil should be fine-grained and moist. Factors like coarse fragment size and humidity are included in the profile of the soil and contribute to the production capacity and fertility of the soil. The profile of the

soil is a combination of the humidity of the soil, the soil type, and the classification of the soil. The humidity of the soil can be divided into dry, fresh, moist, and wet. The class is determined by the level of the ground water. A dry soil has a ground water level of 2-5 meters, a fresh soil 1-2 meters, a moist soil less than 1 meter, and a wet soil has a ground water level at the surface or just below. The soil type depends on the decomposition of the material and the amount of organic material. The coarse fragments in soil can be categorised into clay (< 0.002 mm), silt (0.002-0.06 mm), and sand (0.06-2 mm). The most common soil type in Sweden is till as a consequence of the latest ice age. The reason is that material such as soil residues and plants were dragged with the ice. The result was the soil type till which is a mixture of different soil types and coarse fragments. Peat, which consists of partially decayed plants as a consequence of anaerobic conditions, is also an example of a soil type.

Another example is that for croplands, soil compaction, humus content, and erosion are factors of interest [65]. Soil compaction is related to the structure of the soil and the degree to which the soil is packed. A compact soil makes the soil less rich in nutritions due to the high density. A higher density makes it more difficult for the roots of plants to penetrate the soil [71]. Therefore, a compact soil often leads to use of fertilizers to a greater extent. A reduction in the amount of roots leads to lower plants above ground too. A combination of these consequences of compact soil results in a reduced yield. The main reason for compaction of soil is the use of heavy agricultural machinery in combination with wet soil [65]. It takes a long time to restore land with compacted soil, and therefore it is better to use preventive measures instead. It can, for example, be to use wheels with larger contact surface or only use machinery at some places. To restore compacted soil, mechanical and biological loosening may be used. This means that either tools or root systems loosen up the soil and make it more porous. The humus content is the amount of organic material from animals and plants stored in the soil [6]. A higher humus content contributes to a better capacity to store water and handle nutrients such as nitrogen and phosphorus. The humus content depends on the difference between the amount of organic material that is removed from the soil as a result of, for example, the yield and the amount of organic carbon added to the soil from plants [65]. Methods to increase the humus content are to add more organic material to the soil by leaving some part of the yield at the field and to cultivate over longer periods during the year to increase the production of biomass. Perennial plants increase the humus content because their root system needs more area compared to annual plants. The last factor, erosion, reduces the production capacity because fertile soil is removed from the field. Then, nutrients in the soil are transported to nearby water coarse and increase the risk of eutrophication. The risk for erosion depends on the coarse fragments in the soil and the vegetation [68]. A coarse grained soil erodes on a smaller scale while a soil with a high sand content implies a higher risk of erosion. Even land with less vegetation are at a greater risk for erosion as a consequence of wind. Due to that, covering the land with vegetation during the whole year and a high humus content are good measures to reduce the risk of erosion.

This section has described that the human use of soil is affected by different soil indicators and their relation and they are in turn affected by the human use of the soil. The definition of a healthy soil depends of which perspective that is used.

2.3 Methods to describe soil

As described in Section 1, there are different ways to describe soil. In this section, the methods used to measure the soil indicators used in this report are described.

2.3.1 Methods to measure characteristic soil indicators

During sample collection in LUCAS, approximately 500 g of soil from 5 points, see Figure 2.3, were mixed. The four outer points are located 2 m away along each geographical direction. Stones, vegetation residues, grass and litter were removed from the soil before sampling. Then samples were stored at -20 °C until the laboratory analysis at the University of Tartu in Estonia [29].

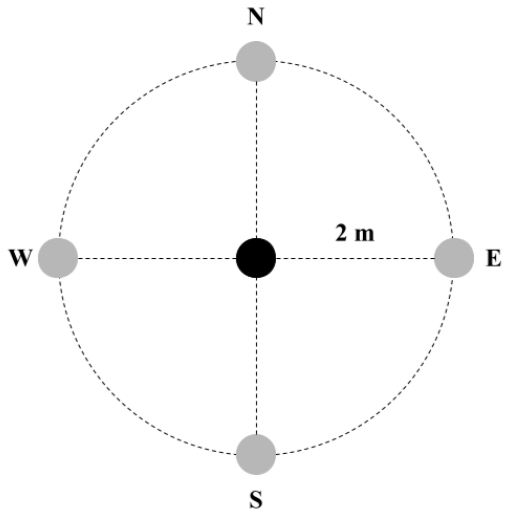

Figure 2.3: Collection pattern for each sample site [29]. The four outer points are located 2 m away along each geographical direction.

Depending on the characteristic soil indicator, different methods can used [29]. In the following, each method is briefly described. To obtain the value of the coarse fragment, the fine fragments are separated by sieving. The coarse fragments are defined as fragments with a diameter between 2 and 60 mm. For the content of clay, sand and silt, laser diffraction is used. The definitions for the different types of grain sizes are that mineral fractions smaller than 0.002 mm belong to clay, between 0.002-0.063 mm belong to silt and between 0.063-2 mm belong to sand. In order to measure pH value, soil is dissolved in $CaCl_2$ or H_2O to give a 1:5 volume ratio. Then pH value is measured in the solution with a glass electrode. To measure the amount of organic carbon in the soil, elementary analysis by dry combustion is used.

With this method, carbon is converted into CO_2 and the CO_2 in the resulting gas is quantified. The content of carbonates is measured with a volumetric method. To determine the content of phosphorus, the phosphorus is dissolved in $CaCO_3$ and after that detected and measured by spectrometry. A modified Kjeldahl method is used to determine the amount of nitrogen in the soil. Atomic absorption spectrometry is used for the content of potassium and finally the bulk densities are calculated based on the mass and volume.

2.3.2 Methods to measure biological soil indicators

The methods that are described in this section are PLFA analysis that is used by Eurofins and eDNA analysis that is used by LUCAS.

2.3.2.1 Phospholipid fatty acids analysis

One method of detecting organisms in soil is Phospholipid fatty acids (PLFA) analysis that is based on Gas Chromatography Mass Spectrometry (GC-MS) [21]. This is a combination of two techniques: gas chromatography (GC) and mass spectrometry (MS). GC separates compounds by their volatilization capacity. If the purpose is to analyse a solid material, the material undergoes solvent extraction, outgassing, or pyrolysis before the gas chromatography. The separation is carried out in a glass column bent as a coil. The column is coated with, for example, silicone, grease, or wax [61]. This phase is called the stationary phase. More volatile molecules in the sample interact less with the stationary phase and go through the column faster, resulting in a shorter retention time. Less volatile molecules interact more with the stationary phase and have a longer retention time instead. Because of the different retention times, the molecules become separated. After the GC part, the molecules continue to the MS part. First, the molecule becomes ionized and then goes through an electric field and starts to oscillate [45]. Because the oscillations are related to the mass-to-charge ratio, the identification of the molecule can be done. The result from the GC-MS is compared with published libraries in order to identify the molecules.

The molecules of interest during soil biodiversity measurements are PLFA, which are found in the cell membrane of living organisms [22]. The molecule is polar and the hydrophilic part is located in contact with the water outside the cell and inside the cell while the hydrophobic part is in contact with each other, see Figures 2.4 and 2.5 [9]. The PLFAs are different for groups of fungi, bacteria, and protozoa, and therefore, this method can be used to detect and identify the different organisms in soil samples [22]. PLFA measurements only measure the living organisms present because PLFAs quickly degenerate after the organisms have died.

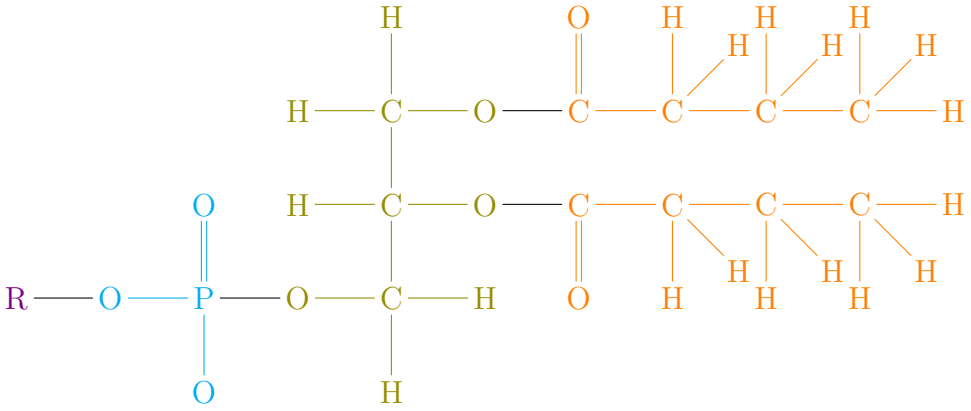

Figure 2.4: Example of molecule structure of a phospholipid [9]. The blue part is the phosphate group, the green part is glycerol and the orange part is the fatty acids. The structure of phospholipid can changes by different R-molecules (in purple) and different fatty acids. The R-molecule is often a smaller polar molecule, for example, choline. The fatty acids can have other lengths and there can also be other combinations of saturated and unsaturated molecules. In this example, both acids are saturated. If the acid is unsaturated, there are at least one double bond between two carbon atoms. The left side of the molecule is hydrophilic and want to be in contact with water. Instead, the right part of the molecule is hydrophobic and want not to be in contact with water.

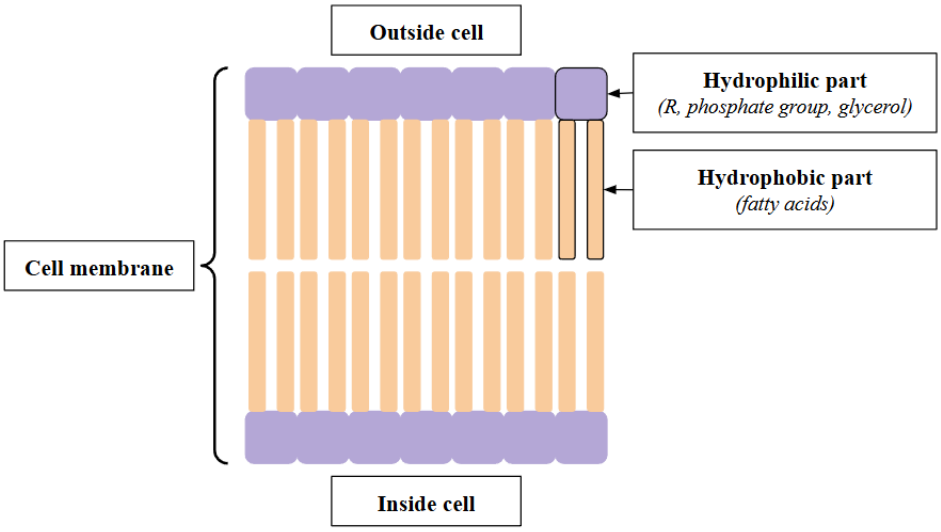

Figure 2.5: Structure of a cell membrane built up by phospholipids [9]. The hydrophilic parts are in contact with the water outside the cell and inside the cell while the hydrophobic part is in contact with each other.

2.3.2.2 Environmental DNA

Another method for measuring biodiversity is the use of eDNA [67]. This implies that genetic traces in the environment are detected and analysed. Genetic traces

can be detected in the environment through, for example, skin cells, dead cells, plant material, or eggs. The general steps in an eDNA analysis are shown in Figure 2.6 [66]. The concepts in Figure 2.6 are described in the following section.

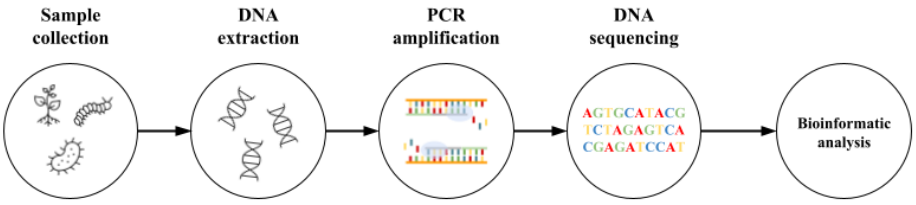

Figure 2.6: The steps involved in an eDNA analysis [67].

Samples can be collected from water, air, sediment, or soil depending on the purpose of the research [66]. During sample collection, it is important to protect the sample from contamination. Therefore, new decontaminated equipment is used for each new sampling site. The equipment can be decontaminated with, for example, soapy water, distilled water, ethanol, or bleach [43]. During the collection of soil samples, soil from different places in the site is used [67]. The collected soil samples are mixed together before laboratory analysis.

After collection, DNA is extracted in a laboratory and the DNA can be analysed. The DNA extraction is done to obtain clear DNA molecules from the collected soil. Steps in the DNA extraction can, for example, be shaking, incubation, or centrifugation [5]. When DNA is extracted, there are many molecules that are not of interest. Therefore, Polymerase Chain Reaction (PCR) amplification is used to mass copy the relevant DNA fragments that relate to the organisms of interest. Components that need to be available to perform PCR amplification are extracted DNA, a primer for the specific set of organisms and genomic region, DNA polymerase, and free nucleotides. The first step is called denaturation, where the temperature is increased to around 90-95 °C, causing the two DNA strains to separate [12], see Figure 2.7b. During a reduction in temperature, which is called annealing, the primer is connected to one of the DNA pieces, see Figure 2.7c. The primer is a part of the DNA sequence that is unique for the specific set of organisms. Other parts of the DNA sequence have high variations between individuals. During elongation, the enzyme polymerase starts to copy the remaining code of the DNA sequence, see Figure 2.7d.

There are different methods to read the DNA depending on the purpose of the analysis. During metabarcoding, a set of primers are used to amplify specific marker genes to detect specific organism groups. In metagenomics, primers are not used and all DNA is analysed instead of only some specific molecule parts such as in metabarcoding [67].

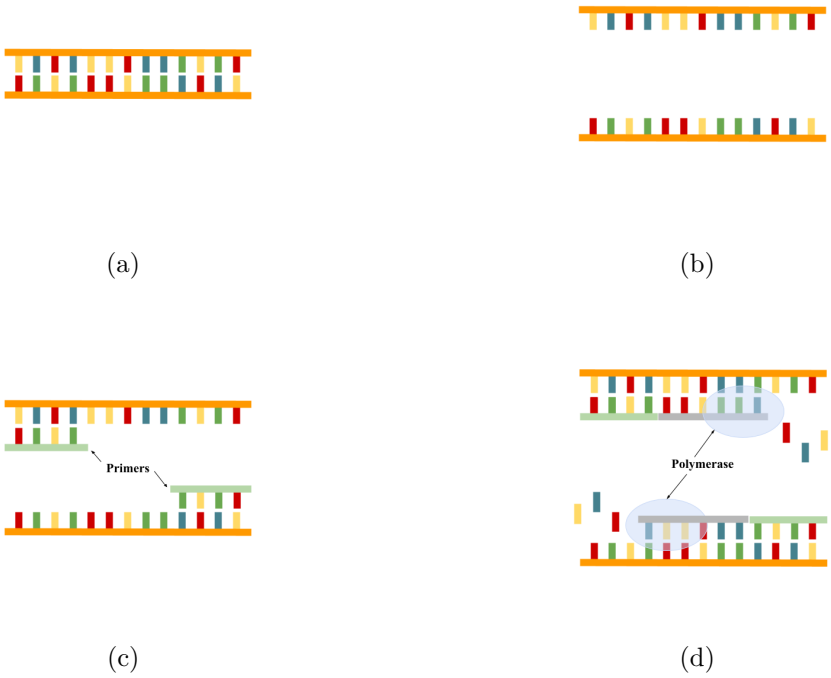

Figure 2.7: The procedure in PCR amplification [5]. (a) Initial state of the DNA sequence. (b) Denaturation around 90-95 °C. Division of the DNA sequence because of increased temperature. (c) Annealing around 50-60 °C. Connection between primers and the DNA sequences. (d) Elongation around 70 °C. The polymerase continues to construct the remaining code of the DNA sequence.

Only the DNA part between the primers is of interest for the analysis. The polymerase continues to construct the copied DNA sequence to the end of the first strain, see Figure 2.8. Because of that, the DNA sequences get the length of interest first after three cycles. Each cycle of PCR amplification results in a duplication of the number of DNA sequences. This means that the amplification is exponentially related to the number of cycles of the amplification.

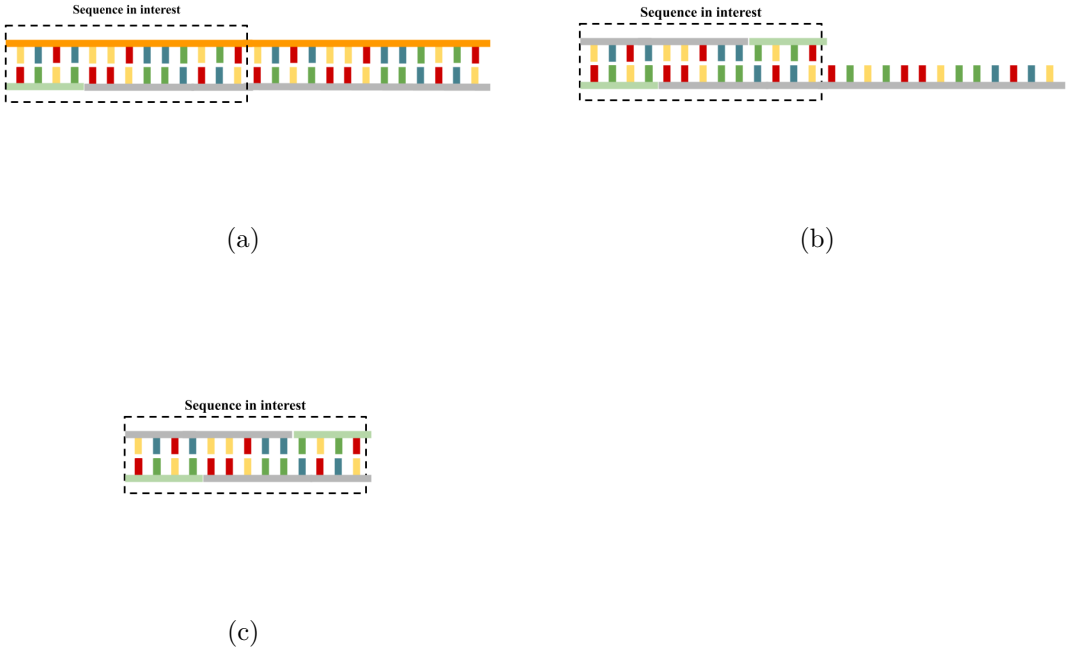

Figure 2.8: Visualisation of the change in length of the DNA sequence that is amplified during PCR amplification. The sequence has the length of interest first after three cycles (a) After cycle one. (b) After cycle two. (c) After cycle three.

Sequence identification can be performed with different methods. Historically, Operational Taxonomic Units (OTUs) has been the most commonly used method, but today other more modern methods are used too [16]. The method with OTUs is based on clustering of genes by sequence similarities. The threshold value for similarity is often set to 97%. A more modern method is to use Amplicon Sequence Variants (ASVs) [15]. This method detects sequence variations at a single-nucleotide level. It is more sensitive and detects and adjusts errors in the DNA sequences. During the production of ASVs, the tool DADA2 is used to denoise data.

The eDNA is classified into different taxonomic ranks depending on the available information. Taxonomy is used to organize the relation between organisms into a hierarchy [10]. In biology, the binomial system, created by Carl von Linné, is internationally used. Organisms are classified into taxa after a taxonomic hierarchy, see Figure 2.9. There are three domains; Bacteria, Archaea, and Eukaryote. They are then divided into six kingdoms; Animalia, Plantae, Fungi, Protista, Archaea, Bacteria, and Eubacteria. For each step in the hierarchy, there are more divisions and the resolution becomes more detailed. In this report the rank *Domain* is defined as the lowest rank and *Species* as the highest rank. The taxonomy is used to make it easier to organise organisms and to facilitate communication. The binomial system is based on this taxonomic hierarchy. According to the nomenclature, the name of the organism is a combination of genus and the species epithet [10].

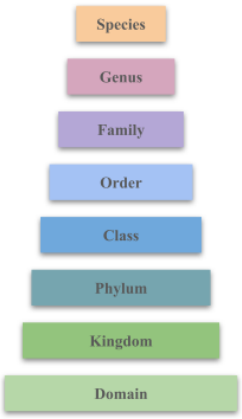

Figure 2.9: The hierarchy of biological taxonomy [10]. Domain is the widest rank divided into Bacteria, Archaea, and Eukaryote. For each step in the hierarchy, there are more divisions and the resolution becomes more detailed. Here, *Domain* is defined as the lowest rank and *Species* as the highest rank.

Figure 2.10 shows an example of taxonomic ranking for the organism *Niallia circulans* [51]. This is also an example of the binomial system, where the name of an organism is a combination of genus and species epithet.

Domain	**Kingdom**	**Phylum**	**Class**	**Order**	**Family**	**Genus**	**Species**
Bacteria	Bacillati	Bacillota	Bacilli	Bacillales	Bacillaceae	Niallia	Circulans

Figure 2.10: Example of the taxonomy ranking for the organism *Niallia circulans* [51]. The name of the organism is a combination of genus and species epithet according to the binomial system developed by Carl von Linné [10].

There are some advantages with eDNA compared to other methods for biodiversity measurements, such as, for example, PLFA analysis [5]. eDNA is more precise in organisms determination, more sensitive and therefore a better measurement method for detecting rare organisms, and more cost-effective. There are also some disadvantages with eDNA. The method gives information about which organism are detected, but without knowledge about, for example, number of individuals and if they are healthy or not. According to [19], the amount of DNA from a specific organism is proportional to the amount of biomass for the same organism. The difficulty is to calculate the proportional constants because they are different for different species. There is also a risk that organisms that are present are not detected (false negative) or that organisms that are not present are reported (false positive), this is called errors. Potential errors can be divided into if they are related to sample collection or laboratory work [11]. If samples are taken from a watercourse,

there is a risk that flows of water from other watercourses have brought DNA to the watercourse where the samples have been taken [5]. Another risk is that there still remains DNA from populations of organisms that are extinct. This gives errors in the result. There are also risks of failure during the laboratory analysis. Early failure in the DNA analysis can give large differences in the final result.

2.4 Principal component analysis

To search for linear correlations between a set of indicators, a Principal Component Analysis (PCA) may be used [35]. The aim of the method is to find correlations between linear combinations of the original datasets. The method for this analysis is shown here with an example in two dimensions. In the example, an analysis between two datasets X and Y is described. Both datasets are from *Land Use/Cover Area frame statistical Survey* (see Section 3.1.1) and describe the pH in topsoil measure with $CaCl_2$ and H_2O, respectively. In Figure 2.11, a correlation is shown between the pH based on $CaCl_2$ and the pH based on H_2O.

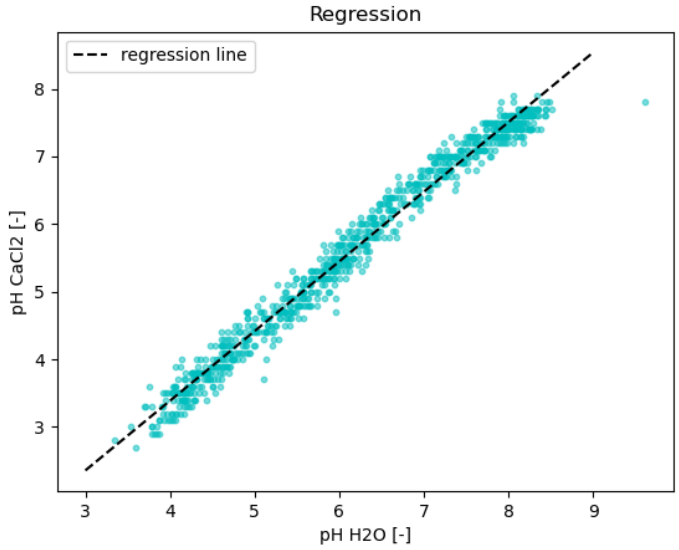

Figure 2.11: Linear correlation between the pH based on $CaCl_2$ and the pH based on H_2O. Data from [29].

According to the regression line shown in Figure 2.11, it can be assumed to be a linear correlation between the two indicators with the constants a and b, see Eq. 2.1.

$$pH_{CaCl_2} = a \cdot pH_{H_2O} + b \tag{2.1}$$

The first step in a principal component analysis is to autoscale the data and remove the offset. This means that data is subtracted with the mean of the data set ($\bar{x}$) and divided with the standard deviation of the data set (σ), see Eq 2.2.

$$\frac{x_i - \bar{x}}{\sigma} \rightarrow \tilde{x}_i \tag{2.2}$$

If Eq. 2.2 is implemented for the data sets used in Figure 2.11, the variables are described by the system of equations shown in Eq. 2.3. The new variables are centered around origo.

$$\begin{cases} X = \widetilde{pH}_{CaCl_2} = \frac{pH_{CaCl_2} - \overline{pH_{CaCl_2}}}{\sigma(pH_{CaCl_2})} \\ Y = \widetilde{pH}_{H_2O} = \frac{pH_{H_2O} - \overline{pH_{H_2O}}}{\sigma(pH_{H_2O})} \end{cases} \tag{2.3}$$

The correlation in Figure 2.11, may be written with Eq. 2.4, where $\mathbf{v}_1$ is a normalized vector along the regression line, σ_1 is the standard deviation along the regression line and R_1 is a stochastic variable with variance one. The vector $\mathbf{v}_2$ is a normalized vector perpendicular to the regression line describing the noise added to the correlation, σ_2 is the standard deviation of the noise and R_2 is a stochastic variable with variance one.

$$(X, Y) = \mathbf{v}_1 \cdot \sigma_1 \cdot R_1 + \mathbf{v}_2 \cdot \sigma_2 \cdot R_2 \tag{2.4}$$

In Figure 2.11, the data points stretch along the correlation line according to σ_1 but the data point stretch little in the perpendicular direction according to σ_2. So $\sigma_1 >> \sigma_2$, which is a fact when a strong correlation exists.

In order to calculate the correlation vectors ($\mathbf{v}_1$ and $\mathbf{v}_1$) and their respective standard deviations (σ_1 and σ_2), the PCA method uses the covariance matrix, see Eq. 2.5. The formula to calculate the covariance is shown in Eq. 2.6, where E is the expectation value operation. Due to the autoscaling of the data sets, $E[X]$ and $E[Y]$ are zero.

$$\begin{bmatrix} cov[X,X] & cov[X,Y] \\ cov[Y,X] & cov[Y,Y] \end{bmatrix} \tag{2.5}$$

$$cov[X,Y] = E[(X - E[X])(Y - E[Y])] \tag{2.6}$$

The correlation vectors are the eigenvectors of the covariance matrix and the standard deviations are obtained from the eigenvalues (λ_i) according to Eq. 2.7.

$$\sigma_i = \sqrt{\lambda_i} \tag{2.7}$$

For the data with $\widetilde{pH}_{CaCl_2}$ and $\widetilde{pH}_{H_2O}$ the calculated eigenvalues are shown in Eq. 2.8 and the respective eigenvectors are shown in Eq. 2.9. The directions of the eigenvectors compared to the datasets are shown in Figure 2.12.

$$\begin{cases} \lambda_1 = 1.99 \Rightarrow \sigma_2 = 1.41 \\ \lambda_2 = 0.01 \Rightarrow \sigma_1 = 0.10 \end{cases} \tag{2.8}$$

$$\begin{cases} \mathbf{v}_1 = [-0.71, -0.71] \\ \mathbf{v}_2 = [-0.71, 0.71] \end{cases} \tag{2.9}$$

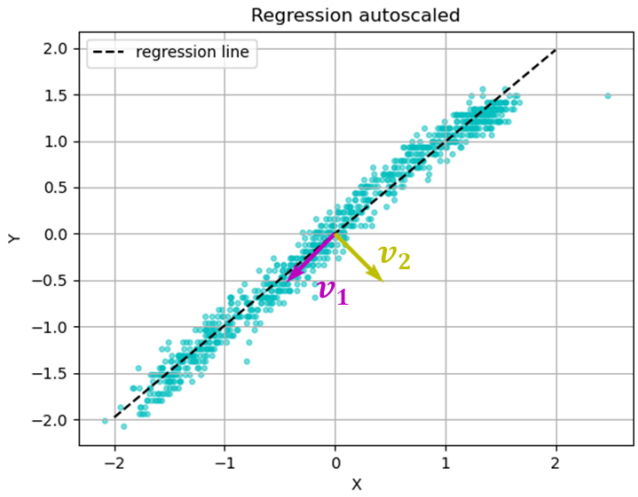

Figure 2.12: Linear correlation between the scaled pH based on $CaCl_2$ and the scaled pH based on H_2O. The eigenvector with largest standard deviation is shown in purple and the eigenvector with smallest standard deviation is shown in yellow.

The result of the analysis indicates that there is a strong correlation along the eigenvector v_1 because the standard deviation for that direction is much larger than for the direction related to the eigenvector v_2. Since $\sigma_1 >> \sigma_2$, this means that the principal component, i.e. the main correlation is given by the eigenvector v_1.

With two or three datasets, correlation can be detected by regression analysis or simply plotting data points. One of the advantages with PCA is that correlations can be detected even when more datasets are included in the analysis.

In a multi-component situation, the components are again given by the normalized eigenvectors $\mathbf{v}_i = [v_{i1} \quad v_{i2} \quad v_{i3} \quad ...]$. If the corresponding standard deviation (σ_i, calculated from the eigenvalue λ_i) is $>> 1$, a true correlation is found. The correlation may include more than two variables, but the datasets j with $|v_{ij}| << 1$ are considered not to be part of the correlation.

2.5 Soil health

A concept that can be used to try to assess soils is soil health. With this concept, an overall picture is used to try to get a diagnosis about if a soil is healthy or not. Soil health is a concept enclosed with confusion and uncertainties. Due to the complexity of the soil system, it is still a research area with many questions. Neither, there is no common definition of the concept of soil health, and it can have different meanings for different people and purposes. As a consequence of that, there are difficulties to measure and indicate soil health. The defintion of soil health according to *Natural*

Resources Conservation Service [52] is:

> *"The continued capacity of soil to function as a vital living ecosystem that sustains plants, animals, and humans."*

The expectation is that the soil will contribute with clean water and food for the entire population [23]. Due to the increase of the global population to an expected number of 10 billion in 2050, the demand of soil contributions to humanity becomes higher. The functions that can be connected to soil ecosystem services are biodiversity, water regulation, soil structure, nutrient cycling, and pollution management [52], see Figure 2.13.

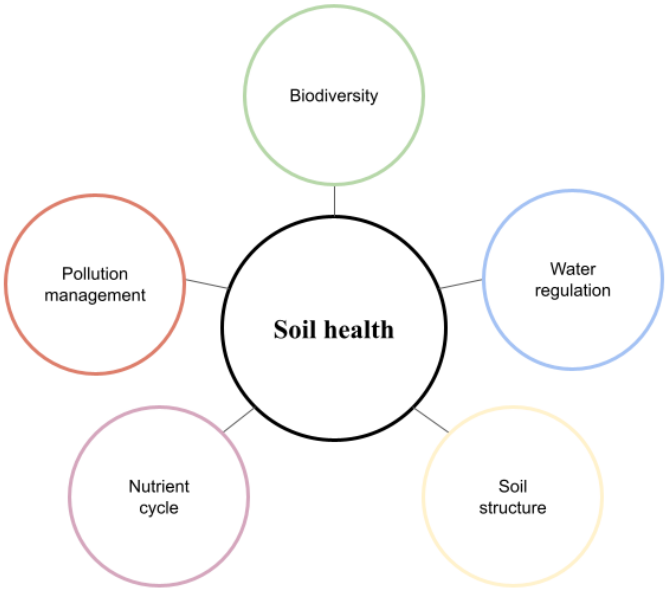

Figure 2.13: The five functions that are connected to soil ecosystem services according to the Natural Resources Conservation Service [52].

The functions in Figure 2.13 are related and influence each other [52]. Biodiversity refers to that there is a diversity of different organisms such as plants, microorganisms, and animals in the soil [14]. This is a broad concept that includes both the variance of ecosystems, species, and genetic differences [53]. The diversity in ecosystems is related to different types of nature, for example, forests, deserts, and lakes. Depending on the type of nature, different organisms are adapted to live in that arbitrary place. The diversity of species means that there are many different species that live in the same area. Instead, genetic differences focus on changes in DNA for the same species, according to [53]. In common for these three types of biodiversity is that parts of the system have different functions and complement each other. This project will focus on diversity of species since that can be measured by eDNA (see chapter 2.3.2.2). Sometimes biodiversity leads to that other organisms can replace vanished functions in the system, as a consequence of loss of organisms. In other cases, there are functions that only a few organisms can perform [56]. These types

of functions are called specialized functions and are more sensitive to biodiversity loss compared to generalized functions that many organisms can implement. The biodiversity is a result of millions of years of development and changes all the time [14]. The biodiversity is affected by humans in many ways, for example, by agriculture. One difference between now and the time before the agricultural revolution is the scale of human exploitation of ecosystem services. In modern agriculture, mass production and optimisation are important factors. An example is that fewer rotations of cultivated crops lead to less biodiversity. In addition, deforestation to create farmland is one cause of the loss of biodiversity. The extinction rate of species is nowadays 50-100 times higher than the natural rate [14]. Water regulation includes control of water flows through the soil. It can, for example, be related to erosion, water infiltration, and water cycling. Water regulation is, among other things, related to roots in the rhizosphere and the structure of the soil that is affected by the cultivation methods. The meaning of soil structure is that soil provides the essential conditions for plant and human use. The structure affects the conditions for living organisms and the production capacity. Cultivation methods such as tilling and the use of cover crops are essential factors for soil structure. The soil also influences the nutrient cycle in the soil because substances such as carbon, nitrogen, and phosphorus are decomposed and transformed by the soil biota. Optimisation of naturally occurring nutrients in the soil also leads to less need for added nutrients in terms of fertilizers. This is good both from a climate perspective and an economic perspective. The last function, pollution management, contains reduction of pollutants through, for example, filtration and decomposition by soil biota.

The understanding of the concept of soil health is a part of the work for sustainability. In 2015, the United Nations agreed upon the Sustainable Development Goals (SDGs) with the aim of peace and prosperity for people and the planet [70]. The 17 goals relate in different ways to the three dimensions of sustainable development; ecological, economical, and social. A factor that contributes to many of the goals is soil health due to its societal impact through their natural contributions to humanity, such as crop production and water infiltration [52]. The connection with the ecological dimension is that increased soil health leads to increased resilience and protection against the effects of climate changes [23]. The relation to the economy is both the value of ecosystem services and the influence of the market price of land. Information about soil health may affect the possibility of selling land and at what price. The social dimension includes, for example, the mental health related to being outside in nature and the availability of food and water. Some goals with a strong connection to soil health are: zero hunger (SDG 2), clean water and sanitation (SDG 6), reduced inequalities (SDG 10), responsible consumption and production (SDG 12), climate action (SDG 13), and life on land (SDG 15). To make it possible to develop strategies to improve soil health, it is necessary to understand which factors that influence soil health and in what way.

3
Method

This chapter includes a description of the datasets that have been used for the analysis and the methodology for the data management and data analysis. The script used for data management was written in Python and the libraries *Numpy*, *Seaborn*, *Tabulate*, and *Matplotlib* were used. The areas that are described in this chapter are the available data, how data was sorted, which data was removed from the data sets, and which types of graphs were used in the data analysis. Finally, the statistical tests used in the analysis are described.

3.1 Available data

To make it possible to investigate potential correlations between some chosen biological soil indicators (see Section 3.2) and characteristic soil indicators (see Table 4.5), a dataset from LUCAS has been used. This dataset is described in Section 3.1.1.

3.1.1 LUCAS

The *Land Use/Cover Area frame statistical Survey* (*LUCAS*) is a collaboration between the *European Statistical office* (*EUROSTAT*), the *Directorate General responsible for Agriculture* and the *Joint Research Centre* (*JCR*) as a result of a decision of the *European Parliament* [24]. The survey, including measurements in the topsoil, mapping, and investigating the changes in land use in the members of the European Union. Since 2009, new measurements have been made every few years to make it possible to monitor changes over time. In this report, data from the latest published survey from 2018 are used. The sampling period was between April and October 2018 and 18 984 sampling points were successfully completed [29]. In 2018, soil biodiversity was added as a new type of analysis in the survey. In total, 885 sample points were used for this purpose. The distribution of the sample sites is shown in Figure 3.1. The different regions are based on the geoscheme of the *United Nations* [64].

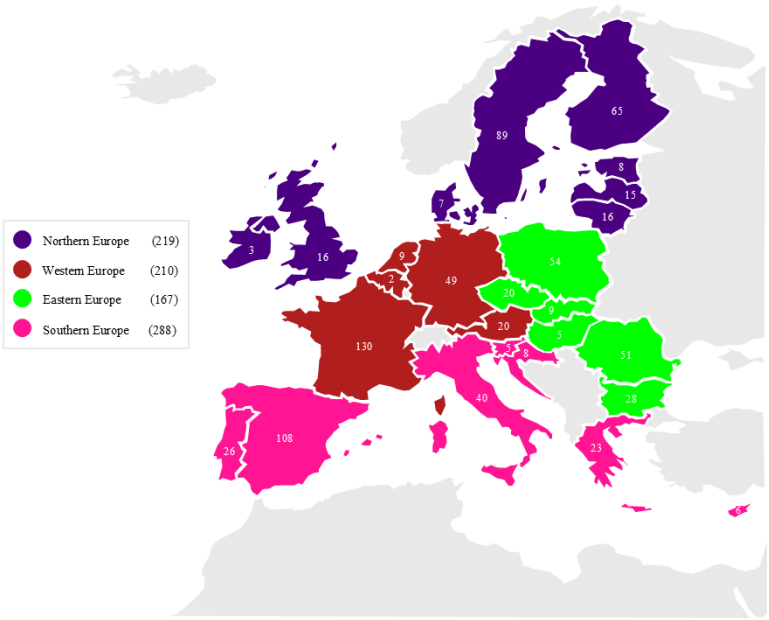

Figure 3.1: Number of samples in LUCAS 2018 [29].

The biodiversity data from LUCAS consisted of DNA sequences that afterwards have been managed and finally presented in an Excel file with the structure visualised in Table 3.1. The file with eDNA data has the structure that each row shows a detected DNA sequence, and the columns show classifications and the number of ASVs for each sample.

Table 3.1: File structure for eDNA from LUCAS. The blue cells represent classification group followed by phylum, class, order, family, genus, and species (see Figure 2.9). The red cells represent each sample site and related number of ASVs. The last column includes the total number of ASVs summed over all samples for the specific ASV ID.

ASV ID	Domain	...	Species exact	confidence	ASV ID	Sample 1	...	Total
⋮	⋮	...	⋮	⋮	⋮	⋮	⋮	⋮

The eDNA data related to LUCAS that were used gave information on the number of ASVs for each detected DNA sequence per sample. One file for data related to bacteria and one file for data related to fungi. There was taxonomic information for each DNA sequence, including the name of domain, phylum, class, family, genus, and species (see Table 3.1). There were differences in resolution between different DNA sequences. This means that some DNA sequences were only classified to, for example, phylum level and others to, for example, species level. The number of copies of identified ASVs for a specific organism was not directly correlated with the amount of biomass of that organism. Instead, the numbers of copies of ASVs for different organisms were comparable with each other. Therefore, a considerably higher number of ASVs for one DNA sequence indicated a higher presence of that organism in that specific sample. In total, for the 885 sample sites, 144 634 different

DNA sequences were available for bacteria and 65 479 were available for fungi. For each sample site, additional information and soil indicators were also available in another file (see Table 3.2). The different kinds of land types that are handled in this report are woodland, grassland, bareland, cropland, shrubland, and wetland. Woodland is defined as an area of more than 0.5 hectares and with at least 10% tree crown cover [27]. Grassland is an area with a dominant cover of grass and forbs. An area covered with moss and lichens or with no dominant vegetation in 90% is called bareland. Cropland is covered with annual or perennial agricultural crops. The definition of shrubland is land covered with shrubs and plants less than 5 meters high for at least 10% of the area. Finally, wetlands are areas that are wet for sufficiently long periods. This means that animals and plants are adapted to or dependent on the wet conditions.

Table 3.2: File structure for metadata from LUCAS. The red cells represent the sample number. They are comparable with the red cells in Table 3.1 for eDNA. The other columns give additional information of each sample.

Barcode ID	Country	Land use	Coarse fragments	Clay content	Sand content	Silt content	pH
1	FI	Woodland	5	1	84	15	3.4 4.16
2	PL	Grassland	-	-	-	-	6.5 6.87
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Table 3.2: Continued

Barcode ID	Electrical conductivity	Organic C	CO_3^{2-} content	P content	N content	K content	Bulk density
1	10.96	346.6	0	44	10.4	193.5	0.43 0.56
2	118.7	401.3	3	16.6	29.6	65.9	- -
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

The file types for the initial data files were .xlsm for eDNA and .xlsx for metadata. All data files were converted to the file type .csv before data management.

3.2 Calculation of biological soil indicators

The following section describes the biological soil indicators that have been used in this report. The data that have been used to calculate the indicators is the total number of copies of the detected ASVs. For a more detailed description, see Section 2.3.2.2. In the following equations, the total number of copies of the detected ASVs

is denoted as *the number of ASVs*.

Bacteria ratio
Gram+ and gram− bacteria are sensitive to different stressors. A balanced ratio or just above 1 can be assumed to be an indicator of a healthy soil [22]. A value below 1 is often an indicator of stress in terms of, for example, tillage and pesticides. In contrast, a value above 1 can be an indicator of stress in terms of, for example, drought or floods. Human activities, such as agriculture, stress gram+ bacteria. Therefore, a ratio below 1 is expected for the soils analysed in this report. The indicator is calculated by Eq. 3.1.

$$gram+/gram- = \frac{the\ number\ of\ ASVs\ of\ gram+}{the\ number\ of\ ASVs\ gram-} \tag{3.1}$$

Actinomycetes
Actinomycetes is a specific group of gram+ bacteria that is important for soils. They are bacteria with a fungi-like structure [22]. They play an important role in resistance to soil disease due to their secretion of antibiotics and competition with pathogenic fungi. The indicator is calculated by Eq. 3.2.

$$actinomycetes = \frac{the\ number\ of\ ASVs\ of\ actinomycetes}{the\ total\ number\ of\ ASVs} \cdot 100\% \tag{3.2}$$

Saprotrophs
Saprotrophic fungi is an important group of fungi that contribute to the soil nutrition cycle [22]. The organic material is broken down by enzymes that are excreted by saprotrophs. The nutrients that become available from the organic material contribute to increased plant growth and can also be taken up by fungi. The indicator is calculated by Eq. 3.3.

$$saprotrophs = \frac{the\ number\ of\ ASVs\ of\ saprotrophs}{the\ total\ number\ of\ ASVs} \cdot 100\% \tag{3.3}$$

Mycorrhiza
Mycorrhiza forming fungi increase the root surface and stabilise the structure of the soil due to their symbiosis with plants [13]. They also increase the absorption of nutrients (such as nitrogen and phosphorus) for plants and protect the soil from stresses such as drought, salt, and pathogens. Because of their symbiosis with plants, organic agriculture and crop variation affects the growth of mycorrhiza. Also, reduction of chemical pesticides, tillage, and addition of phosphorus are beneficial for the growth of mycorrhiza [13]. The indicator is calculated by Eq. 3.4 [22].

$$mycorrhiza = \frac{the\ number\ of\ ASVs\ of\ mycorrhiza}{the\ total\ number\ of\ ASVs} \cdot 100\% \tag{3.4}$$

Diversity index
The diversity index is an indicator of the biodiversity in the soil [22]. The indicator is based on the Shannon-Wiener index that both takes the number of different organisms and the total number of organisms into account. The indicator is calculated by Eq. 3.5, where i loops through all detected organism groups. H is the diversity index, n is the number of groups of organisms that are analysed, p_i is the proportion of the entire community made up of the organism group i. A higher value means higher diversity, and a value of zero means that there is only one group of organisms. Higher diversity is often related to healthier soils.

$$H = -\sum_{i=1}^{n} p_i \cdot ln(p_i) \tag{3.5}$$

3.3 Types of analyses

To investigate possible correlations between the characteristic soil indicators and the biological soil indicators, different types of analyses were done. The first one was an analysis to investigate if the biological soil indicators were related to geographical location. The reason for this analysis was to see if larger geographical differences could be ruled out from the rest of the analysis or not. If not, geographical location can have an underlying impact on the following analyses. The followed analyses were even performed for data related to Spain and France, solely. The reason for this was to reduce the risk of regional climate impact (see more in Section 3.4).

The second analysis was to investigate whether the biological soil indicators depended on land use. For this and the geographical analysis, a hypothesis test was implemented to analysis differences in mean values (see more in Section 3.7.1).

The correlation analysis included one part related to linear correlations in two dimensions between one biological soil indicator and one characteristic soil indicator. In that analysis, R^2 values for each linear correlation was calculated (see more in Section 4.3). The second part was the PCA. In that part, linear correlations in multi dimensions were investigated. Then, a PCA was performed for each biological soil indicator.

The last analysis was to investigate possible correlations between the biological soil indicators. Even in this analysis, R^2 for each correlation was calculated.

3.4 Sorting of data

In the first step of data management in *Python*, the datafile was converted to a list with each row as an element. The structure of the list of data is shown in Eq. 3.6 for the metadata and in Eq. 3.7 for the eDNA data.

$$meta_{list} = [['Barcode\ ID',\ 'Country',\ 'Land\ use',\ldots],['1',\ 'FI',\ 'Woodland'\ldots],\ldots] \tag{3.6}$$

$$eDNA_{list} = [['ASV\ ID',\ 'Domain',\ 'Phylum',\ldots,\ 'Sample\ 1',\ldots],['d184\ldots',\ 'Bacteria',\ 'Proteobacteria'\ldots,\ '0',\ldots],\ldots] \tag{3.7}$$

The general method to sort out the relevant information from the eDNA was to first make a function that calculated the values of the biological soil indicators for a specific sample. This was done by looping through the $eDNA_{list}$ and registering the number of ASVs related to the specific sample. Depending on the biological soil indicator, the number of ASVs was registered for the relevant rows and the indicator value was calculated (see Section 2.5). With another function, which looped through all samples, a indicator value for each sample was calculated, and all values were returned as a list. Below, an example of a python script to calculate the value of an arbitrary biological soil indicator for a specific sample is shown. The input parameter *eDNA_list* was the list shown in Eq. 3.7 and the input parameter *sample* was the column number in the eDNA file entered as an integer for the sample in interest.

```
def indicator(eDNA_list, sample):

    """
    # eDNA_list: List there each element is a list representing each row in the eDNA file.
    # All cells in each row is represented as a string element.

    # sample: Column number in the eDNA file entered as an integer for the sample in interest.

    # indicator_value: The float value of the biological soil indicator in interest.
    """

    # Initial value of the DNA detections
    relevant_DNA_detections = 0
    iteration = 0


    # Loop through all elements in eDNA_list
    for element in eDNA_list:


        # The first row headlines
        if iteration == 0:
            iteration += 1

        else:
            iteration += 1


            # Registrate the taxonomic name (here phylum)
            # Registrate the number of ASVs
            taxonoic_name = element[2]
            detections = int(element[sample])


            # Control if the taxonic name indicates the relevant name
            if phylum == 'relevant name':
                relevant_AVS += detections


    # Calculation of the indicator value
    # Implement the number of ASVs in related function
    indicator_value = function(relevant_AVS)

    return indicator_value
```

To sort out gram+ and gram− bacteria the phylum names were used. According to [3] and [1], DNA sequences with the phylum names *Actinobacteriota* or *Firmicutes* together form the group of gram+ bacteria. The rest of the detected bacteria were classified as gram− bacteria. The actinomycetes bacteria were sorted in the same way, but only the phylum name *Actinobacteriota* was used as keyword. Due to the high taxonomic richness and diversity of fungi, the *FungalTraits* database was used during the fungi classification progress [59]. The database connected taxonomic ranks to the primary lifestyle of fungi such as *parasites*, *wood saprotrophs*, and *arbuscular mycorrhizal*. The highest resolution in the database was *Genus*, and therefore the comparison with the eDNA data was done by that resolution. The groups of fungi that were of interest for the biological soil indicators were saprotrophs and mycorrhiza. The primary lifestyles related to saprotrophs and mycorrhiza are shown in Table 3.3. All elements in $eDNA_{list}$ were looped through and the genus names were registered. The genus names were compared with the names in the database. If a match was found, the primary lifestyle for that name was compared

with the lifestyles of interest, see Table 3.3. If the lifestyle found in the database was consistent with a lifestyle for saprotrophs, the organism was classified as a saprotroph. In the same way, if the lifestyle found in the database was consistent with a lifestyle for mycorrhiza, the organism was classified as a mycorhhiza. The DNA sequences in the eDNA data with a taxonomic resolution lower than genus or with a genus name not included in the database were not classified as neither saprotrophs nor mycorrhiza.

Table 3.3: Primary lifestyles in the FungalTraits database related to saprotrophs and mycorrhiza fungi [59].

Saprotrophs	Mycorrhiza
Litter saprotrophs	Ectomycorrhizal
Wood saprotrophs	Arbuscular mycorrhizal
Soil saprotrophs	
Dung saprotrophs	
Pollen saprotrophs	
Nectra/tab saprotrophs	
Unspecified saprotrophs	

The geographical information about the samples was the country name. To make the statistical analysis more general, the samples were instead grouped into geographical regions [64], those shown in Figure 3.1. The reason why the geographical analysis is interesting is that the climate should have an impact on soil indicators and the geographical regions in Europe have different climate [40]. Even the regions north, south, east and west are large and have some climate variations. Therefore even smaller regions were analysed. Correlation analyses were performed individually for France and Spain. By doing a more narrow analysis concentrated on one country, the risk to climate impacts is reduced. The reasons why France and Spain were used for the individual analyses were that they were the countries with the most data points distributed over different types of land use. Sweden, for example, had many data points too, but most of them were for woodland and other categories of land uses had too few data points to do a statistical analysis.

The diversity index can be calculated for a specific taxonomic rank. Therefore, a relevant rank was decided prior to the calculation. A higher taxonomic rank gives a more detailed analysis but if the number of ASVs that are classified to that rank is low, the results can be misleading due to a reduction of input data. Therefore, a taxonomic rank as high as possible, but without too few ASVs is desirable. The diversity index have been calculated for both *genus* and *species* for the correlation analyses. This is done to reduce the risk of missing important correlations because of the used data.

3.5 Removed data

Some groups of data were removed from the analysis. For the geographical analysis, data points without geographical information were removed before plotting. *WETLAND* had only one data point and *OTHERS* had only four data points, and therefore these sets of data were not included in the analysis.

Depending on the purpose, the limit for sufficient resolution of the DNA classification was set to different taxonomic ranks. For example, DNA sequences only classified to a lower taxonomic rank than Phylum, were not possible to use for the calculation of the bacterial ratio and the presence of actinomycetes. This limit was used since the branching in the taxonomic hierarchy that distinguish these bacteria groups are at that rank (see Section 2.5). DNA sequences with lower taxonomic resolution were removed from the calculations. For the other biological soil indicators, the limit was set to the taxonomic rank Genus (or species for one diversity index). The reason for this was not related to the taxonomic hierarchy, but instead related to the available data. For the biological soil indicators based on fungi, the reason was that the FungalTraits database [59] used the taxonomic rank Genus as the limit (see Section 3.4). In the eDNA data for fungi, a set of DNA sequences was only classified to domain with the name Fungi. These data points were classified as noise because of the low resolution and the risk that the DNA sequences actually belong to another eukaryote.

For analysis that included more than one soil indicator, only data points having information about all soil indicators were included.

3.6 Types of graphs

For the geographical analysis, a combination of a bar chart and a scatter plot was used. For this graph, the libraries *Seaborn*, and *Matplotlib* were used. All data points were classified by country name and geographical region in Europe (*NORTH*, *EAST*, *SOUTH*, *WEST*, or *NO DATA*). There was only one data point in the group *NO DATA*, and thus this group was omitted before the plotting. In one dictionary the data points were grouped after geographical region in Europe and in another dictionary the data points were grouped after both geographical region in Europe and country name. The mean value and the standard deviation were calculated for each region based on the first dictionary. The heights of the bars were based on the mean value for each group and the error bars were based on the standard deviation. The scatter plot of the data points was based on the second dictionary and was plotted by the library *Seaborn*. The key information about the geographical region in Europe determined the class in the histogram and the country name the colour of the data point.

The graph for the analysis of the land use was done in a similar way as the geographical analysis. The differences were that the data points were classified into land use

(*WOODLAND*, *GRASSLAND*, *BARELAND*, *CROPLAND*, *SHRUBLAND*, *WETLAND*, or *OTHERS*) and that the data points were plotted in the same colour as the bar for the related land use. There were less than five data point in the groups *WETLAND* and *OTHERS*, and these groups were omitted before the plotting (see Section 3.5). This means that the dictionary with two information parts as key used for the geographical analysis, was not necessary for this type on analysis.

For the correlation analysis, scatter plots were used. These were plotted with the library *Matplotlib*. All combinations of biological soil indicators were plotted to make it possible to do a comparison. The colour of the data point was related to the land use.

3.7 Statistical test

To make it possible to mathematically draw conclusions about the research questions, two statistical tests were implemented. The hypothesis test was used to analyse the differences between the mean values of two datasets. The regression test and the PCA were used to search for correlations between datasets.

3.7.1 Hypothesis test

To mathematically drawn conclusions about whether the biological soil indicators were dependent on geographical location and land use, a hypothesis test was implemented. The test investigated the probability that two data sets with the same population mean value get two different mean values with the available data points. The null hypothesis was that the two datasets had the same population mean values and the confidence level was set to 99%. The data sets are assumed to be normally distributed. The result of the test was an interval in which the difference in the mean value was located with a confidence level of 99%. If the interval included the value zero, this indicated that the null hypothesis could not be rejected with a confidence of 99%. The confidence interval was calculated with Eq. 3.8. $\overline{X}_1$, $\overline{X}_2$ and s were stochastic variables that were dependent on the data sets. $\overline{X}_1 - \overline{X}_2$ was the difference between the mean values of the data sets. $\mu_1 - \mu_2$ was the difference between the population mean values. n_1 and n_2 were the numbers of data points for each data set. x was set to 2.326 according to the value in the t-distribution related to the confidence level (99%) and the number of degrees of freedom (∞) [42]. s was the pooled variance calculated with Eq. 3.9, there s_1 and s_2 were the standard deviations of the two data sets. This statistical test was done for the mean values two by two.

$$\mu_1 - \mu_2 = \overline{X}_1 - \overline{X}_2 \ \pm \ xs\sqrt{\frac{n_1 + n_2}{n_1 n_2}} \tag{3.8}$$

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \tag{3.9}$$

3.7.2 Regression test

The regression analysis was done with the commands *polyfit* and *poly1d* from the library *NumPy*. The command *polyfit* returned a list of the coefficient values for the regression line, and the command *poly1d* returned the function expression. The regression line was plotted by looping through all the x values in the data set and calculating the related y values according to the function expression. To calculate the value of R^2, Eq. 3.10 was used [55]. $\hat{y}$ was the value according to the function expression, y_i was the value according to the data set, and $\bar{y}$ was the mean value of the y values in the data set.

$$R^2 = 1 - \frac{\sum (y_i - \hat{y})^2}{\sum (y_i - \bar{y})^2} \tag{3.10}$$

3.7.3 Principal component analysis

To search for linear correlations between the biological soil indicators and the characteristic soil indicators in the dataset from LUCAS a principal component analysis (see Section 2.4) was performed. For each analysis all soil indicators shown in Table 3.2 and one biological soil indicator individually were included. The eigenvalues and the corresponding standard deviations were calculated for each covariance matrix. To find potential correlations, the magnitudes of the standard deviations were compared. A large ratio indicated a possible correlation. Because the purpose of the principal component analysis was to find correlations related to the biological soil indicators, only correlations where the biological soil indicator had a significant impact were of interest. If the analysis showed strong correlations where the biological soil indicator had low impact, these correlations were not relevant for the purpose. The principal component analyses were done for each biological soil indicator for France and Spain. The dataset of pH_{H_2O} was not included due to the strong linear correlation with pH_{CaCl_2}, see Section 2.4.

3.8 Relation between the presence of actinomycetes and the ratio of gram+ and gram−

In order to display the ratio of firmicutes and gram−, a scatter plot with the bacterial ratio on the x axis and the presence of actinomycetes on the y axis was plotted (see Eq. 3.11).

$$\begin{cases} x = \frac{actinobacteriota + firmicutes}{gram-} \\ y = \frac{actinobacteriota}{actinobacteriota + firmicutes + gram-} \end{cases} . \tag{3.11}$$

Eq. 3.11 can be converted to Eq. 3.12

$$y = \frac{x - \frac{firmicutes}{gram-}}{x + 1} \tag{3.12}$$

Thus, a data point with no presence of firmicutes falls on the curve shown in Eq. 3.13.

$$y = \frac{x}{x+1} \tag{3.13}$$

The bigger the fraction $\frac{firmicutes}{gram-}$, the more below that curve the data point falls.

4
Results

In this chapter, the results of the data analysis are shown. Firstly, the results related to land use, and the associated statistical tests are shown. Finally, the results from the regression analyses and the PCA are shown.

4.1 Geographical location

The result of the geographical analysis is shown in Figure 4.1. The mean values and standard deviations related to Figure 4.1 are compiled in Table 4.1.

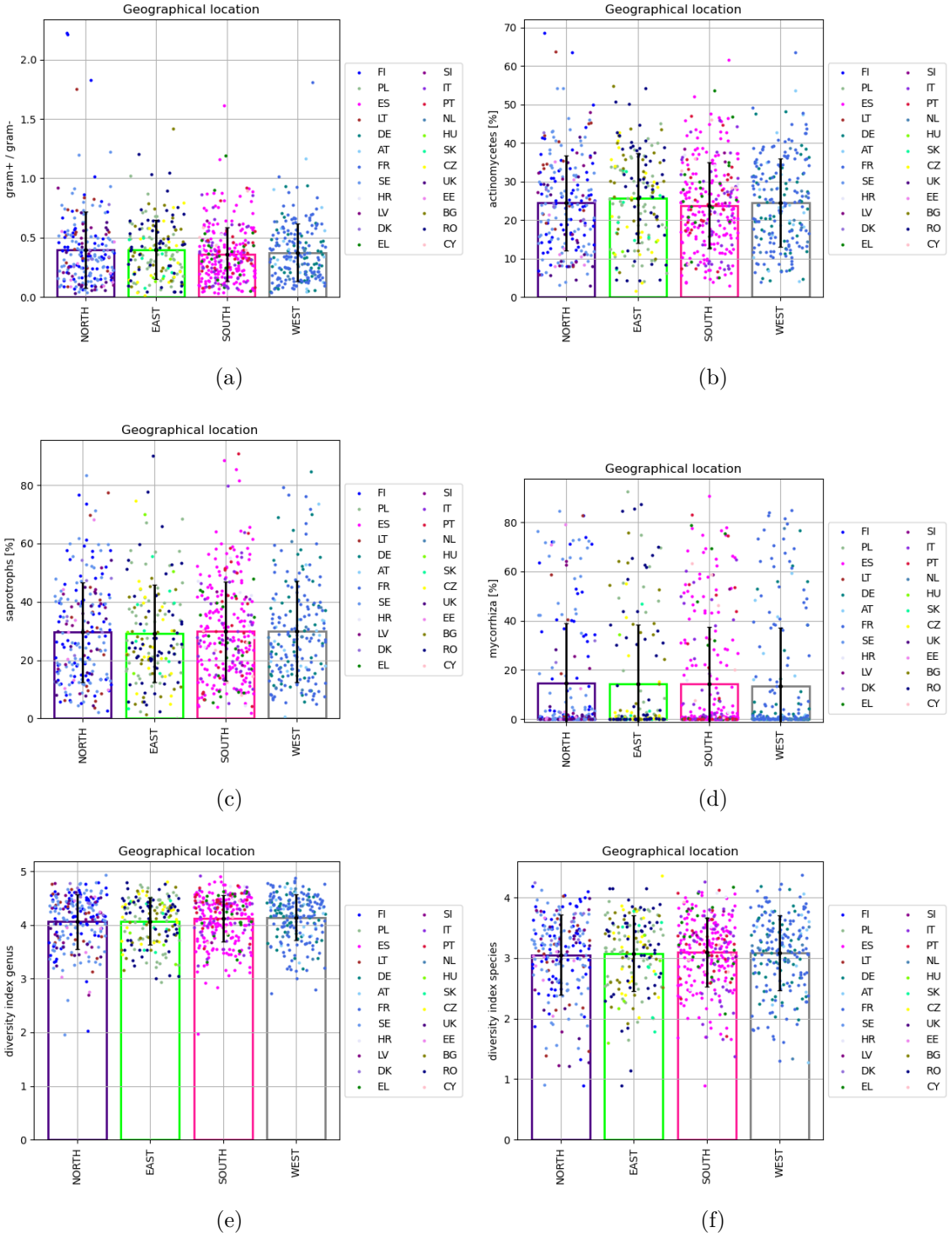

Figure 4.1: Visualisation of how the ratio of gram+ and gram− bacteria and the presence of actinomycetes, saprotrophs and mycorrhiza depend on geographical location. The last two graphs show the diversity index calculated for genus and species, respectively. The indicators without shown unit are unitless. The data points are grouped into the European regions of north, east, south and west. The bars represent the mean values and the black errorbars are the standard deviation of each group of data. The colours of the data points are related to the country names shown in the description box to the right of the graph. (a) Variation in ratio of gram+ and gram− bacteria. (b) Variation in presence of actinomycetes bacteria. (c) Variation in presence of saprotrophic fungi. (d) Variation in presence of mycorrhiza forming fungi. (e) Variation in diversity index based on genus. (f) Variation in diversity index based on species.

Table 4.1: Compilation of the mean values and standard deviations related to the geographical analysis for the ratio of gram+ and gram− bacteria and the presence of actinomycetes, saprotrophs and mycorrhiza. Also the values for diversity index calculated for genus and species, respectively. The values relate to Figure 4.1. The last row (rejection) indicates whether the null hypothesis was rejected for any of the pair wise comparisons of each region. A rejected null hypothesis means that the compared data sets do not have the same population mean value with a confidence level of 99% (see Section 3.7.1). If no pairwise comparison was rejected, the result shows *No*, otherwise *Yes*.

		gram+/gram−	Actinomycetes [%]	Saprotrophs [%]	Mycorrhiza [%]
Mean	North	0.40	24.49	29.55	14.60
	East	0.40	25.72	29.10	14.32
	South	0.36	23.80	29.96	14.32
	West	0.38	24.51	29.79	13.43
Std	North	0.32	12.32	17.13	24.39
	East	0.25	11.69	16.66	23.99
	South	0.23	11.13	16.91	23.10
	West	0.24	11.40	17.44	23.59
Rejection		No	No	No	No

Table 4.1: Continue

		Diversity index genus	Diversity index species
Mean	North	4.06	3.05
	East	4.07	3.08
	South	4.13	3.10
	West	4.14	3.08
Std	North	0.51	0.67
	East	0.43	0.62
	South	0.43	0.57
	West	0.43	0.62
Rejection		No	No

4.2 Land use

The result of the analysis with general data of land use is shown in Figure 4.2. The values of the mean values and standard deviations related to Figure 4.2 are compiled in Table 4.2.

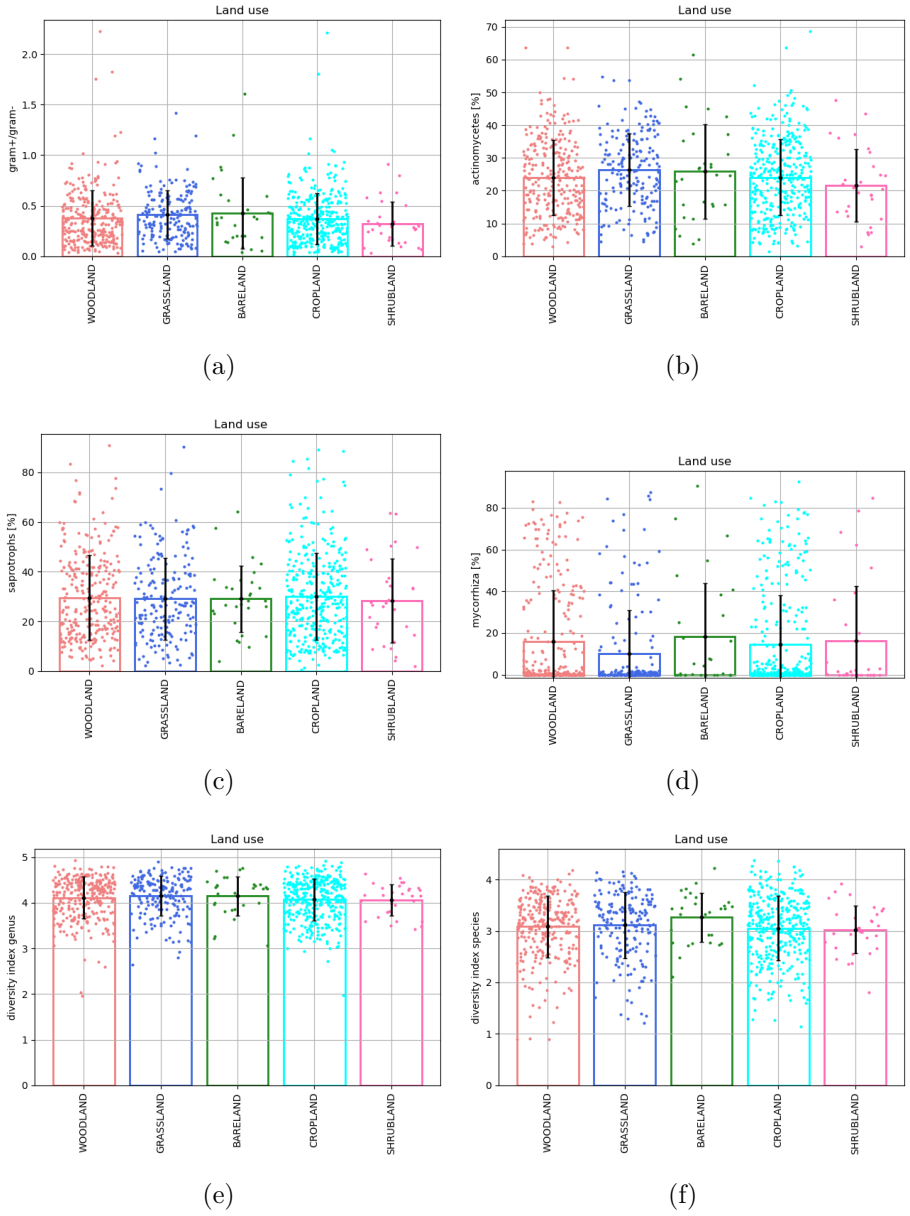

Figure 4.2: Visualisation of how the ratio of gram+ and gram− bacteria and the presence of actinomycetes, saprotrophs and mycorrhiza depend on the land use. The last two graphs show the diversity index calculated for genus and species, respectively. The indicators without shown unit are unitless. The data points are grouped into the land types woodland, grassland, bareland, cropland, and shrubland. The bars represent the mean values and the black errorbars are the standard deviation of each data group. (a) Variation in ratio of gram+ and gram− bacteria. (b) Variation in presence of actinomycetes bacteria. (c) Variation in presence of saprotrophic fungi. (d) Variation in presence of mycorrhiza fungi. (e) Variation in diversity index based on genus. (f) Variation in diversity index based on species.

Table 4.2: Compilation of the mean values and standard deviations related to the analysis of land use for the ratio of gram+ and gram− bacteria and the presence of actinomycetes, saprotrophs and mycorrhiza. Also the values for diversity index calculated for genus and species, respectively. The values relate to Figure 4.2. The last row (rejection) indicates whether the null hypothesis (see Section 3.7.1) was rejected for any of the pair wise comparisons of each land uses. A rejected null hypothesis means that the compared data sets do not have the same population mean value with a confidence level of 99% (see Section 3.7.1). If no pairwise comparison was rejected, the result shows *No*, otherwise *Yes*.

		gram+/gram−	Actinomycetes [%]	Saprotrophs [%]	Mycorrhiza [%]
Mean	Woodland	0.38	24.02	29.62	15.85
	Grassland	0.41	26.44	29.15	10.20
	Bareland	0.43	25.87	29.08	18.16
	Cropland	0.37	24.12	30.09	14.40
	Shrubland	0.32	21.65	28.43	16.09
Std	Woodland	0.28	12.32	17.07	24.47
	Grassland	0.24	11.69	16.47	20.81
	Bareland	0.35	11.13	13.27	25.78
	Cropland	0.25	11.13	17.52	23.76
	Shrubland	0.22	11.40	16.99	26.28
Rejection		No	No	No	Yes

Table 4.2: Continued

		Diversity index genus	Diversity index species
Mean	Woodland	4.11	3.08
	Grassland	4.15	3.11
	Bareland	4.15	3.27
	Cropland	4.07	3.05
	Shrubland	4.07	3.02
Std	Woodland	0.46	0.60
	Grassland	0.43	0.63
	Bareland	0.43	0.45
	Cropland	0.46	0.63
	Shrubland	0.34	0.46
Rejection		No	No

4.2.1 Land use analysis in France

Figure 4.3 shows the values for the biological soil indicators for data from France. The data points are grouped after land use.

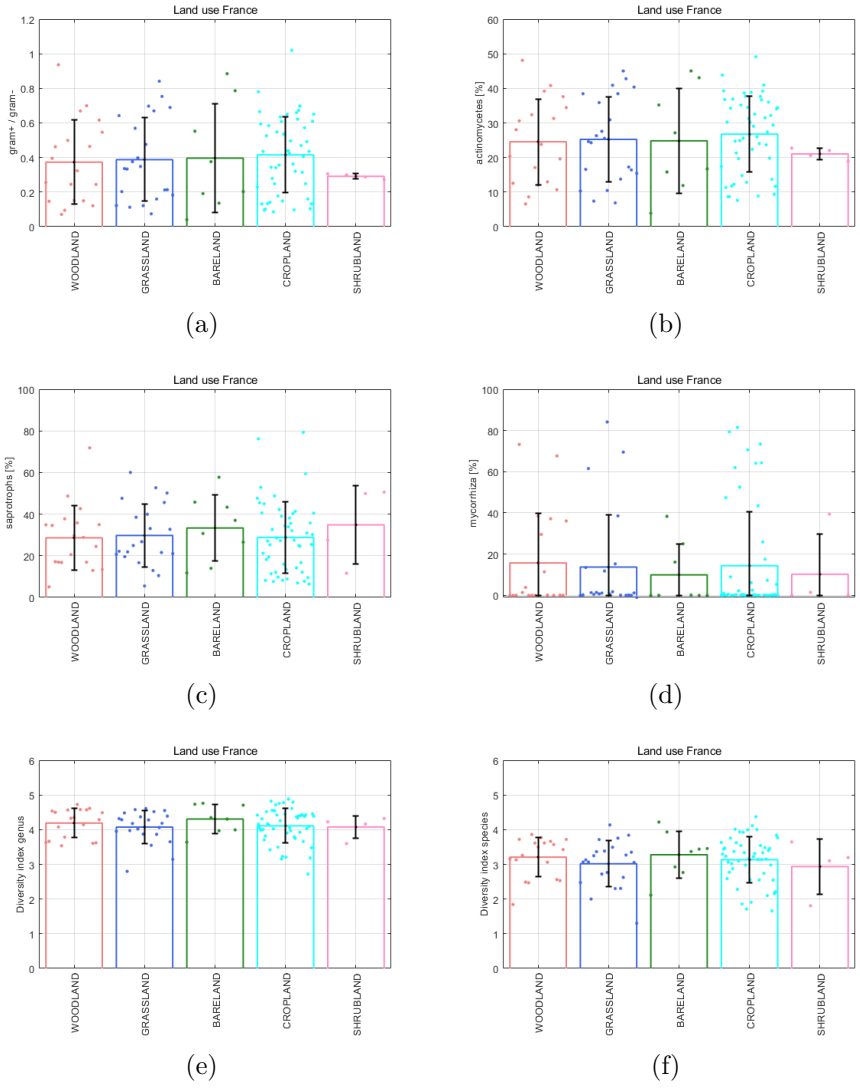

Figure 4.3: Visualisation of how the ratio of gram+ and gram− bacteria and the presence of actinomycetes, saprotrophs and mycorrhiza depend on the land use in France. The last two graphs show the diversity index calculated for genus and species, respectively. The indicators without shown unit are unitless. The data points are grouped into the land types woodland, grassland, bareland, cropland, and shrubland. The bars represent the mean values and the black errorbars are the standard deviation of each data group. (a) Variation in ratio of gram+ and gram− bacteria. (b) Variation in presence of actinomycetes bacteria. (c) Variation in presence of saprotrophic fungi. (d) Variation in presence of mycorrhiza fungi. (e) Variation in diversity index based on genus. (f) Variation in diversity index based on species.

4.2.2 Land use analysis in Spain

Figure 4.4 shows the values for the biological soil indicators for data from Spain. The data points are grouped after land use.

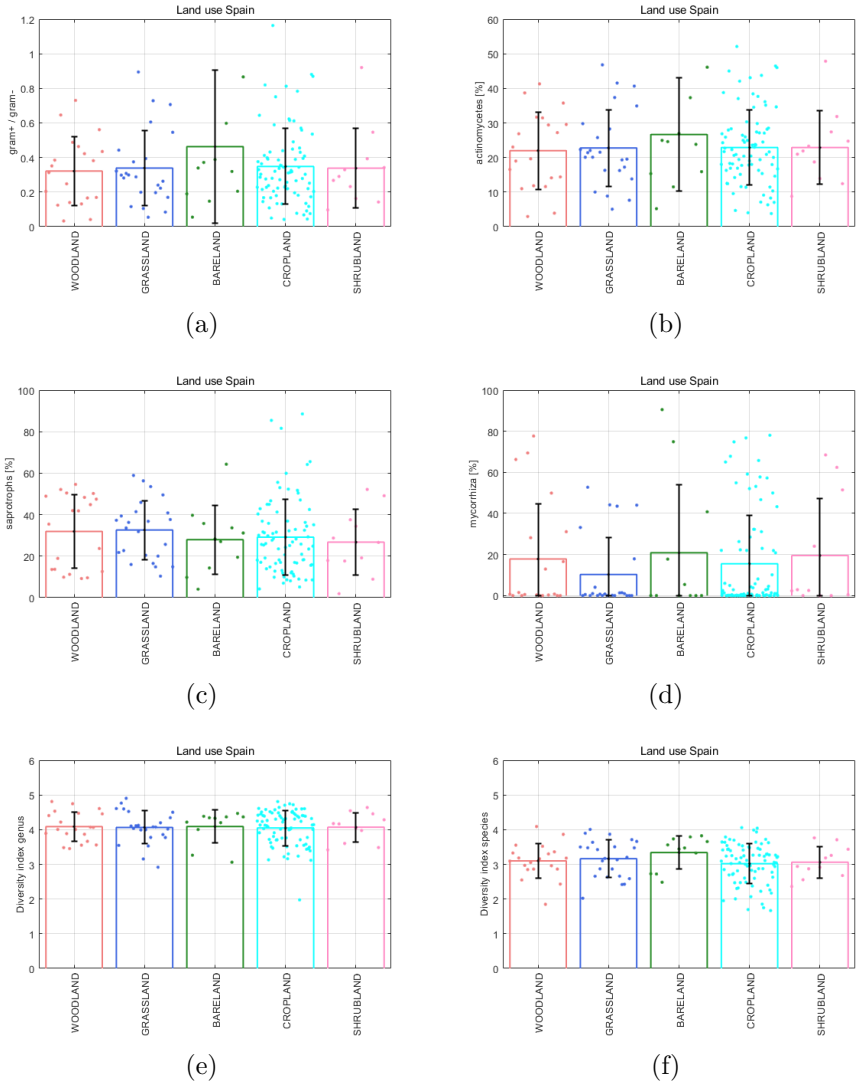

Figure 4.4: Visualisation of how the ratio of gram+ and gram− bacteria and the presence of actinomycetes, saprotrophs and mycorrhiza depend on the land use in Spain. The last two graphs show the diversity index calculated for genus and species, respectively. The indicators without shown unit are unitless. The data points are grouped into the land types woodland, grassland, bareland, cropland, and shrubland. The bars represent the mean values and the black errorbars are the standard deviation of each data group. (a) Variation in ratio of gram+ and gram− bacteria. (b) Variation in presence of actinomycetes bacteria. (c) Variation in presence of saprotrophic fungi. (d) Variation in presence of mycorrhiza fungi. (e) Variation in diversity index based on genus. (f) Variation in diversity index based on species.

4.3 Correlation analysis

For the correlation analysis, the countries France and Spain were analysed. For each biological soil parameter, correlation with the characteristic soil indicators (see Table 3.2) were visualised and the R^2 values for a linear regression were calculated.

4.3.1 Correlation analysis for France

In Figures 4.5 - 4.10, the correlation or correlations with highest R^2 between each biological soil indicator and the characteristic soil indicators (see Table 3.2) measured in France are shown (see Appendix B for the rest of the correlation graphs). The specific biological soil indicator is on the y axis, and the characteristic soil indicator is on the x axis. The R^2 values are shown in the upper right corner of each graph. In Table 4.3, all R^2 values for the linear correlations between each biological soil indicator and the characteristic soil indicators are compiled.

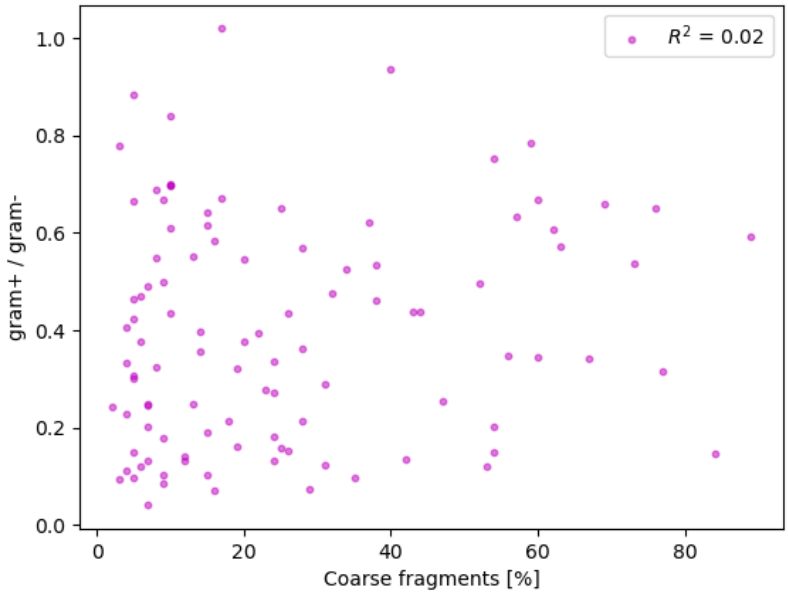

Figure 4.5: Visualisation of the strongest linear correlation between the biological soil indicator of the ratio of gram+ and gram− bacteria and the characteristic soil indicators showed in Table 3.2. All the data is related to France. The R^2 value for the linear correlation between the ratio of gram+ and gram− bacteria and the characteristic soil indicator of coarse fragment is 0.02.

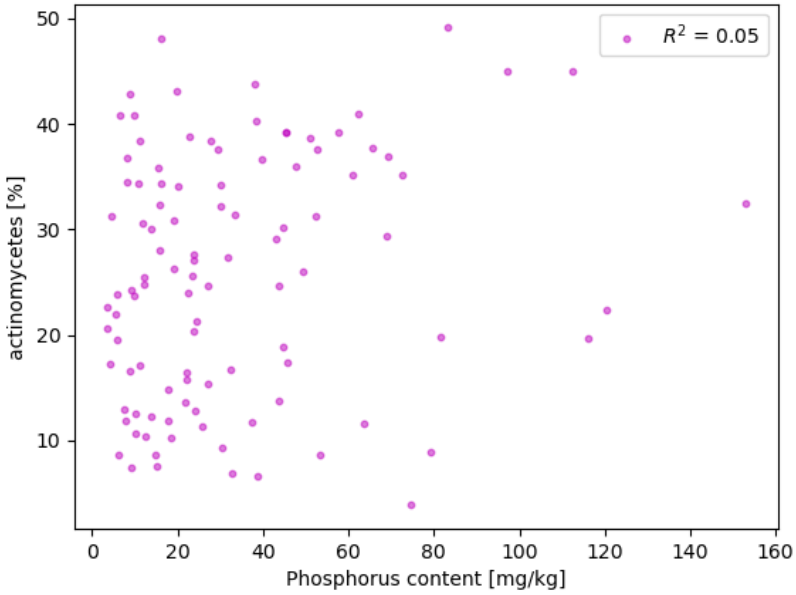

Figure 4.6: Visualisation of the strongest linear correlation between the biological soil indicator of the presence of actinomycetes bacteria and the characteristic soil indicators showed in Table 3.2. All the data is related to France. The R^2 value for the linear correlation between the presence of actinomycetes bacteria and the characteristic soil indicator of phosphorus content is 0.05.

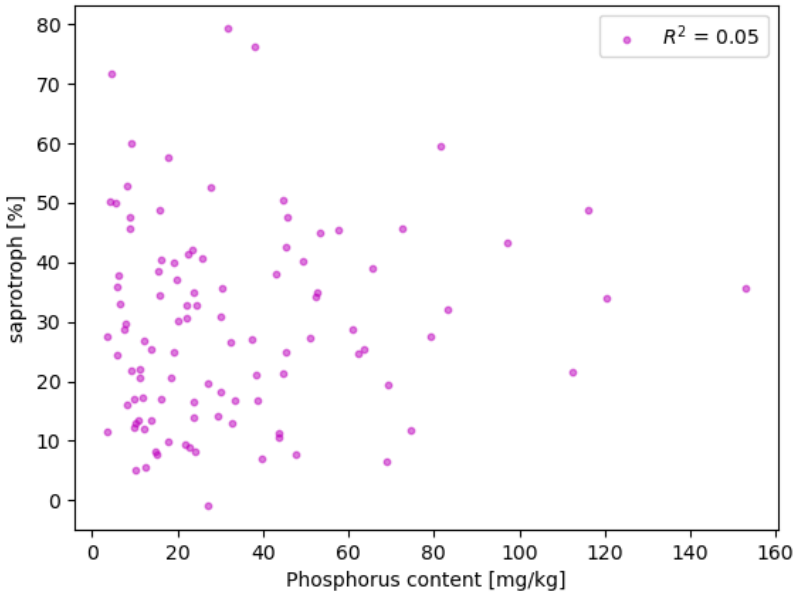

Figure 4.7: Visualisation of the strongest linear correlation between the biological soil indicator of the presence of saprotrophs and the characteristic soil indicators showed in Table 3.2. All the data is related to France. The R^2 value for the linear correlation between the presence of saprotrophs and the characteristic soil indicator of phosphorus content is 0.05.

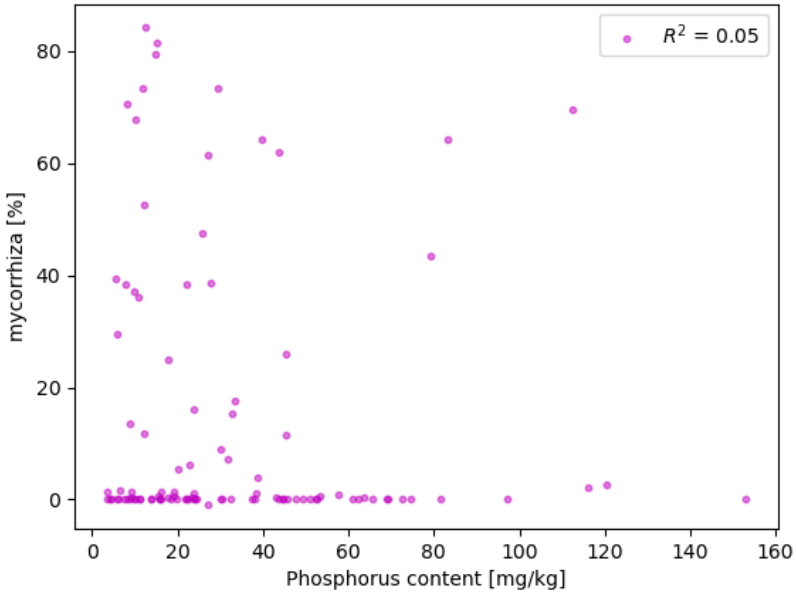

Figure 4.8: Visualisation of the strongest linear correlation between the biological soil indicator of the presence of mycorrhiza and the characteristic soil indicator showed in Table 3.2. All the data is related to France. The R^2 value for the linear correlation between the presence of mycorrhiza and the characteristic soil indicator of phosphorus content is 0.05.

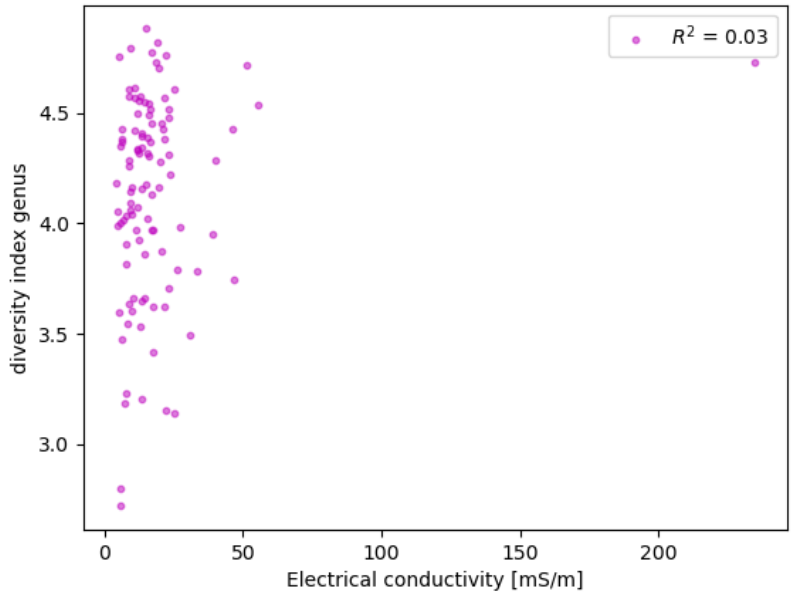

Figure 4.9: Visualisation of the strongest linear correlation between the biological soil indicator the diversity index calculated for genus and the characteristic soil indicators showed in Table 3.2. All the data is related to France. The R^2 value for the linear correlation between the the diversity index calculated for genus and the characteristic soil indicator of electrical conductivity is 0.03.

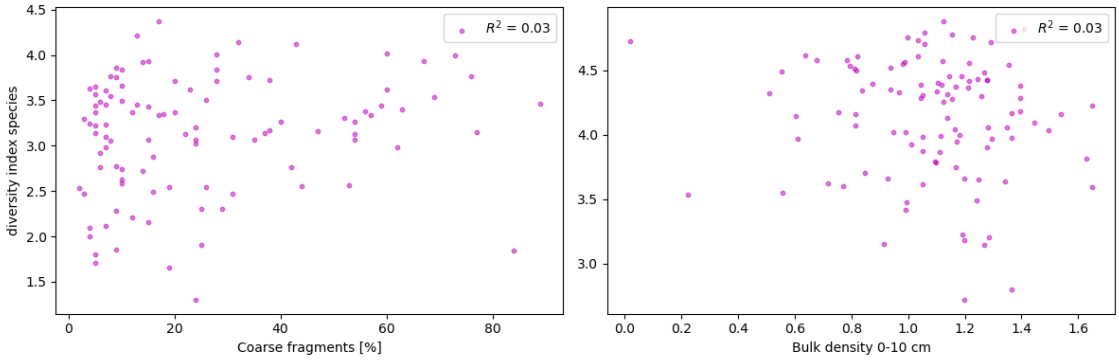

Figure 4.10: Visualisation of the strongest linear correlations between the biological soil indicator the diversity index calculated for species and the characteristic soil indicators showed in Table 3.2. All the data is related to France. The R^2 values for the linear correlation between the diversity index calculated for species and the characteristic soil indicators of coarse fragments and bulk density 0-10 cm are 0.03.

Table 4.3: Compilation of the R^2 values for the linear correlations for the data related to France.

	gram+/gram−	Actinomycetes	Saprotrophs	Mycorrhiza
Coarse fragments	0.02	0.02	0.02	0.02
Clay content	0.0	0.0	0.0	0.0
Sand content	0.0	0.0	0.0	0.0
Silt content	0.0	0.0	0.0	0.0
pH $CaCl_2$	0.0	0.0	0.0	0.0
pH H_2O	0.0	0.0	0.0	0.0
Electrical conductivity	0.0	0.0	0.0	0.0
Organic carbon	0.01	0.01	0.05	0.01
Carbonate content	0.01	0.01	0.0	0.01
Phosphorus content	0.05	0.05	0.05	0.05
Total nitrogen	0.0	0.0	0.01	0.0
Extractable potassium content	0.01	0.01	0.01	0.01
Bulk density (0-10 cm)	0.0	0.0	0.0	0.0
Bulk density (10-20 cm)	0.01	0.01	0.01	0.01

Table 4.3: Continued

	Diversity index genus	Diversity index species
Coarse fragments	0.01	0.03
Clay content	0.0	0.01
Sand content	0.0	0.0
Silt content	0.0	0.0
pH $CaCl_2$	0.02	0.01
pH H_2O	0.02	0.01
Electrical conductivity	0.03	0.02
Organic carbon	0.01	0.0
Carbonate content	0.0	0.0
Phosphorus content	0.01	0.01
Total nitrogen	0.01	0.0
Extractable potassium content	0.0	0.0
Bulk density (0-10 cm)	0.01	0.03
Bulk density (10-20 cm)	0.01	0.01

4.3.2 Correlation analysis for Spain

In Figures 4.11 - 4.16, the correlations between each biological soil indicator and the characteristic soil indicators (see Table 3.2) measured in Spain are shown (see Appendix B for the rest of the correlation graphs). The specific biological soil indicator is on the y axis, and the characteristic soil indicator is on the x axis. The R^2 values are shown in the upper right corner of each graph.

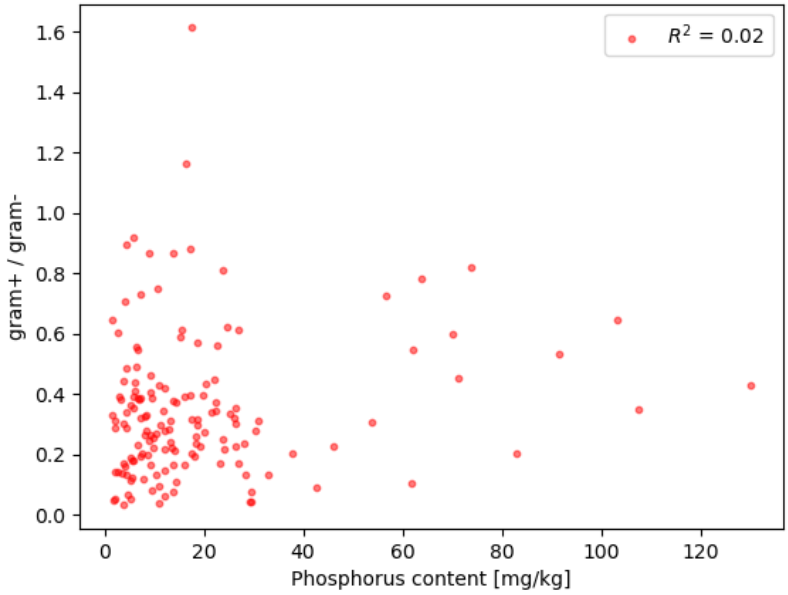

Figure 4.11: Visualisation of the strongest linear correlation between the biological soil indicator of the ratio of gram+ and gram− bacteria and the characteristic soil inicators showed in Table 3.2. All the data is related to Spain. The R^2 value for the linear correlation between the ratio of gram+ and gram− bacteria and the characteristic soil indicator of phosphorus content is 0.02.

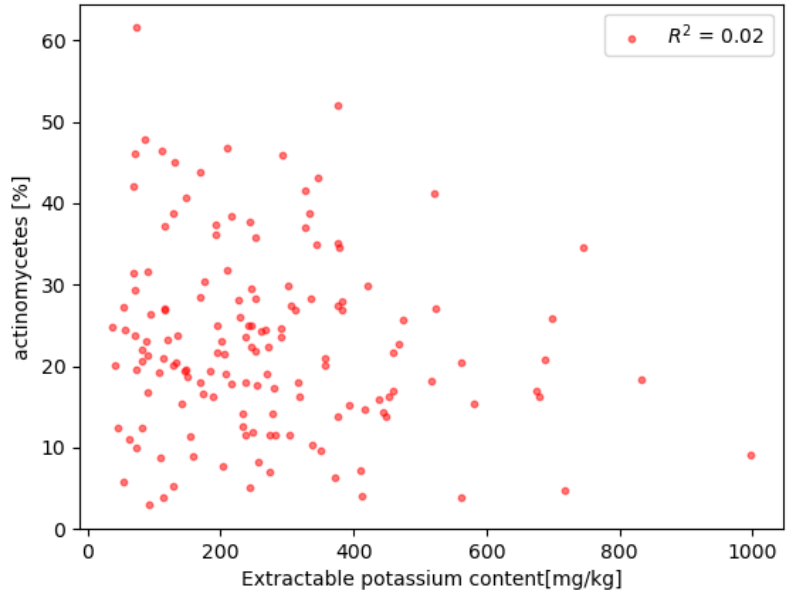

Figure 4.12: Visualisation of the strongest linear correlation between the biological soil indicator of the presence of actinomycetes bacteria and the characteristic soil indicators showed in Table 3.2. All the data is related to Spain. The R^2 value for the linear correlation between the presence of actinomycetes bacteria and the characteristic soil indicator of extractable potassium content is 0.02.

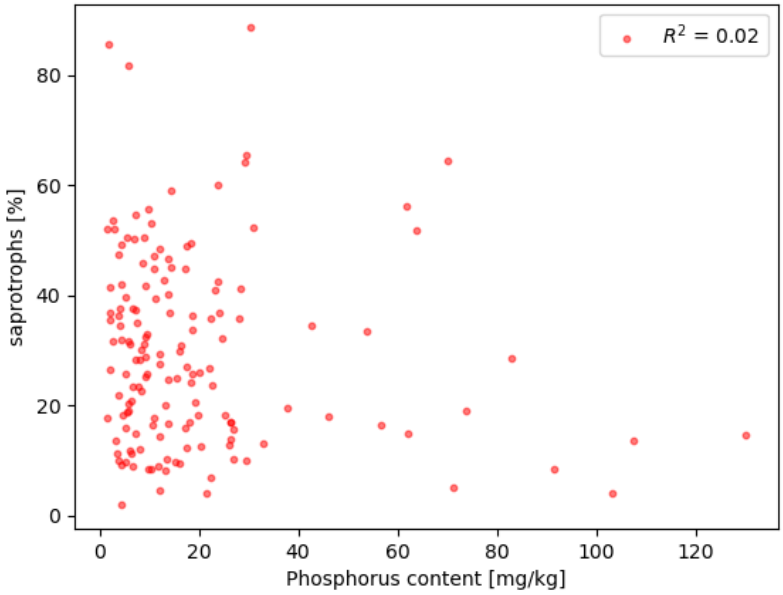

Figure 4.13: Visualisation of the strongest linear correlation between the biological soil indicator of the presence of saprotrophs and the characteristic soil indicators showed in Table 3.2. All the data is related to Spain. The R^2 value for the linear correlation between the presence of saprotrophs and the characteristic soil indicator of phosphorus content is 0.02.

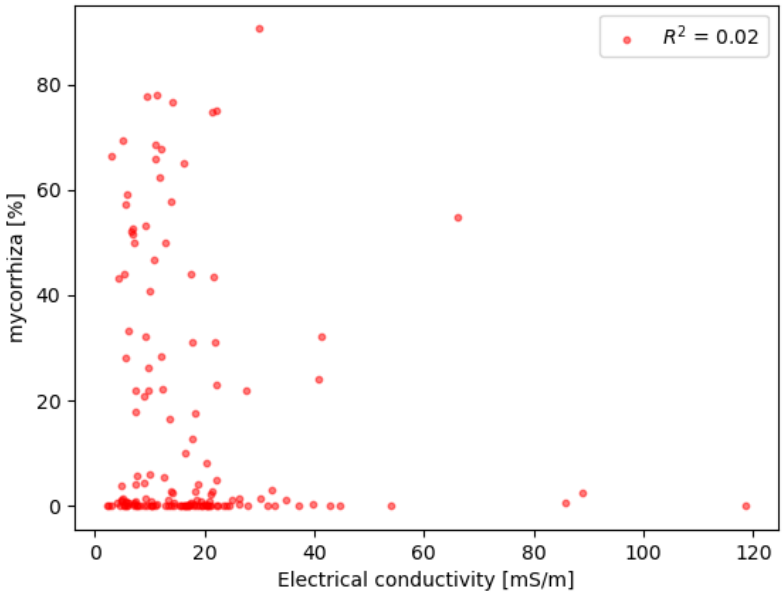

Figure 4.14: Visualisation of the strongest linear correlation between the biological soil indicator of the presence of mycorrhiza and the characteristic soil indicators showed in Table 3.2. All the data is related to Spain. The R^2 value for the linear correlation between the presence of mycorrhiza and the characteristic soil indicator of electrical conductivity is 0.02.

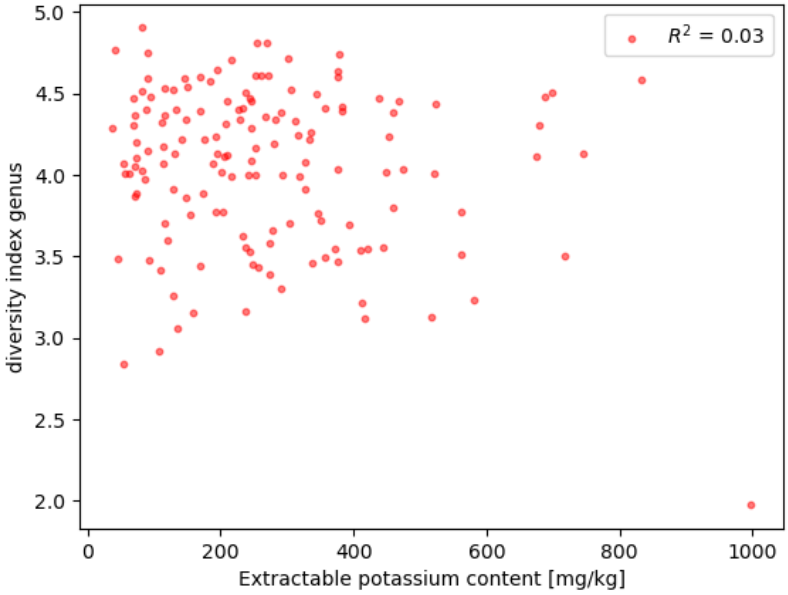

Figure 4.15: Visualisation of the strongest linear correlation between the biological soil indicator the diversity index calculated for genus and the characteristic soil indicators showed in Table 3.2. All the data is related to Spain. The R^2 value for the linear correlation between the the diversity index calculated for genus and the characteristic soil indicator of extractable potassium content is 0.03.

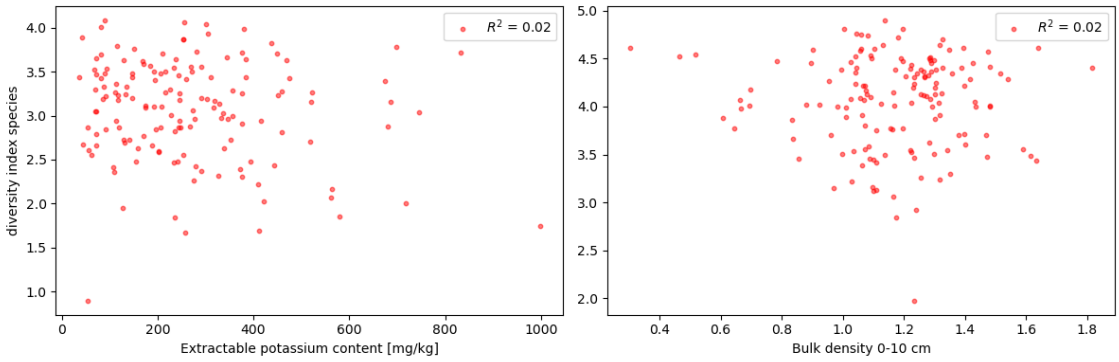

Figure 4.16: Visualisation of the strongest linear correlations between the biological soil indicator the diversity index calculated for species and the characteristic soil indicators showed in Table 3.2. All the data is related to Spain. The R^2 values for the linear correlation between the the diversity index calculated for species and the characteristic soil indicators of extractable potassium content and bulk density 0-10 cm are 0.02.

Table 4.4: Compilation of the R^2 values for the linear correlations for the data related to Spain.

	gram+/gram−	Actinomycetes	Saprotrophs	Mycorrhiza
Coarse fragments	0.0	0.01	0.01	0.0
Clay content	0.01	0.01	0.01	0.0
Sand content	0.0	0.0	0.0	0.0
Silt content	0.0	0.0	0.01	0.0
pH $CaCl_2$	0.0	0.0	0.0	0.0
pH H_2O	0.0	0.0	0.0	0.0
Electrical conductivity	0.01	0.0	0.0	0.02
Organic carbon	0.0	0.0	0.0	0.0
Carbonate content	0.0	0.0	0.0	0.0
Phosphorus content	0.02	0.01	0.02	0.01
Total nitrogen	0.0	0.0	0.0	0.0
Extractable potassium content	0.01	0.02	0.0	0.0
Bulk density (0-10 cm)	0.0	0.0	0.0	0.0
Bulk density (10-20 cm)	0.0	0.0	0.0	0.0

Table 4.4: Continued

	Diversity index genus	Diversity index species
Coarse fragments	0.0	0.0
Clay content	0.0	0.0
Sand content	0.0	0.0
Silt content	0.0	0.0
pH $CaCl_2$	0.0	0.0
pH H_2O	0.0	0.0
Electrical conductivity	0.01	0.0
Organic carbon	0.01	0.01
Carbonate content	0.0	0.0
Phosphorus content	0.0	0.0
Total nitrogen	0.01	0.01
Extractable potassium content	0.03	0.02
Bulk density (0-10 cm)	0.0	0.02
Bulk density (10-20 cm)	0.0	0.0

4.4 Relation between actinomycetes and the ratio of gram+ and gram− bacteria

In Figure 4.17, the relation between the ratio of gram+ and gram− and the presence of actinomycetes bacteria is shown. The graph follows Eq. 3.13 and is plotted as a black dashed line.

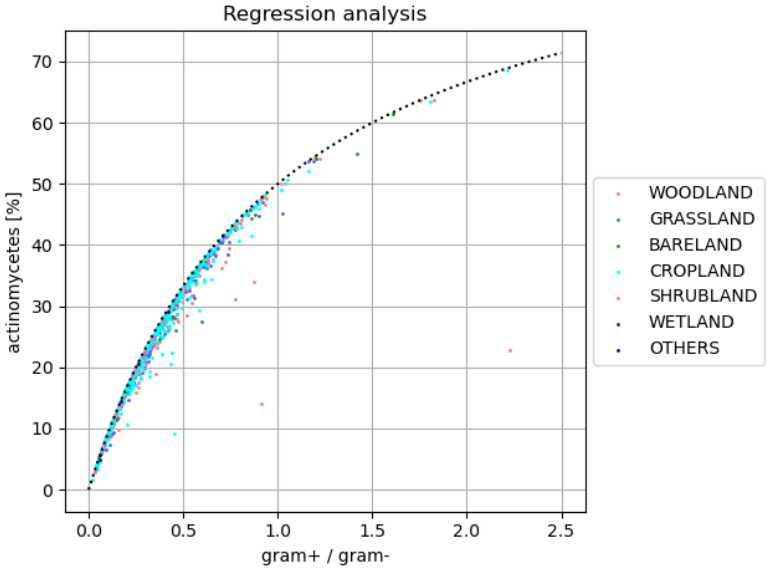

Figure 4.17: Relation between the ratio of gram+ and gram− bacteria and the presence of actinomycetes bacteria. The graph follows the equation $y = \frac{x}{x+1}$, which are shown with a black dashed line.

4.4.1 Correlations between biological soil indicators

In Figure 4.18, some correlations between one biological soil indicator related to fungi and another biological soil indicator related to bacteria are shown (see Appendix A for more graphs).

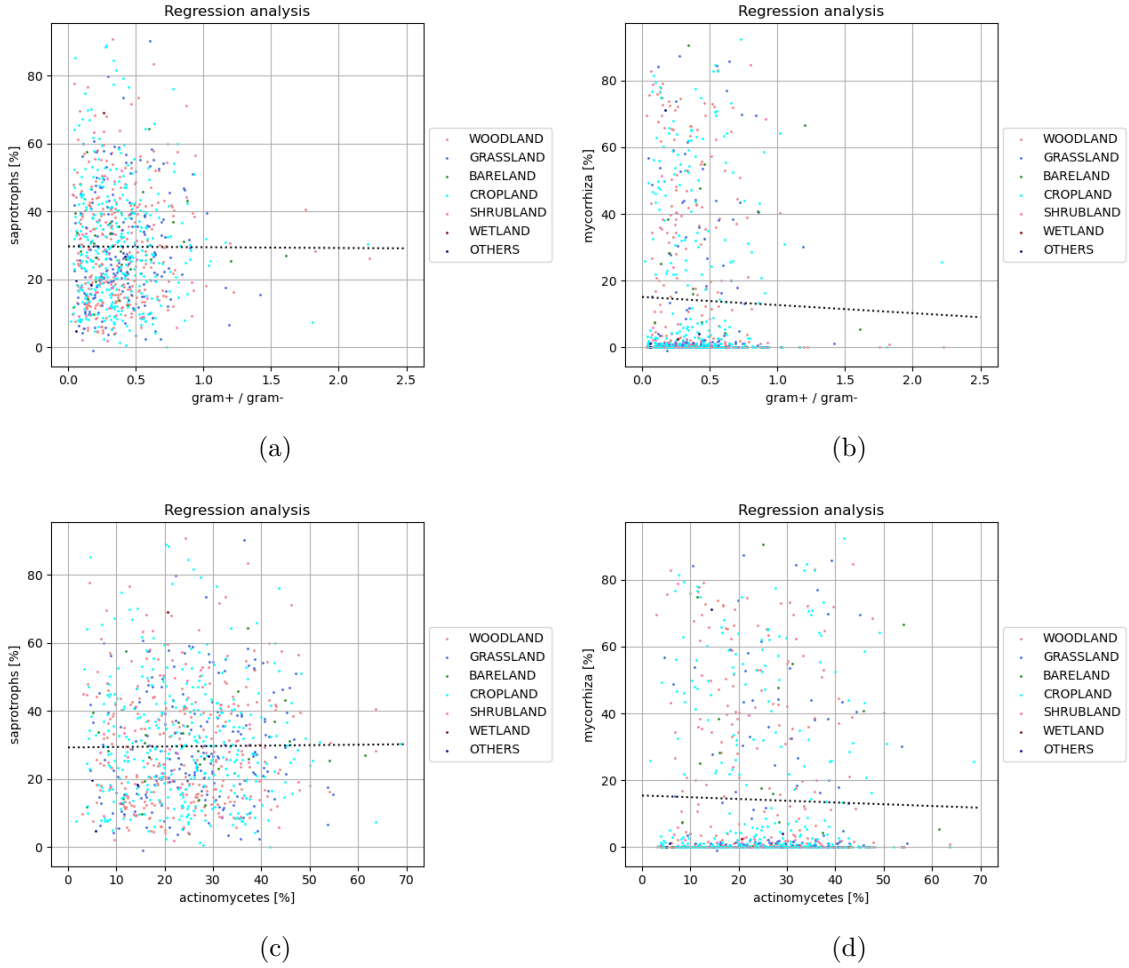

Figure 4.18: Visualisation of some correlations between one biological soil indicators related to fungi and another biological soil indicator related to bacteria. (a) Correlation between the ratio of gram+ and gram− bacteria and the presence of saprotrophs. (b) Correlation between the ratio of gram+ and gram− bacteria and the presence of mycorrhiza. (c) Correlation between the presence of actinomycetes bacteria and the presence of saprotrophs. (d) Correlation between the presence of actinomycetes bacteria and the presence of mycorrhiza.

4.5 Principal component analysis

In the following section, the results of the principal component analysis are shown. For each biological soil indicator, the eigenvalues (λ_i), standard deviations (σ_i) and the first value in the eigenvector are shown. Only the first value of the vector is shown because that value describes the connection of the biological soil indicator and the correlation. If that absolute value is low, the correlation is related to other characteristic soil parameters in the analysis. The order of the characteristic soil indicators from LUCAS is shown in Table 4.5.

Table 4.5: The order of the characteristic soil indicators from LUCAS that are included in the PCA.

i	Soil indicator
2	Coarse fragments
3	Clay content
4	Sand content
5	Silt content
6	pH_{CaCl_2}
7	Electrical conductivity
8	Organic carbon
9	Carbonate content
10	Phosphorus content
11	Total nitrogen
12	Extractable potassium content
13	Bulk density (0-10 cm)
14	Bulk density (10-120 cm)

4.5.1 Principal component analysis for France

Tables 4.6-4.7 show the results of the principal component analyses for data from France. Table 4.6 shows the linear combinations with strongest correlations and Table 4.7 shows the linear combinations where the specific biological soil indicator has strongest impact.

Table 4.6: Compilation of the linear combinations with strongest correlations (highest absolute value of the standard deviation σ and corresponding eigenvalue λ and eigenvector v). The data is from France.

Biological soil indicator	λ	σ	v
gram+/gram−	3.40	1.84	[-0.07 ...]
Actinomycetes	3.40	1.65	[-0.07 ...]
Saprotrophs	3.39	1.84	[-0.02 ...]
Mycorrhiza	3.39	1.84	[0.03 ...]
Diversity index species	3.40	1.84	[0.05 ...]

Table 4.7: Compilation of the linear combinations where the biological soil indicator has the strongest impact (highest absolute value of the first element in the eigenvector v). The data is from France.

Biological soil indicator	λ	σ	v
gram+/gram−	0.71	0.84	[0.61 ...]
Actinomycetes	1.11	1.05	[-0.59 ...]
Saprotrophs	0.87	0.93	[-0.74 ...]
Mycorrhiza	0.82	0.91	[-0.64 ...]
Diversity index species	1.11	1.05	[-0.52 ...]

4.5.2 Principal component analysis for Spain

Tables 4.8-4.9 show the results of the principal component analyses for data from Spain. Table 4.8 shows the linear combinations with strongest correlations and Table 4.9 shows the linear combinations where the specific biological soil indicator has strongest impact.

Table 4.8: Compilation of the linear combinations with strongest correlations (highest absolute value of the standard deviation σ and corresponding eigenvalue λ and eigenvector v). The data is from Spain.

Biological soil indicator	λ	σ	v
gram+/gram−	3.86	1.96	[-0.02 ...]
Actinomycetes	3.85	1.96	[0.03 ...]
Saprotrophs	3.39	1.84	[-0.00 ...]
Mycorrhiza	3.86	1.96	[0.01 ...]
Diversity index species	3.87	1.97	[-0.07 ...]

Table 4.9: Compilation of the linear combinations where the biological soil indicator has the strongest impact (highest absolute value of the first element in the eigenvector v). The data is from Spain.

Biological soil indicator	λ	σ	v
gram+/gram−	1.12	1.06	[-0.75 ...]
Actinomycetes	1.08	1.04	[0.67 ...]
Saprotrophs	1.16	1.17	[-0.79 ...]
Mycorrhiza	0.11	1.07	[-0.74 ...]
Diversity index species	1.09	1.04	[-0.71 ...]

5
Analysis

In this section, the results from Section 4 is analysed.

5.1 Geographical location

For all biological soil indicators, the mean values are almost similar for all geographical regions. Even if the mean values do not have a large difference, there are both a large spread and large standard deviations of the data points. In Figure 4.1a, the ratio of gram+ and gram− bacteria is shown. According to the graph, all the regions have a mean value below 0.5 and the general values of the ratio are below 1. Data also show that the presence of mycorrhiza has a large spread and a large proportion of zero values (see Figure 4.1d). The last rows in Table 4.1, show the results of the hypothesis test (see Section 3.7.1). The results indicate that the population mean values are equal for the different regions to 99% confidence for all biological soil indicators.

5.2 Land use

For all biological soil indicators, the mean values are almost similar for all type of land uses. Even if the mean values do not have a large difference, there are both a large spread and large standard deviations of the data points. Even here, the noticeable results are that the ratio of gram+ and gram− bacteria in general is lower than 1 (see Figure 4.2a) and that there is a large proportion of zero values for the presence of mycorrhiza (see Figure 4.2d). The last rows in Table 4.2, show the results of the hypothesis test (see Section 3.7.1). The results show that the population mean values are equal for the different land uses to 99% confidence for all biological soil indicators.

Even the distributions of the data points for France and Spain separately (see Figure 4.3 and Figure 4.4) show no clear correlation between land use and the indicator value.

5.3 Correlation analysis

The visualised distributions of the data points in Section 4.3, do not indicate strong linear correlations between the soil indicators. Since the largest R^2 values are 0.05

for data related to France and 0.03 for data related to Spain, the analysis indicate that there are no linear correlations between the biological soil indicators and the characteristic soil indicators in the Lucas for these data sets.

5.4 Relation between biological soil indicators

In Figure 4.17, the correlation between the ratio of gram+ and gram− and the presence of actinomycets follows the line $y = \frac{x}{x+1}$. Data points located at the line mean that the presence of firmicutes is zero. Data points located further below the line indicates a higher presence of firmicutes. Figure 4.17, indicates that the presence of firmicutes in Europe in general is low.

The regression analysis that are done in Figure 4.18 indicate that there may be no linear correlations between that biological soil indicators. Even the correlation graphs in shown in Appendix A with similar types of analyses between bacteria and fungi result in no strong linear correlations.

5.5 Principal component analysis

From the results in Section 4.5, the analysis is similar for both France and Spain. For all biological soil indicators, there are no σ values that are signifiant larger than the other. For the largest σ for all biological soil indicators, the participation of the biological soil indicator in the correlation is low. This means that with the data from France and Spain, there are no strong linear correlations between none of the biological soil indicators and the characteristic soil indicators.

5.6 Overall analysis

From both analyses in two dimensions and multi dimensions, the results in Section 4 indicate that there are no strong linear correlations between the biological soil indicators and the characteristic soil indicators. The results neither shows strong linear correlations between the different biological soil indicators.

6
Discussion

This chapter discusses the results of the data analysis and compares it with previous research. Also, limitations of the study, ethical aspects of the project and potential future research are discussed.

6.1 Discussion of results

In the following section, the results from Chapter 4 are discussed and compared with previous research.

6.1.1 Correlations

The linear regression analyses (see Section 4.3), indicate that the chosen biological soil indicators may not have linear correlations with the characteristic soil indicators in the dataset. The maximum R^2 values are $0.05 << 1$ and $0.03 << 1$ for France and Spain, respectively. Even the results of the PCA (see Section 4.5), indicate that there may be no strong linear correlations between any of the biological soil indicators and the characteristic soil indicators. The linear combinations with possible correlations contain a low contribution of the biological soil indicators, and linear combinations with a high contribution of the biological soil indicators do not indicate any linear correlations. In the regression analysis and the PCA, the data is grouped according to national borders and only data from France and Spain is used. The purpose of this data division was to reduce differences in, for example, vegetation and climate. Even if some potential differences are reduced, there is still a potential for differences within individual countries. A future alternative to grouping according national borders could be to group according to, for example, vegetation or precipitation.

6.1.2 Presence of mycorrhiza

For example, Figures 4.1d, 4.2d, and 4.8 show that there are many samples with a low or zero proportion of mycorrhiza. The high number of zero values for mycorrhiza makes it difficult to find correlations with characteristic soil indicators. Due to the symbiosis between mycorrhiza and plants, the presence of mycorrhiza depends on the available plants. As described in Section 2.5, measures such as organic agriculture, crop variation, and reduction of chemical pesticides, tillage, and addition of phosphorus are beneficial for mycorrhiza growth. According to EUROSTAT, the use of chemical pesticides and fertilizers based on nitrogen and phosphorus has a

descending trend in the EU, but they are still used. The sale of pesticides in the EU was reduced to 29 200 tonnes in 2023, but the countries that are the main producers in the area of agriculture (France, Spain, Germany and Italy) represent the majority of the sales [25]. The use of fertilizers in the EU decreased with 10.3% between 2021 and 2022, but also here the main agricultural producers represented approximately half of the use of phosphorus-based fertilizers [28]. The number of organic farms in the EU increases. In 2020, 9.1 million farms had partial organic production and 245 000 farms were fully organic [26]. Even if the number of fully organic farms doubled during the 2010-2020 period, the fraction of farms with organic work was only 2.7%. By implementing more types of actions that promote the growth of mycorrhiza, the potential for an improved presence of mycorrhiza in European soils increases. Since a healthy soil need high values of both mycorrhiza and saprotrophs, this may lead to an optimization problem. The mission is to find the optimal proportion of mycorrhiza and saprotrophs to support soil health. The potential demands and target of this type of optimization problem can be customized depending on the area of usage of the soil. According to previous research, the relation between saprotrophic fungi and mycorrhiza is complex and difficult to study. In general, soil is divided into different layers with different characteristics [4]. Due to the demands from the fungi, even the fungi are located in different layers. The upper layer is richer in plant litter, which is necessary for saprotrophic fungi. The humus layer with well-composed organic matter is located underneath. Here, mycorrhiza is located to a greater extent due to their opportunity to get carbon from the symbiosis with plants. Under the humus layer, the mineral soil is found. The fact that saprotrophic fungi is dominant in the upper layer can be a reason for the low values of mycorrhiza in the analysed data. As said before, there are uncertainties of the relation between saprotrophic fungi and mycorrhiza but in most cases it was discovered that they have an antagonistic interaction [58]. The sectioning of the different fungi is a consequence of competition of both nutrition and space. It is also important to point out that there are studies that show opposite results and more research is needed in this research area.

6.1.3 Ratio of gram+ and gram− bacteria

In Figure 4.1a and Figure 4.2a, the ratio of gram+ and gram− bacteria is shown. Something interesting with these graphs is that the mean value is significantly lower than unity. Good soil health is according to Eurofins [22], characterized by a ratio close to 1 or just above 1. Values below 1 were expected since the soils in the data are largely cultivated or used by humans. As described in Section 2.5, stresses such as tilling and pesticides inhibit gram+ bacteria and therefore reduce the ratio. The interesting thing is that the values on average are below 0.5, which is much lower than 1. The results show clearly that actions are required to increase the ratio and to get healthier soils in Europe.

6.1.4 Relation between biological soil indicators

Figure 4.17, shows the relation between the ratio of gram+ and gram− bacteria and the presence of actinomycetes. A correlation between these indicators was expected

because of the underlying calculation (see Eq. 3.2, Eq. 3.1, and Section 3.8). As seen in the equations, the number of ASVs related to the phylum *actinobacteriota* is a variable in the equations for both biological soil indicators. Knowledge about the relation simplifies the data analysis and makes it more time efficient. The calculation of the ratio is more time consuming because the eDNA data for bacteria needs to be sorted into gram+ and gram− bacteria. Instead, by only sorting out the ASVs related to actinomycetes bacteria and calculating the ratio based on that, the Python script becomes simplified. It is also important to note that the data from LUCAS follows the graph $y = \frac{x}{x+1}$, which indicates that the presence of firmicutes bacteria in general is close to zero in Europe.

6.1.5 Connection to soil health

As described in Section 2.5, the concept of soil health focuses on a general assessment of soils. To make a fair and realistic assessment, a good understanding of how soil can be described is advantageous. The results of this analysis indicate no strong linear correlations between the analysed soil indicators. Even if this study only investigates a small part of the concept of soil health, the result in turn shows the complexity of the soil system and that more research is needed about the system to improve understanding.

6.1.6 Correlation between bacteria and fungi

Another aspect discovered during the analysis of correlations is the relation between biological soil indicators based on bacteria and biological soil indicators based on fungi. The indicators based on bacteria are the ratio of gram+ and gram− bacteria and the presence of actinomycetes. The indicators based on fungi are the presence of saprotrophs and the presence of mycorrhiza. For the scatter plots with an indicator based on bacteria at one axis and an indicator based on fungi at the other axis, there are no strong linear correlations (see Figure 4.18 and Appendix A). This means that the concept of soil health may be divided into two separated parts. One part is only connected to bacteria and one part is only connected to fungi. The insight that there may be no connection between the indicators based on bacteria and the indicators based on fungi may be important in the development of strategies to improve soil health. Without a correlation, actions to improve the bacteria indicators may be implemented without the risk of disturbing the fungi, and vice versa. If they are connected, more caution may be needed during implementation. It is important to have in mind that this hypothesis is based on one data set and one analysis. Therefore, more research is needed to obtain more knowledge and understanding.

6.2 Limitations of the study

There are some limitations in the study that increase the risk of inaccuracies in the result. For example, potential errors during the eDNA analysis are not taken into account. Examples of this are contamination during sample collection, mistakes during the laboratory analysis, or sequencing errors. One sequencing error during

the eDNA analysis, results in another taxonomic classification. A more thorough analysis should include investigations about how this and the confidence interval for the data affect the results.

In the analysis of land use for the whole dataset, the data points were grouped according to land use. Compared to the groups in the geographical analysis, the difference in the number of data points per group is larger for the land use analysis. Even if some groups were removed from the analysis (because of fewer than 5 data points), there are still groups with few data points (e.g. bareland and shrubland). This leads to less reliable values for the mean and standard deviation. There were also much fewer data points per dataset in the analyses for France and Spain separately.

During the classification of fungi, the FungalTraits database was used [59]. The classification was done by looping through the genus names in the data with sequence classifications and checking if the same name was found in the database. If the name was found and the primary lifestyle was related to the classification of fungi of interest (saprotrophs or mycorrhiza), the DNA detection was classified as that type of fungi. If no match was found or if the primary lifestyle was not related to the group of fungi in interest, the DNA sequence was not classified in that group. The classification method should be more thorough if the detections of DNA with a resolution less than *Genus* had been compared with another taxonomic rank in the database. Even the usage of the information on the secondary lifestyle during the classification should give a more thorough result.

An important factor according to the data is the large standard deviation. There are large standard deviations and big differences in indicator values for different samples. From the analysis done in this report, it is difficult to draw conclusions about the causes of good soil health according to the spread. The large spread is also a confirmation for the complexity of the soil system. There are uncertainties about how the system works and how it reacts to changes. Because of the complexity of the soil system, the method of analysing not all indicators is a simplification.

6.3 Ethical aspects of the project

An advantage of analysing eDNA is that sample collection does not disturb the environment or the organisms that live there. Only genetic traces of organisms are collected and analysed. Another aspect is that more research on eDNA and soil health can lead to a higher availability of the method. With more knowledge as well as analysis methods that are easier to use and cheaper, more eDNA data can be collected and analysed. By including soil samples for more different land uses, climate zones, and usage, the research can be more general, inclusive, and realistic.

The work according to eDNA includes data storage. In this report, data from bacteria and fungi have been used and analysed, which are not related to personal

information. The measuring instruments used during sample collection are not sterile and, therefore, there is a risk for human DNA collection. These markers are not of interest for these types of data analysis and it is difficult to connect the DNA to a specific person. Despite that, it is important to discuss the consequences related to collection of human DNA. Since the implementation of GDPR, the storage of personal data has been more regulated. There are still ambiguities about the type of data that will be classified as personal data. If human eDNA are collected and genetic information is classified as personal data, this can be regulated according to GDPR. Due to the relative new usage of eDNA on a larger scale, it is necessary to review the regulation on the storage of eDNA data.

It is also necessary to discuss the obligations related to DNA detections from eDNA. Examples are if some kind of pathogen is detected in the eDNA or some rare or threatened organism is detected. Should the responsible person be obliged to release that information or not? Detections of pathogens can, for example, entail a risk of a prohibition to use the area. If the area is used as an income for the owner, this type of information can be devastating. There is also a risk of not being informed about detected pathogens because of the potential spread and long-term consequences. Detections of rare or threatened species can also affect the permission to use the area. It can both be important to inform about the organism to increase the possibility of protection and also a risk that the area becomes exposed to sabotage or an increased interest in visits that harm the organism.

6.4 Future research

The data analysis methods and knowledge from this report can be used to increase understanding and indication of soil health and discuss possible strategies to improve the soil health. By adding data from additional data samples, conclusions about the current state for each biological soil indicator can be calculated and evaluated. By comparing the values with the LUCAS data, conclusions can be drawn about whether or not they are expected according to the average in the EU.

The correlation graphs indicate that the biological soil indicators based on bacteria and the biological soil indicators based on fungi, based on this dataset, may not relate to each other. This might be an important aspect in the development of strategies to improve soil health. If more similar analyses are implemented for other datasets, this hypothesis can be investigated further and confirmed or not.

7
Conclusion

The regression analyses and the PCA that have been performed in this report with the dataset from LUCAS 2018 [29] and the described boundaries and limitations, indicate that there may be no linear correlations between the chosen biological soil indicators and the characteristic soil indicators shown in Table 4.5. Other interesting insights from this analysis are that the ratio of gram+ and gram− bacteria is on average much lower than 1 in Europe. It has also been found that relatively many soils have low or 0% mycorrhiza. The reason for finding low presence of mycorrhiza might however be due to that the samples were taken from top soil and not from deeper layers where mycorrhiza is normally found. Finally, the result of this analysis indicates that there may be no correlation between biological soil indicators based on bacteria and biological soil indicators based on fungi.

Due to the complexity of the soil system, limitations are needed to make it more manageable to perform analyses. Even if this is necessary, there is a risk for inaccuracies in the results. There are still uncertainties about how the soil can be described and how this is related to the concept of soil health, and therefore more research is needed to improve understanding and knowledge.

References

[1] *8.8C: Firmicutes - Biology LibreTexts*. [Online]. Available: https://bio.libretexts.org/Bookshelves/Microbiology/Microbiology_(Boundless)/08%3A_Microbial_Evolution_Phylogeny_and_Diversity/8.08%3A_Gram-Positive_Bacteria_and_Actinobacteria/8.8C%3A_Firmicutes.

[2] American society for microbiology, *Soil Actinomycetes*, Jun. 2008. [Online]. Available: https://asm.org/image-gallery/soil-actinomycetes.

[3] E. A. Barka et al., "Taxonomy, Physiology, and Natural Products of Actinobacteria," *Microbiology and Molecular Biology Reviews*, vol. 80, no. 1, pp. 1–43, Mar. 2016, ISSN: 1092-2172. DOI: 10.1128/mmbr.00019-15.

[4] I. T. Bödeker, B. D. Lindahl, Å. Olson, and K. E. Clemmensen, "Mycorrhizal and saprotrophic fungal guilds compete for the same organic substrates but affect decomposition differently," *Functional Ecology*, vol. 30, no. 12, pp. 1967–1978, Dec. 2016, ISSN: 13652435. DOI: 10.1111/1365-2435.12677.

[5] P. Bohman, "Vattenprovtagning av DNA från fisk, kräftor och musslor-en kunskapssammanställning," Tech. Rep., Sep. 2018. [Online]. Available: http://pub.epsilon.slu.se/.

[6] Bracklin Anna, "Jord i god kultur," 2005.

[7] Z. Breijyeh, B. Jubeh, and R. Karaman, *Resistance of gram-negative bacteria to current antibacterial agents and approaches to resolve it*, Mar. 2020. DOI: 10.3390/molecules25061340.

[8] Britannica, *Archaea | Definition, Characteristics, & Examples | Britannica*, 2025. [Online]. Available: https://www.britannica.com/science/archaea.

[9] Britannica, *Phospholipid | Cell Membrane, Lipid Bilayer & Fatty Acids | Britannica*, 2025. [Online]. Available: https://www.britannica.com/science/phospholipid.

[10] Britannica, *Taxonomy - Classification, Naming, Categorizing | Britannica*, 2025. [Online]. Available: https://www.britannica.com/science/taxonomy/Making-a-classification.

[11] A. Burian, Q. Mauvisseau, M. Bulling, S. Domisch, S. Qian, and M. Sweet, *Improving the reliability of eDNA data interpretation*, Jul. 2021. DOI: 10.1111/1755-0998.13367.

[12] W. H. Burke et al., "Primer-Directed Enzymatic Amplification of DNA with a Thermostable DNA Polymerase," Tech. Rep., 1986, p. 813. DOI: 10.1126/science.2448875.

[13] S. Caspersen and C. Oskarsson, "Mykorrhiza i trädgårdsodling : en handbok," Tech. Rep., 2024. DOI: 10.54612/a.1ietj3ct15. [Online]. Available: https://res.slu.se/id/publ/128113.

[14] CBD, *Sustaining life on earth : how the Convention on Biological Diversity promotes nature and humal well-being*. Secretariat of the Convention on Biological Diversity, Apr. 2000, ISBN: 9280719041.

[15] CD Genomics, *Introduction to Amplicon Sequence Variants - CD Genomics*. [Online]. Available: https://www.cd-genomics.com/resource-introduction-amplicon-sequence-variants.html.

[16] CD Genomics, *Operational Taxonomic Unit (OTU) and OTU Clustering - CD Genomics*. [Online]. Available: https://www.cd-genomics.com/microbioseq/operational-taxonomic-unit-otu-and-otu-clustering.html.

[17] Claes Andersson, *Lektion 4.2 0128 - Komplexitet.pptx: SEE075 Evolution och självorganisation i biologiska system*, Jan. 2021. [Online]. Available: https://chalmers.instructure.com/courses/12240/files/1130297?module_item_id=145998.

[18] *Daggmask – Wikipedia*, Apr. 2025. [Online]. Available: https://sv.wikipedia.org/wiki/Daggmask.

[19] A. Diana et al., "eDNAPlus: A Unifying Modeling Framework for DNA-based Biodiversity Monitoring," *Journal of the American Statistical Association*, 2024, ISSN: 1537274X. DOI: 10.1080/01621459.2024.2412362.

[20] P. Drenning, "Soil Functions and Ecosystem Services A Literature Review (Part 2/2)," Gothenburg, Tech. Rep., 2021.

[21] Eurofins, *GC-MS Analytical Testing Lab Services | EAG Laboratories*. [Online]. Available: https://www.eag.com/techniques/mass-spec/gas-chromatography-mass-spectrometry-gc-ms/.

[22] Eurofins, *Soil Life Monitor - Eurofins Agro*. [Online]. Available: https://www.eurofins-agro.com/en/soil-life-in-a-healthy-soil#methodss.

[23] European Commission, "EU:s markstrategi för 2030 - Frisk mark till förmån för människor, livsmedel, natur och klimat," Brussels, Tech. Rep., Nov. 2021. [Online]. Available: https://op.europa.eu/en/publication-detail/-/publication/4ebd2586-fc85-11ea-b44f-01aa75ed71a1/..

[24] European Commission, *LUCAS - ESDAC - European Commission*. [Online]. Available: https://esdac.jrc.ec.europa.eu/projects/lucas.

[25] EUROSTAT, *Agri-environmental indicator - consumption of pesticides - Statistics Explained - Eurostat*, Apr. 2025. [Online]. Available: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Agri-environmental_indicator_-_consumption_of_pesticides.

[26] EUROSTAT, *Fully organic farms in the EU - Statistics Explained - Eurostat*, Mar. 2024. [Online]. Available: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Fully_organic_farms_in_the_EU#Organic_farm_specialisation.

[27] EUROSTAT, *Statistics Explained - Eurostat*. [Online]. Available: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Main_Page.

[28] EUROSTAT, *Use of fertilisers in EU agriculture down 10% in 2022 - News articles - Eurostat*, Jun. 2024. [Online]. Available: https://ec.europa.eu/eurostat/web/products-eurostat-news/w/ddn-20240628-1.

[29] O. ; Fernandez-Ugalde, S. ; Scarpa, A. ; Orgiazzi, P. ; Panagos, M. Van Liedekerke, and M. A. Jones, "LUCAS 2018 Soil Module Presentation of dataset and results," 2022, ISBN: 978-92-76-54832-4. DOI: 10.2760/215013. [Online]. Available: https://ec.europa.eu/jrc.

[30] *Gram-negative bacteria - Wikipedia*, Jun. 2025. [Online]. Available: https://en.wikipedia.org/wiki/Gram-negative_bacteria.

[31] *Gram-positive bacteria - Wikipedia*, Jun. 2025. [Online]. Available: https://en.wikipedia.org/wiki/Gram-positive_bacteria.

[32] "GSBAtlas_ch2_Enchytraeids," *Global Soil Biodiversity Atlas*,

[33] *Hoppstjärtar – Wikipedia*, Feb. 2025. [Online]. Available: https://sv.wikipedia.org/wiki/Hoppstj%C3%A4rtar.

[34] International Soil Reference and Information Centre, *The skin of the Earth: All about soil.* [Online]. Available: https://isric-prod.netlify.app/all-about-soil/.

[35] Jaadi Zakaria, *Principal Component Analysis (PCA): Explained Step-by-Step | Built In.* [Online]. Available: https://builtin.com/data-science/step-step-explanation-principal-component-analysis.

[36] Jernigan Ashley, "SOIL MESOFAUNA," [Online]. Available: www.chaosofdelight.com.

[37] L. Johansson, "Daggmaskar-Hur påverkas de av jordbruk?," 2021.

[38] Johansson Therese, *Från jord till skog – här trivs våra vanligaste trädslag - Skogssällskapet.se*, Mar. 2022. [Online]. Available: https://www.skogssallskapet.se/kunskapsbank/artiklar/2022-03-18-fran-jord-till-skog---har-trivs-vara-vanligaste-tradslag.html.

[39] *Kvalster – Wikipedia*, Mar. 2025. [Online]. Available: https://sv.wikipedia.org/wiki/Kvalster.

[40] M. Labouyrie et al., "Patterns in soil microbial diversity across Europe," *Nature Communications*, vol. 14, no. 1, Dec. 2023, ISSN: 20411723. DOI: 10.1038/s41467-023-37937-4.

[41] J. Ladyman, J. Lambert, and K. Wiesner, *What is a complex system?* 2013. DOI: 10.1007/s13194-012-0056-8.

[42] Lennart Råde and Bertil Westergren, *Mathematics Handbook for Science and Engineering.* 1988, ISBN: 9789144128436.

[43] M. Liu, L. J. Clarke, S. C. Baker, G. J. Jordan, and C. P. Burridge, *A practical guide to DNA metabarcoding for entomological ecologists*, Jun. 2020. DOI: 10.1111/een.12831.

[44] Lowenfels Jeff and Lewis Wayne, "Teaming with Microbes," 2010.

[45] Lucas Learns, *GC-MS For Beginners (Gas Chromatography Mass Spectrometry) - YouTube*, Oct. 2022. [Online]. Available: https://www.youtube.com/watch?v=F9NUXkyIl5A&t=190s.

[46] *Membrane transport protein - Wikipedia*, 2025. [Online]. Available: https://en.wikipedia.org/wiki/Membrane_transport_protein.

[47] *Microfauna - Wikipedia*, Jun. 2025. [Online]. Available: https://en.wikipedia.org/wiki/Microfauna.

[48] *Millipede - Wikipedia*, Jun. 2025. [Online]. Available: https://en.wikipedia.org/wiki/Millipede.

[49] *Mullvad – Wikipedia*, Dec. 2024. [Online]. Available: https://sv.wikipedia.org/wiki/Mullvad.

[50] *Mycorrhiza - Wikipedia*, Jun. 2025. [Online]. Available: https://en.wikipedia.org/wiki/Mycorrhiza.

[51] National Center for Biotechnology Information, *Taxonomy browser (Niallia circulans)*. [Online]. Available: https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Info&id=1397.

[52] Natural Resources Conservation Service, *Soil Health | Natural Resources Conservation Service*. [Online]. Available: https://www.nrcs.usda.gov/conservation-basics/natural-resource-concerns/soils/soil-health.

[53] Naturvårdsverket, *Vad är biologisk mångfald?* Dec. 2023. [Online]. Available: https://www.naturvardsverket.se/amnesomraden/biologisk-mangfald/vad-ar-biologisk-mangfald/.

[54] *Nematode | Definition, Description, Diseases, & Facts | Britannica*, Jun. 2025. [Online]. Available: https://www.britannica.com/animal/nematode.

[55] Newcastle University, *Numeracy, Maths and Statistics - Academic Skills Kit*. [Online]. Available: https://www.ncl.ac.uk/webtemplate/ask-assets/external/maths-resources/statistics/regression-and-correlation/coefficient-of-determination-r-squared.html.

[56] U. N. Nielsen, D. H. Wall, and J. Six, "Soil Biodiversity and the Environment," *Annual Review of Environment and Resources*, vol. 40, pp. 63–90, Nov. 2015, ISSN: 15435938. DOI: 10.1146/annurev-environ-102014-021257.

[57] C. Pankhurst, B. Doube, and V. Gupta, *Biological indicators of soil health*. CAB International, 1998, ISBN: 0851991580.

[58] E. Pereira, V. Coelho, R. M. Tavares, T. Lino-Neto, and P. Baptista, "Effect of competitive interactions between ectomycorrhizal and saprotrophic fungi on Castanea sativa performance," *Mycorrhiza*, vol. 22, no. 1, pp. 41–49, Jan. 2012, ISSN: 09406360. DOI: 10.1007/s00572-011-0379-x.

[59] S. Põlme et al., "FungalTraits: a user-friendly traits database of fungi and fungus-like stramenopiles," *Fungal Diversity*, vol. 105, no. 1, Nov. 2020, ISSN: 18789129. DOI: 10.1007/s13225-020-00466-2.

[60] *Protozoan - Microbes, Ecology, Industrial | Britannica*, Mar. 2025. [Online]. Available: https://www.britannica.com/science/protozoan/Ecological-and-industrial-importance-of-protozoans.

[61] Quick Biomass Basics, *Gas chromatography | GC - YouTube*, Aug. 2020. [Online]. Available: https://www.youtube.com/watch?v=UycPljfrnWo.

[62] M. Rajagopal and S. Walker, "Envelope structures of gram-positive bacteria," in *Current Topics in Microbiology and Immunology*, vol. 404, Springer Verlag, Jan. 2017. DOI: 10.1007/82{_}2015{_}5021.

[63] *Rundmaskar – Wikipedia*, Nov. 2024. [Online]. Available: https://sv.wikipedia.org/wiki/Rundmaskar.

[64] Statistics Division, *UNSD — Methodology*. [Online]. Available: https://unstats.un.org/unsd/methodology/m49/.

[65] Svensk kolinlagring, *Regenerativt Jordbruk: Svensk Kolinlagring*. [Online]. Available: https://regenerativtjordbruk.nu/.

[66] Svenska Miljöinstitutet, *Analysera biologisk mångfald med miljö-DNA - IVL.se*, Oct. 2024. [Online]. Available: https://www.ivl.se/vart-erbjudande/vara-tjanster/analysera-biologisk-mangfald-med-miljo-dna.html.

[67] Svenska Miljöinstitutet, *eDNA - IVL.se*, Sep. 2024. [Online]. Available: https://www.ivl.se/vart-erbjudande/vara-omraden/biologisk-mangfald/edna.html.

[68] Sveriges geologiska undersökning, *Erosion*, Nov. 2020. [Online]. Available: https://www.sgu.se/om-geologi/jord/fran-istid-till-nutid/erosion-och-igenvaxning/erosion/.

[69] *Sveriges jordar - Naturskyddsföreningen*, Aug. 2021. [Online]. Available: https://www.naturskyddsforeningen.se/faktablad/sveriges-jordar/.

[70] United Nations, *THE 17 GOALS | Sustainable Development*. [Online]. Available: https://sdgs.un.org/goals.

[71] University of Minnesota extension, *Soil compaction | UMN Extension*, 2018. [Online]. Available: https://extension.umn.edu/soil-management-and-health/soil-compaction#nutrient-uptake-1147261.

A

Appendix 1: Supplementary correlations between biological soil indicators

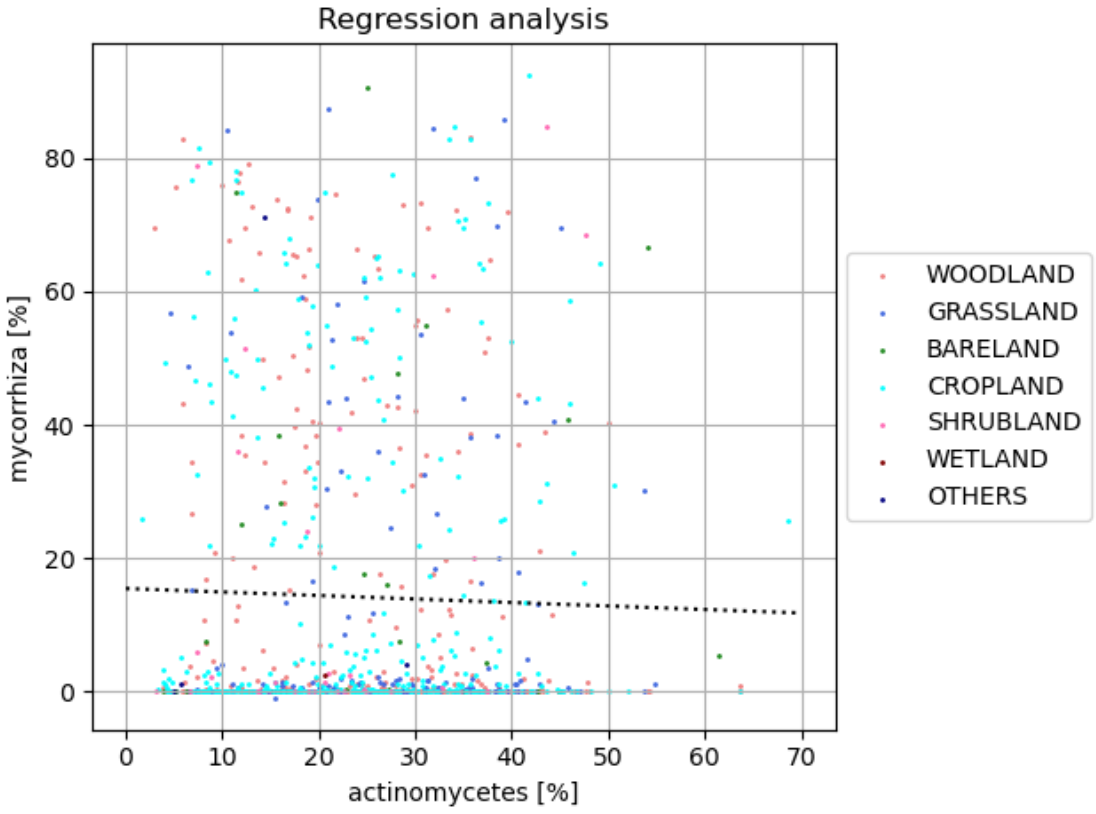

Figure A.1: Correlation between the presence of actinomycetes and the presence of saprotrophs.

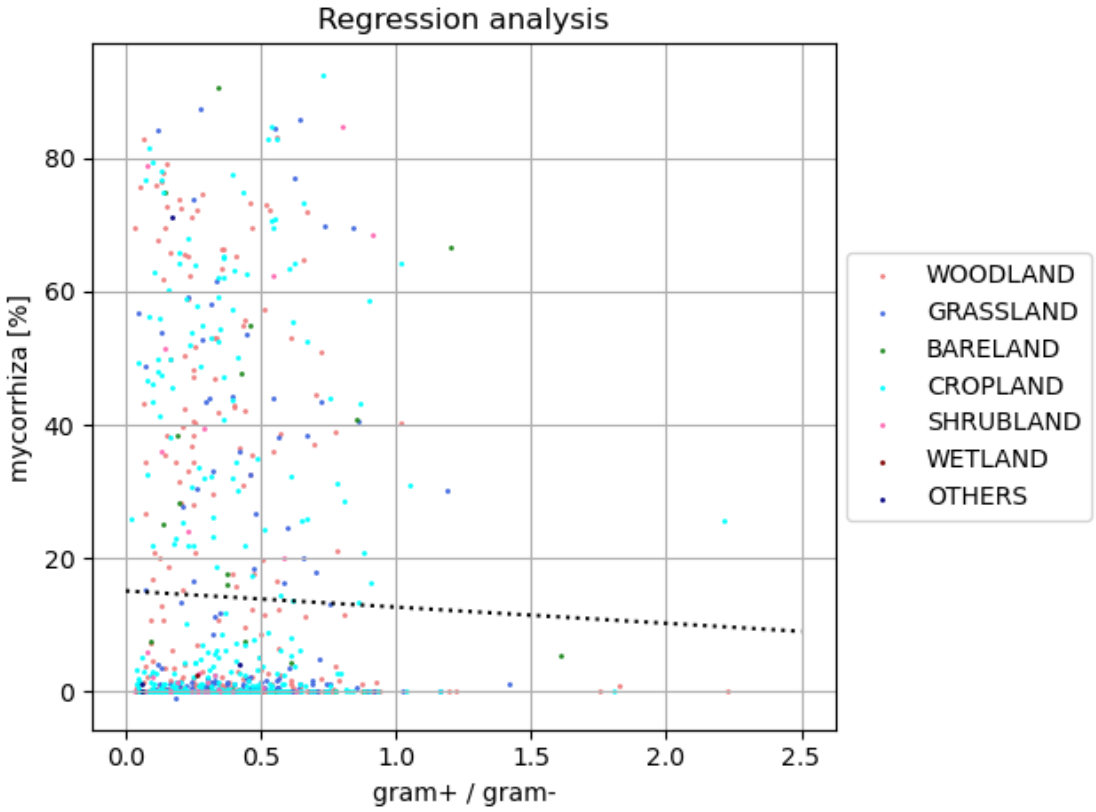

Figure A.2: Scatter plot showing the correlation between the bacterial ratio and the presence of mycorrhiza.

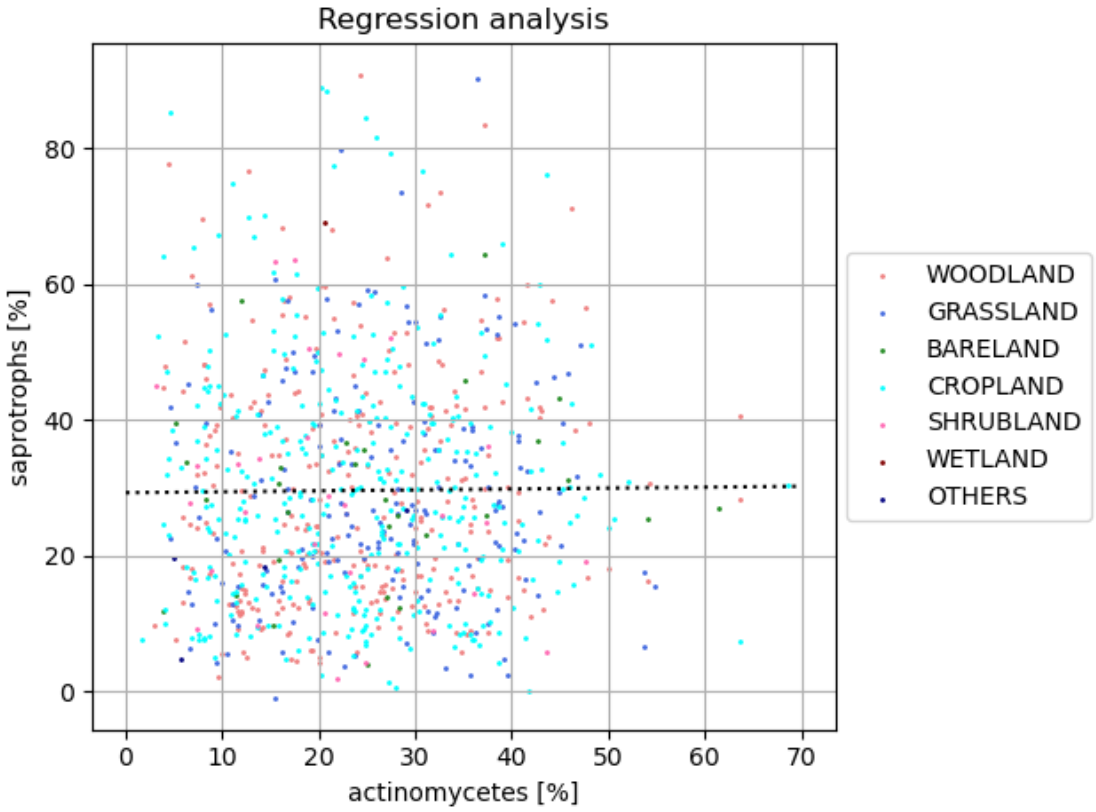

Figure A.3: Correlation between the presence of actinomycetes and the presence of saprotrophs.

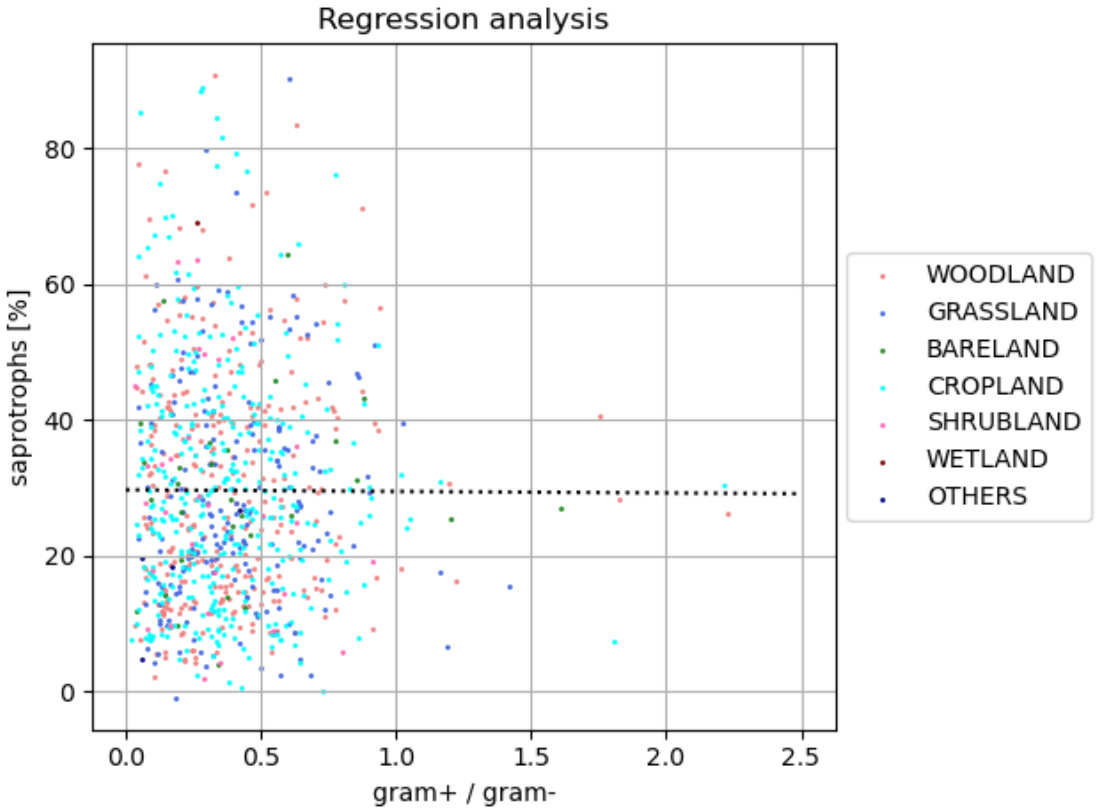

Figure A.4: Correlation between the ratio between gram-positive and gram-negative bacteria and the presence of saprotrophs.

B

Appendix 2: Correlations between biological soil indicators and characteristic soil indicators

In this appendix the supplementary correlation graphs between the chosen biological soil indications and the characteristic soil indicators (see Table 4.5) are shown. Figure B.1-B.6 are related to data from France. Figure B.7-B.12 are related to data from Spain.

B. Appendix 2: Correlations between biological soil indicators and characteristic soil indicators

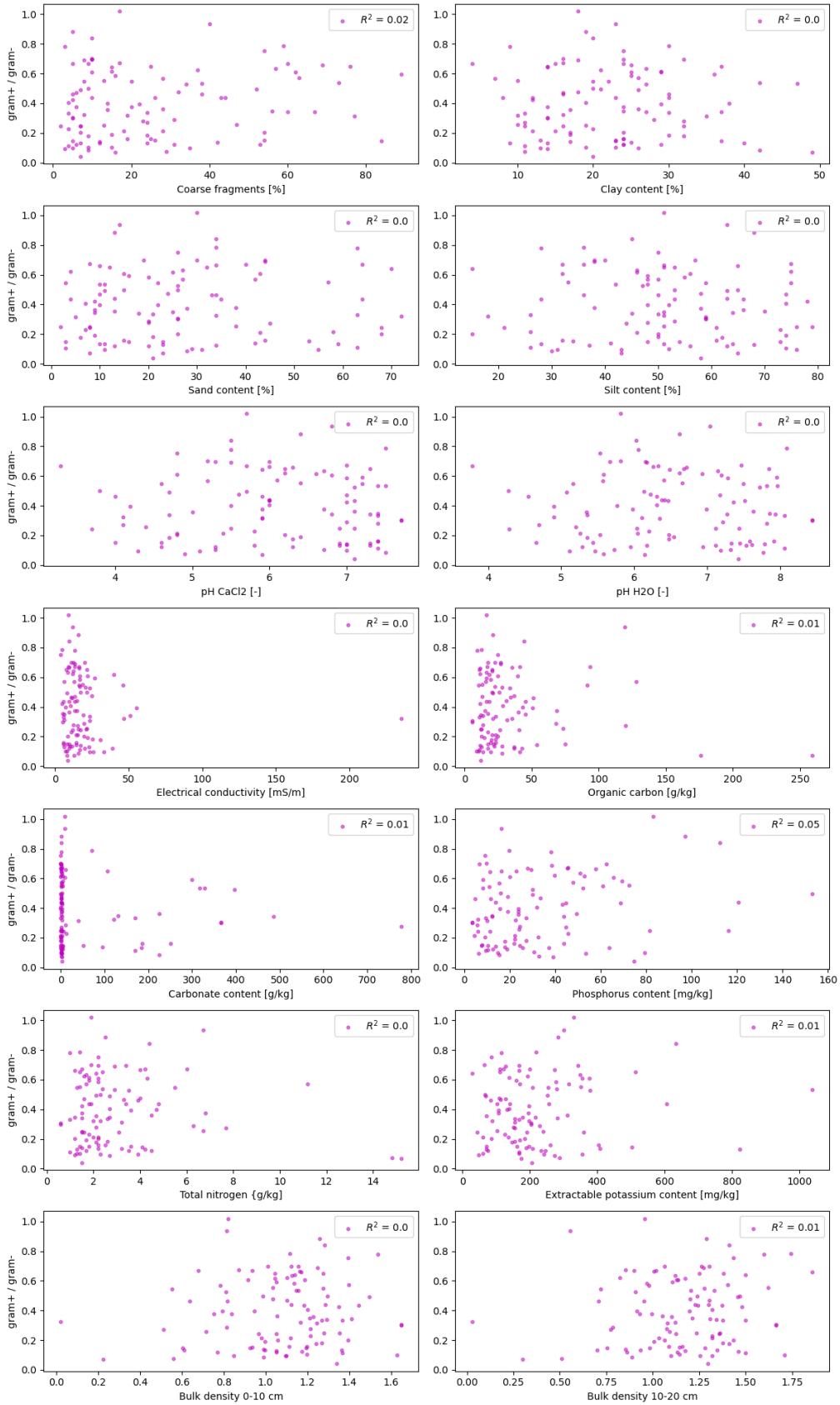

Figure B.1: Visualisation of the correlations between the biological soil indicator of the ratio of gram+ and gram− bacteria and the characteristic soil parameters showed in Table 3.2. All the data is related to France. The R^2 values are shown in the upper right corner of each graph.

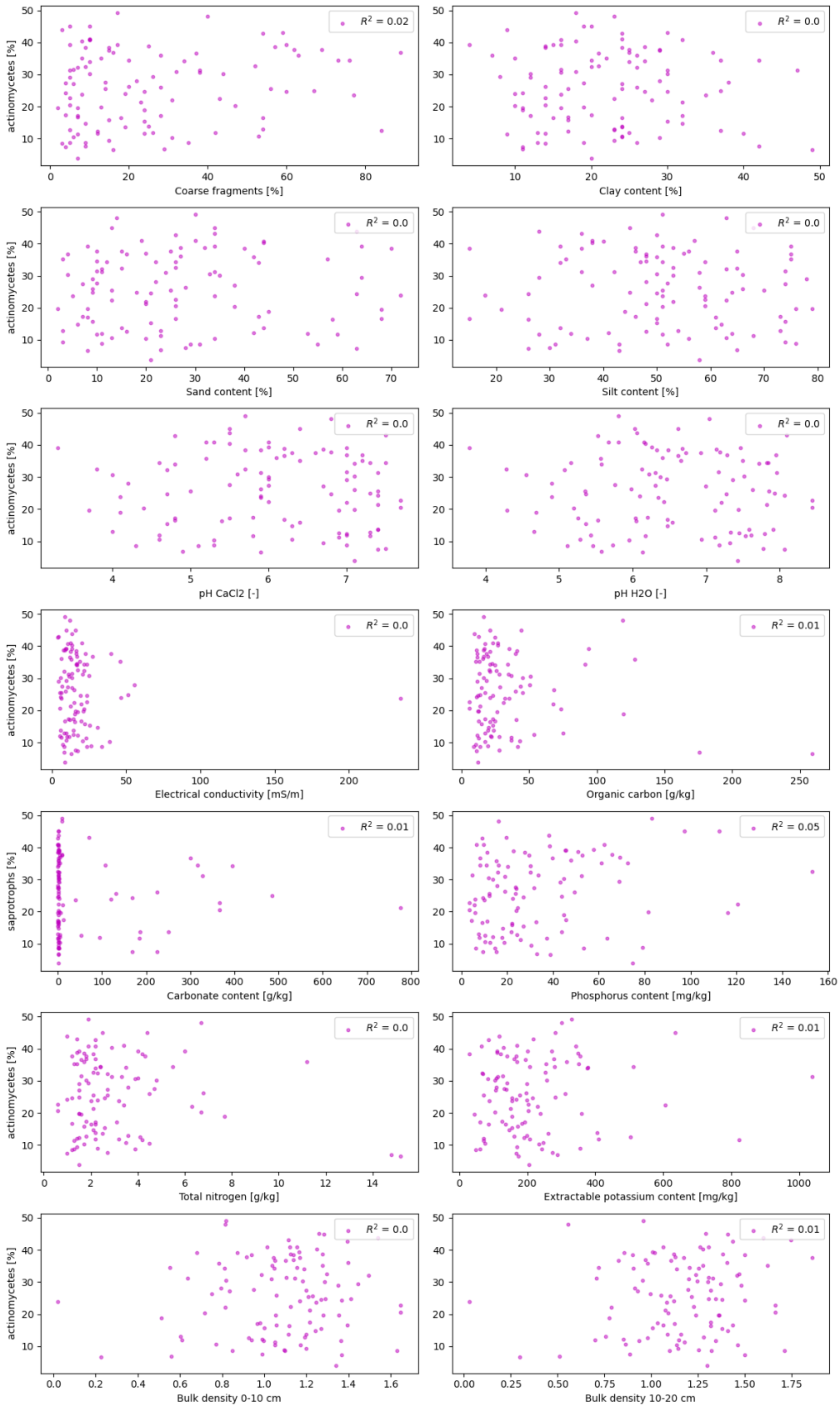

Figure B.2: Visualisation of the correlations between the biological soil indicator of the presence of actinomycetes bacteria and the characteristic soil parameters showed in Table 3.2. All the data is related to France. The R^2 values are shown in the upper right corner of each graph.

B. Appendix 2: Correlations between biological soil indicators and characteristic soil indicators

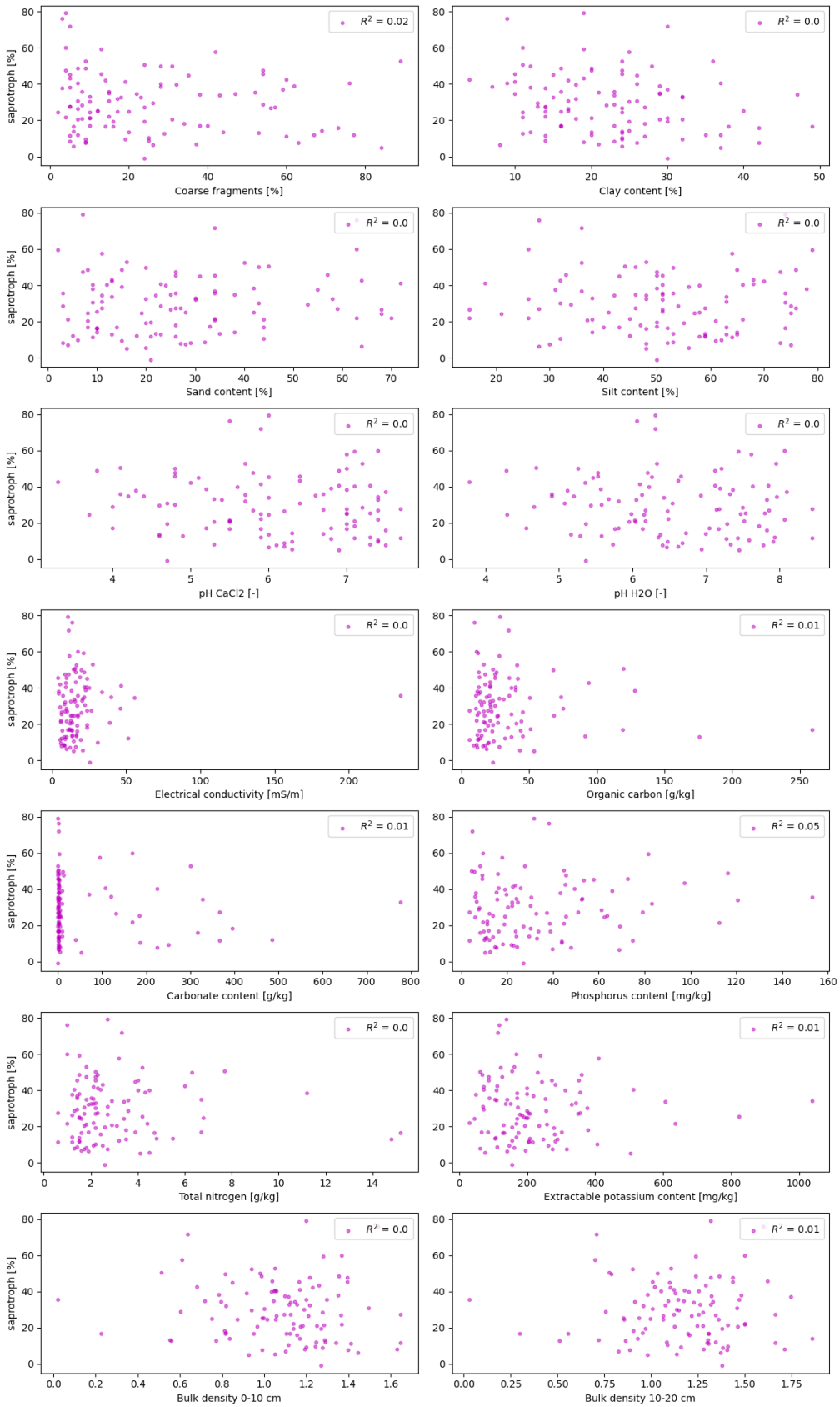

Figure B.3: Visualisation of the correlations between the biological soil indicator of the presence of saprotrophs and the characteristic soil parameters showed in Table 3.2. All the data is related to France. The R^2 values are shown in the upper right corner of each graph.

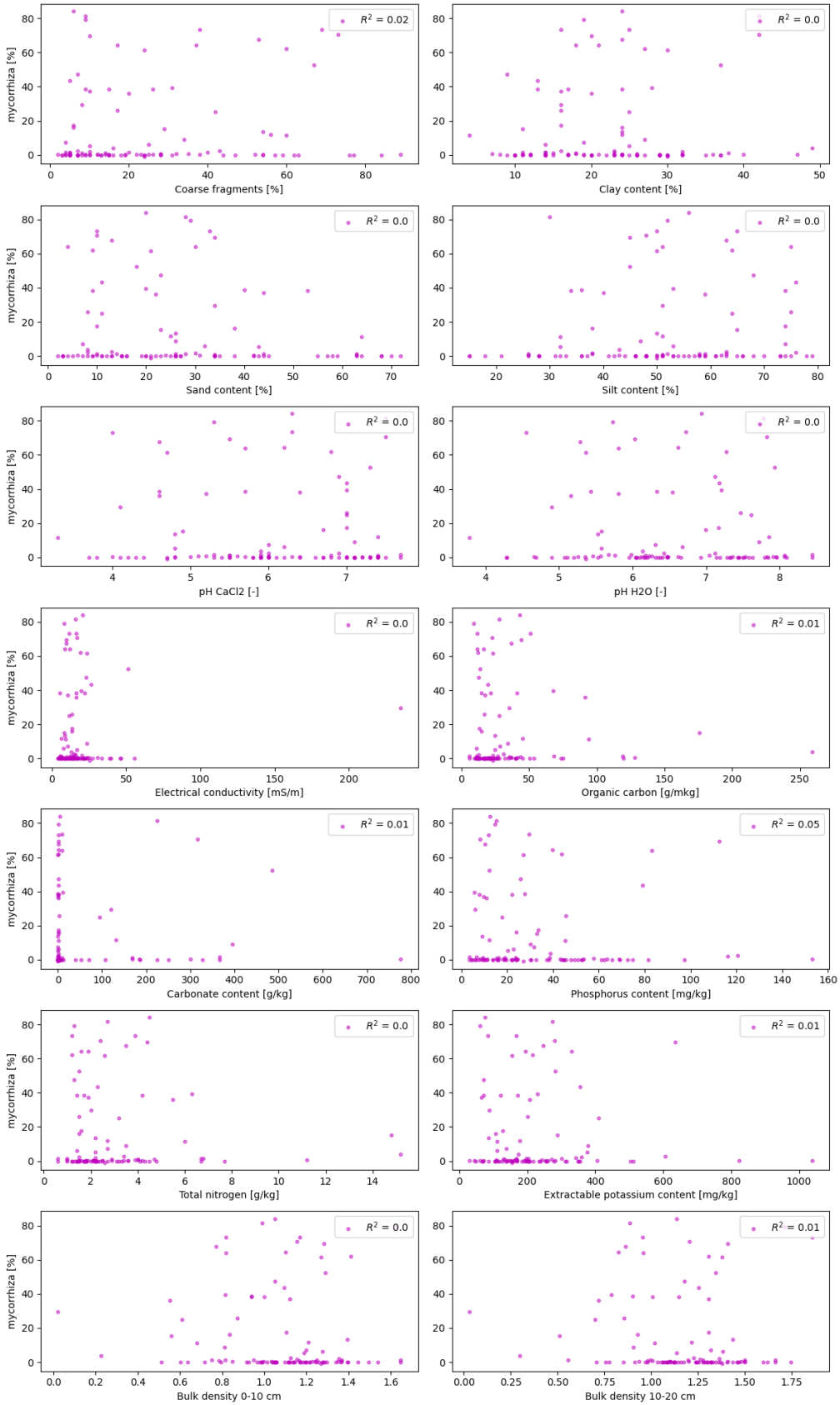

Figure B.4: Visualisation of the correlations between the biological soil indicator of the presence of mycorrhiza and the characteristic soil parameters showed in Table 3.2. All the data is related to France. The R^2 values are shown in the upper right corner of each graph.

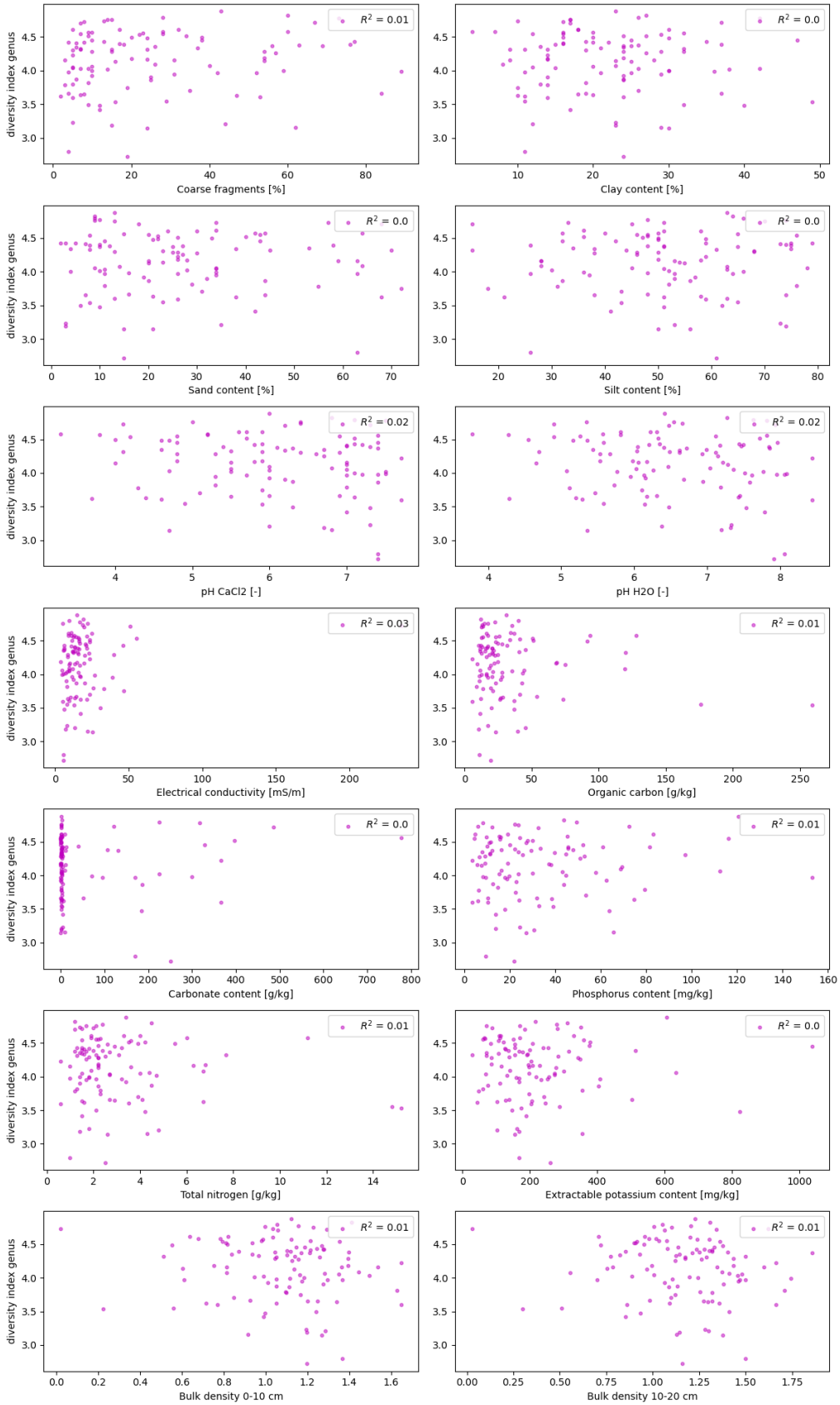

Figure B.5: Visualisation of the correlations between the biological soil indicator of the diversity index based on genus and the characteristic soil parameters showed in Table 3.2. All the data is related to France. The R^2 values are shown in the upper right corner of each graph.

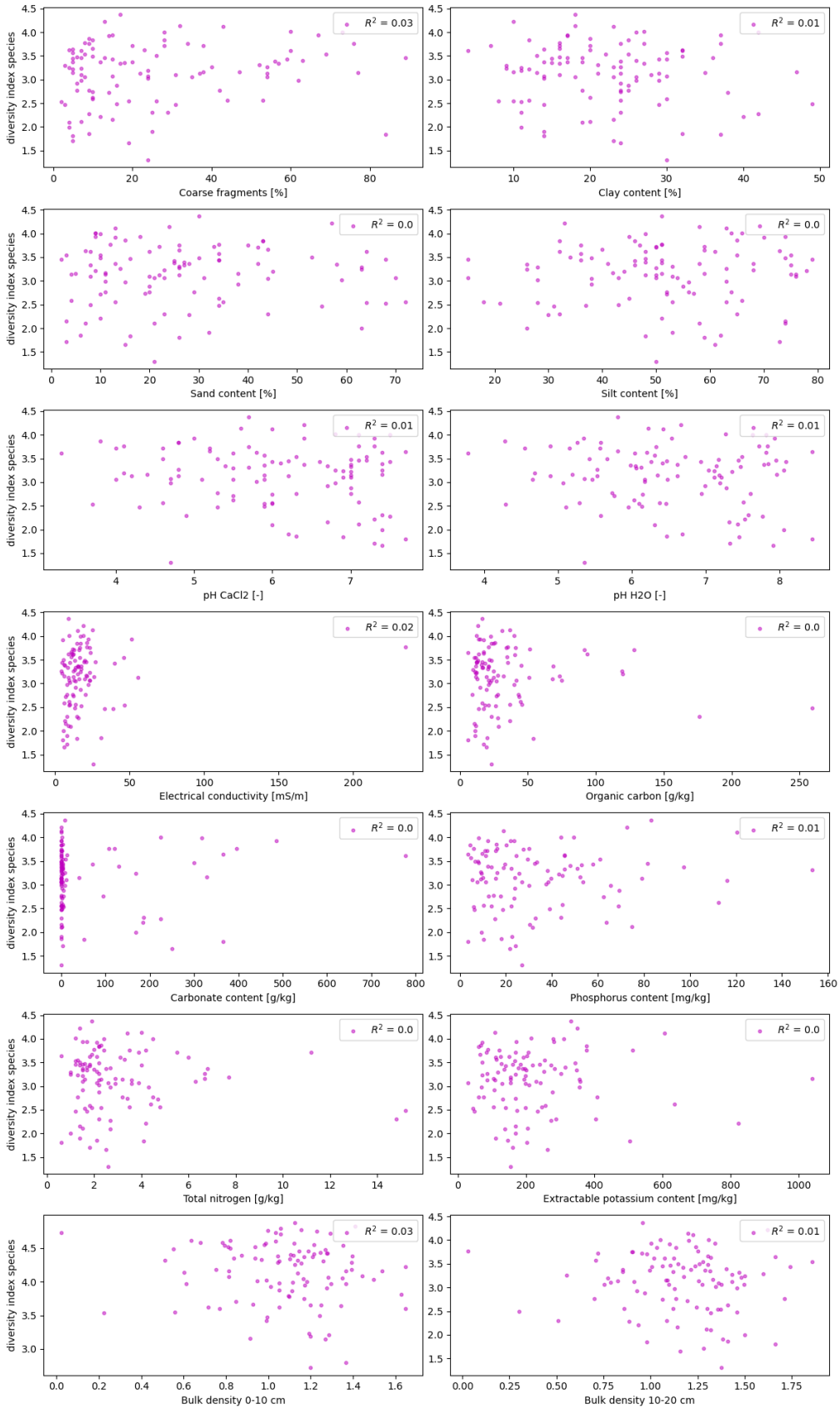

Figure B.6: Visualisation of the correlations between the biological soil indicator of the diversity index based on species and the characterstic soil parameters showed in Table 3.2. All the data is related to France. The R^2 values are shown in the upper right corner of each graph.

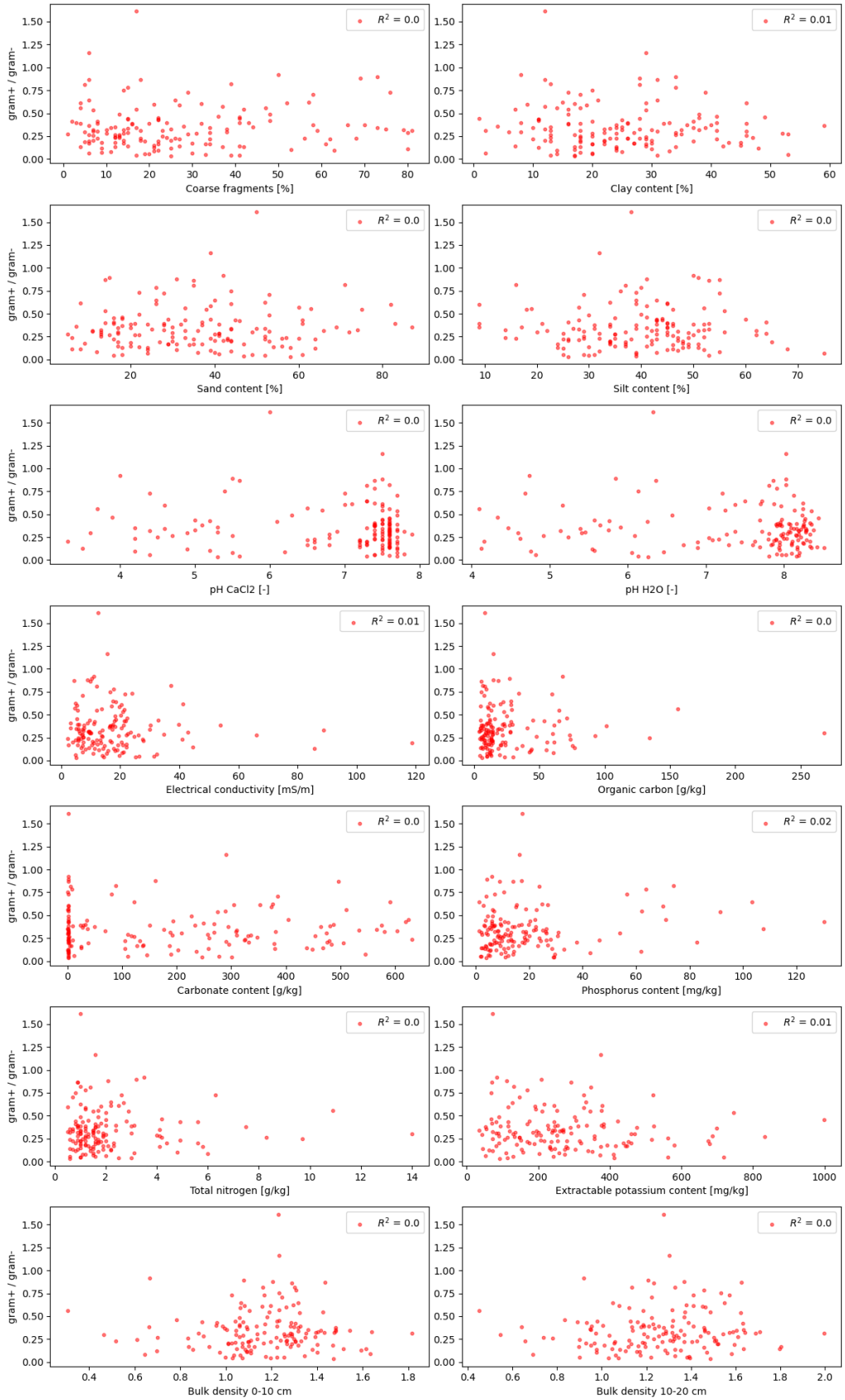

Figure B.7: Visualisation of the correlations between the biological soil indicator of the ratio of gram+ and gram− bacteria and the characteristic soil parameters showed in Table 3.2. All the data is related to Spain. The R^2 values are shown in the upper right corner of each graph.

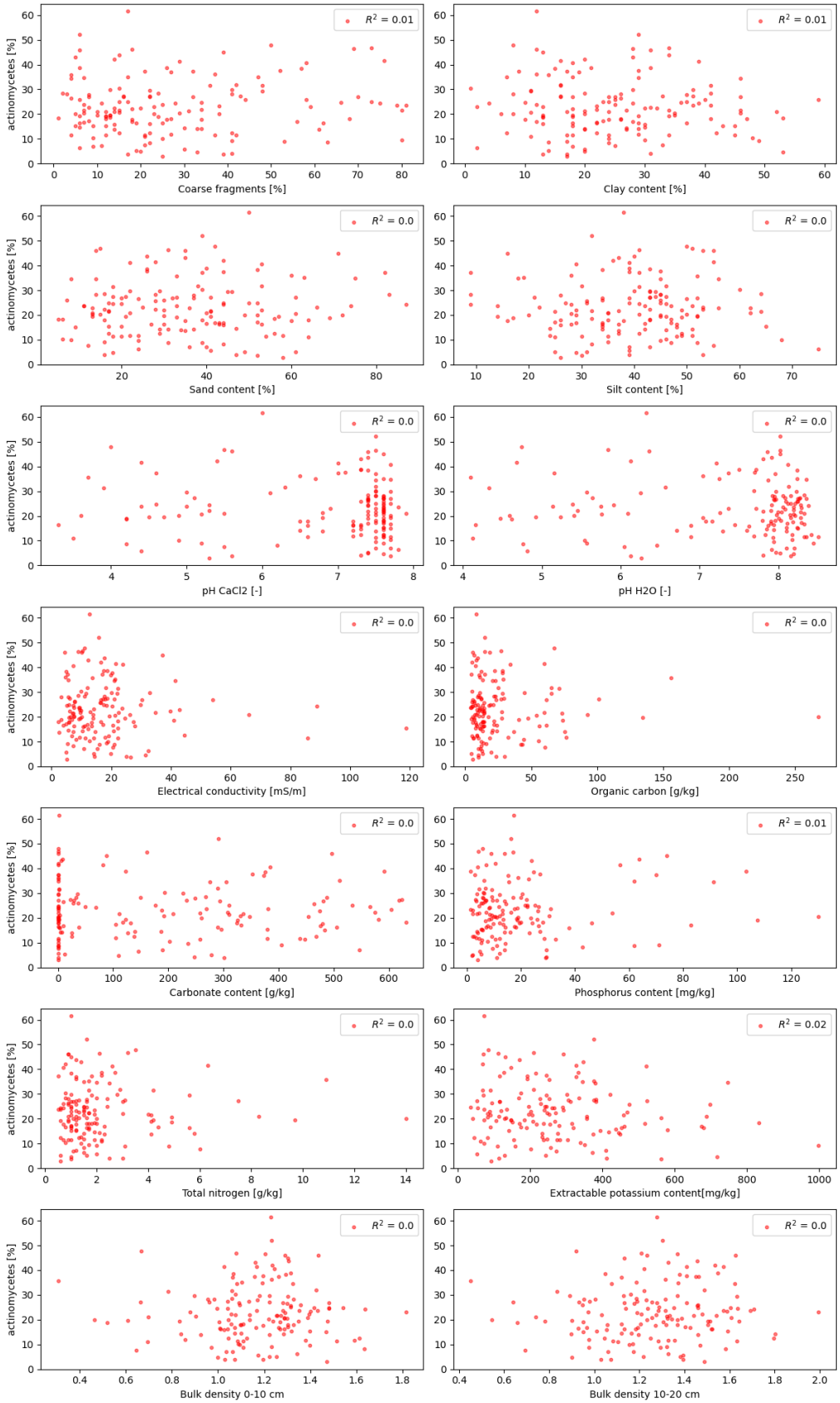

Figure B.8: Visualisation of the correlations between the biological soil indicator of the presence of actinomycetes bacteria and the characteristic soil parameters showed in Table 3.2. All the data is related to Spain. The R^2 values are shown in the upper right corner of each graph.

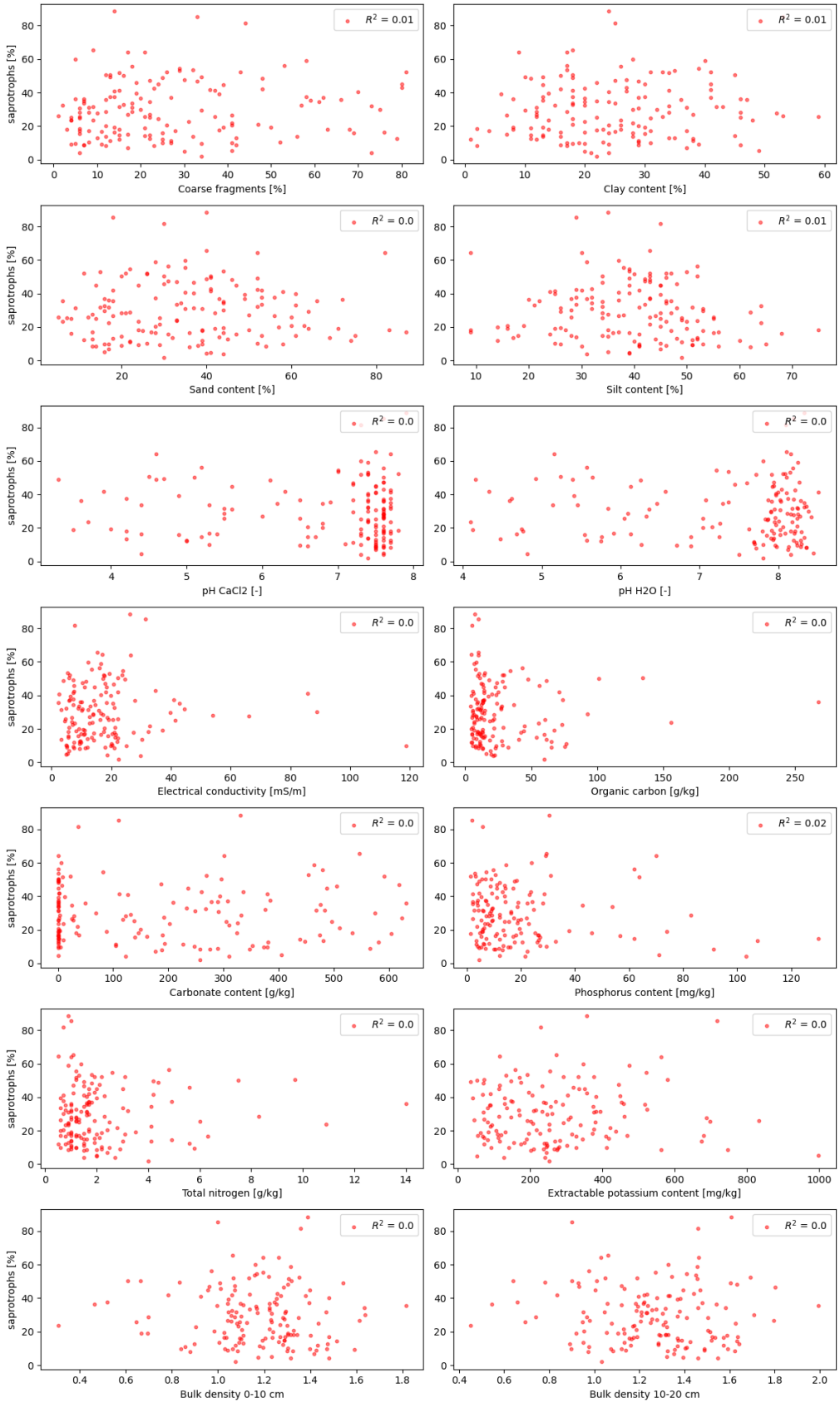

Figure B.9: Visualisation of the correlations between the biological soil indicator of the presence of saprotrophs and the characteristic soil parameters showed in Table 3.2. All the data is related to Spain. The R^2 values are shown in the upper right corner of each graph.

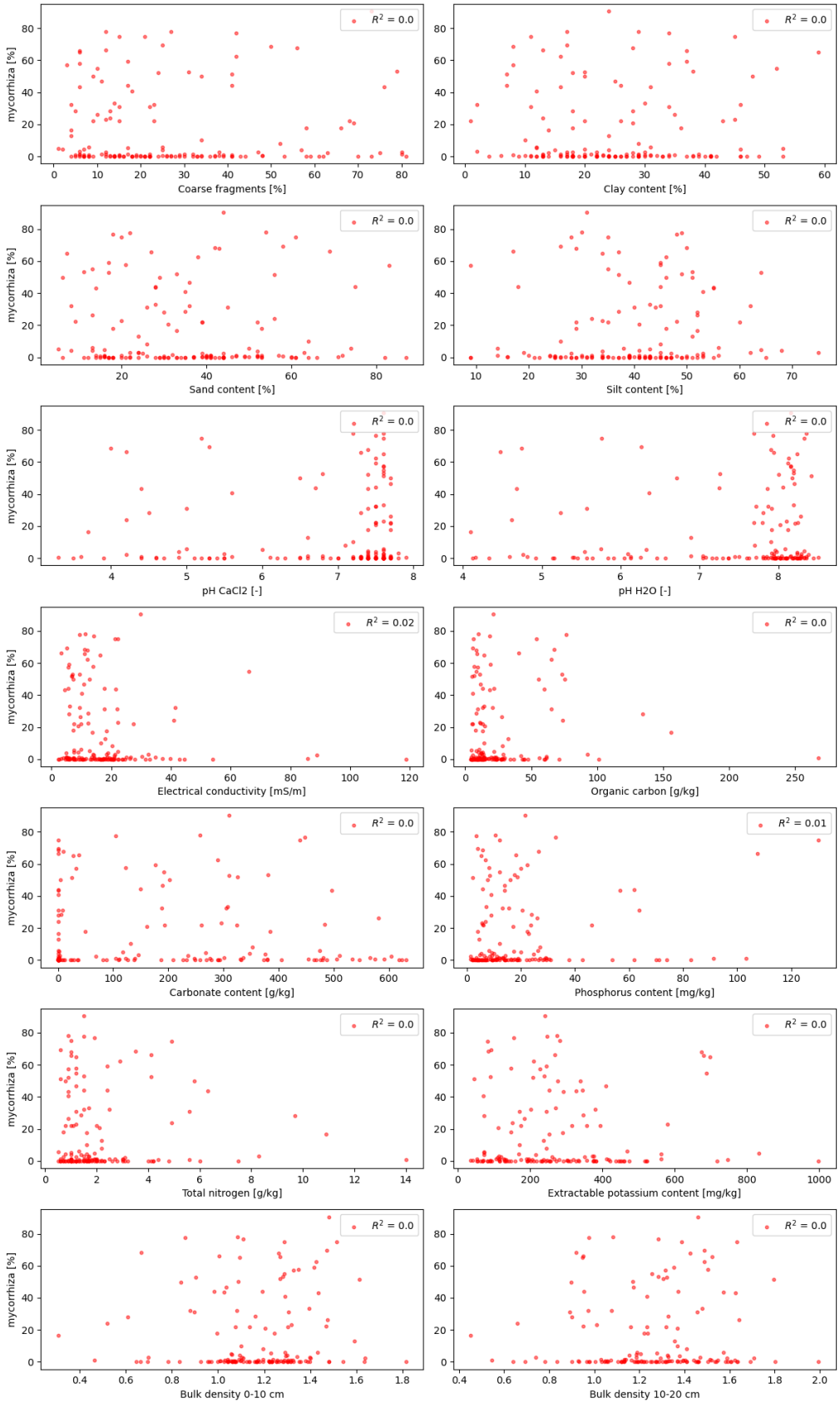

Figure B.10: Visualisation of the correlations between the biological soil indicator of the presence of mycorrhiza and the characteristic soil parameters showed in Table 3.2. All the data is related to Spain. The R^2 values are shown in the upper right corner of each graph.

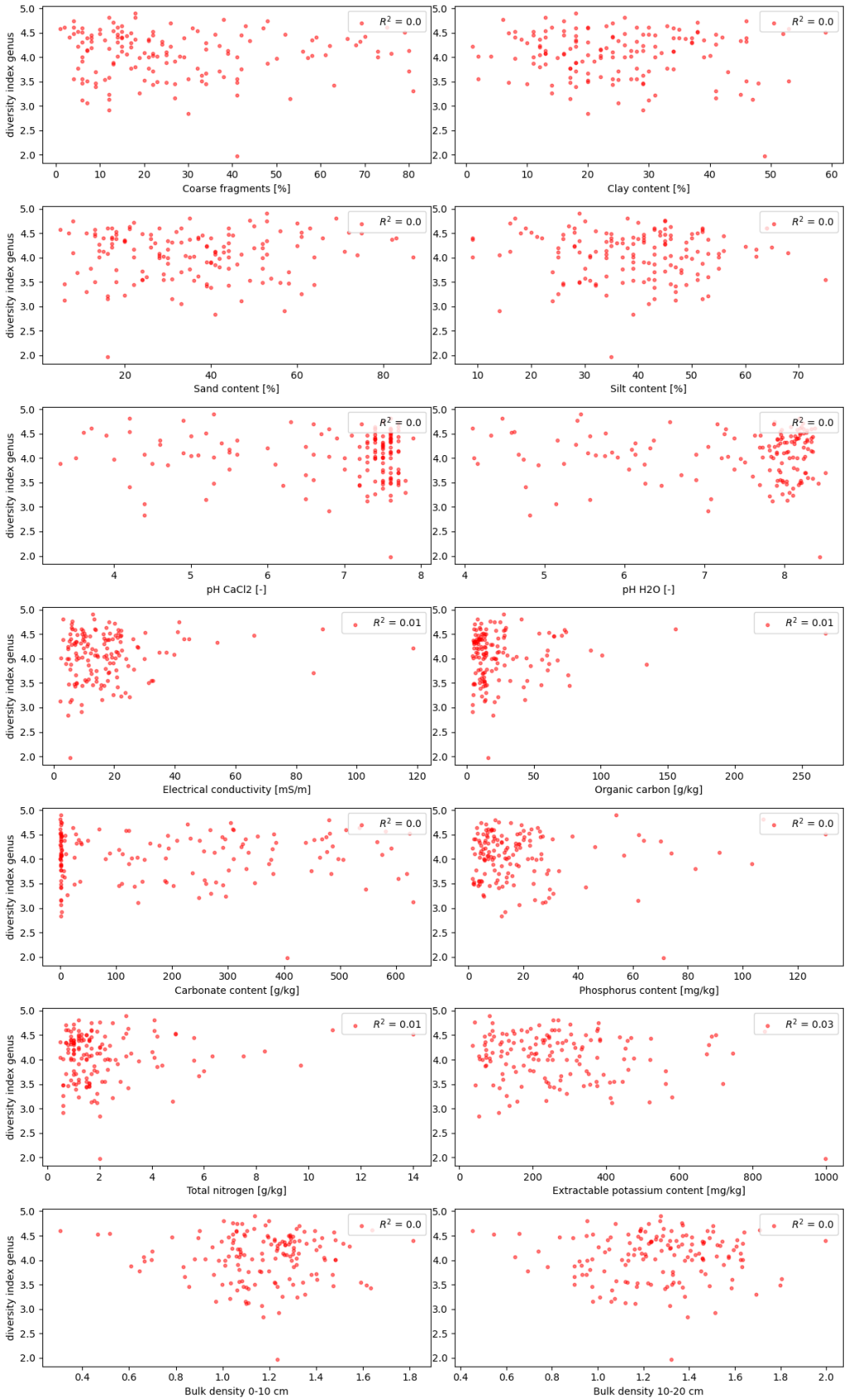

Figure B.11: Visualisation of the correlations between the biological soil indicator of the diversity index based on genus and the characteristic soil parameters showed in Table 3.2. All the data is related to Spain. The R^2 values are shown in the upper right corner of each graph.

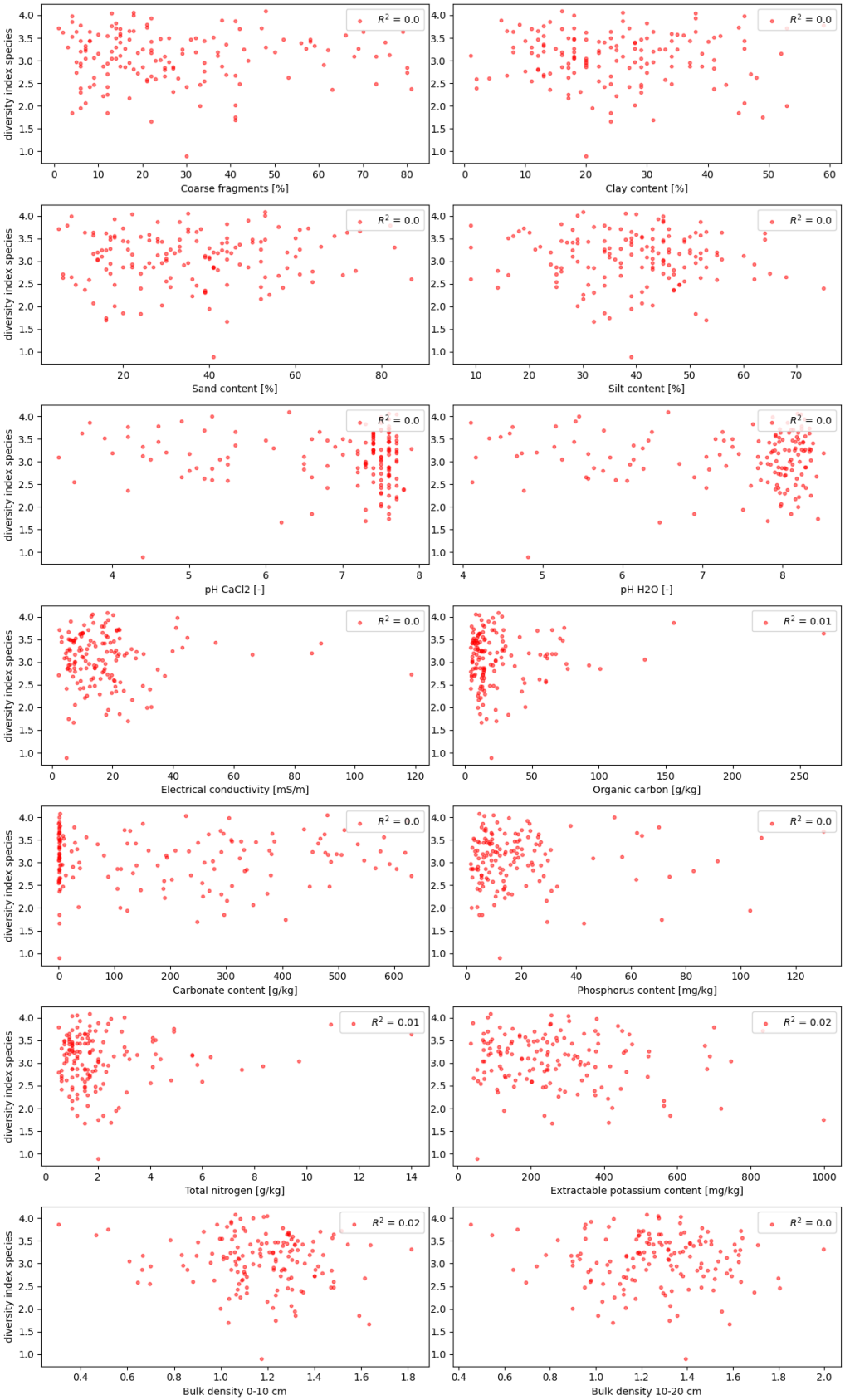

Figure B.12: Visualisation of the correlations between the biological soil indicator of the diversity index based on species and the characteristic soil parameters showed in Table 3.2. All the data is related to Spain. The R^2 values are shown in the upper right corner of each graph.

C

Appendix 3: Results from the PCR analysis

The rows with highest value for λ_i is marked with orange and the rows with highest value of the first element of v_i (the element corresponding to the specific biological soil indicator) is marked with yellow.

C.1 PCR France

Bacterial ratio
Table C.1 shows the results of the principal component analysis for the biological soil indicator of the fraction of gram+ / gram− bacteria.

Table C.1: The results from the principal component analysis for the biological soil indicator of the fraction of $gram+/gram-$ bacteria. λ_i is the eigenvalue, σ_i is the spread and v_i is the eigenvector.

i	λ_i	σ_i	v_i
1	3.40	1.84	[-0.07 ...]
2	2.73	1.65	[-0.01 ...]
3	1.70	1.30	[0.15 ...]
4	1.47	1.21	[0.43 ...]
5	1.21	1.10	[0.27 ...]
6	1.07	1.03	[-0.52 ...]
7	0.71	0.84	[0.61 ...]
8	0.59	0.77	[0.25 ...]
9	0.00	0.02	[-0.00 ...]
10	0.04	0.20	[-0.00 ...]
11	0.13	0.36	[-0.02 ...]
12	0.56	0.68	[-0.05 ...]
13	0.30	0.54	[-0.07 ...]
14	0.33	0.56	[-0.02 ...]

Actinomycetes
Table C.2 shows the results of the principal component analysis for the biological soil indicator of the presence of actinomycetes bacteria.

Table C.2: The results from the principal component analysis for the biological soil indicator of the presence of actinomycetes bacteria. λ_i is the eigenvalue, σ_i is the spread and v_i is the eigenvector.

i	λ_i	σ_i	v_i
1	3.40	1.84	[-0.07 ...]
2	2.73	1.65	[-0.01 ...]
3	1.69	1.30	[0.10 ...]
4	1.43	1.20	[0.38 ...]
5	1.20	1.10	[0.30 ...]
6	1.11	1.05	[-0.59 ...]
7	0.73	0.86	[-0.58 ...]
8	0.58	0.76	[-0.27 ...]
9	0.00	0.02	[-0.00 ...]
10	0.04	0.20	[-0.01 ...]
11	0.13	0.36	[0.02 ...]
12	0.46	0.68	[0.03 ...]
13	0.29	0.54	[-0.08 ...]
14	0.33	0.58	[-0.04 ...]

Saprotrophs
Table C.3 shows the results of the principal component analysis for the biological soil indicator of the presence of saprotrophic fungi.

Table C.3: The results from the principal component analysis for the biological soil indicator of the presence of saprotrophic fungi. λ_i is the eigenvalue, σ_i is the spread and v_i is the eigenvector.

i	λ_i	σ_i	v_i
1	3.39	1.84	[-0.02 ...]
2	2.77	1.66	[0.15 ...]
3	1.71	1.31	[-0.18 ...]
4	1.38	1.17	[-0.11 ...]
5	1.27	1.12	[-0.44 ...]
6	0.91	0.95	[-0.39 ...]
7	0.87	0.93	[-0.74 ...]
8	0.00	0.02	[0.00 ...]
9	0.04	0.20	[-0.01 ...]
10	0.16	0.35	[0.06 ...]
11	0.59	0.77	[-0.17 ...]
12	0.46	0.68	[0.05 ...]
13	0.30	0.54	[-0.06 ...]
14	0.33	0.57	[0.09 ...]

Mycorrhiza
Table C.4 shows the results of the principal component analysis for the biological soil indicator of the presence of mycorrhiza forming fungi.

Table C.4: The results from the principal component analysis for the biological soil indicator of the presence of mycorrhiza forming fungi. λ_i is the eigenvalue, σ_i is the spread and v_i is the eigenvector.

i	λ_i	σ_i	v_i
1	3.39	1.84	[0.03 ...]
2	2.74	1.66	[0.08 ...]
3	1.68	1.30	[0.00 ...]
4	1.46	1.21	[-0.40 ...]
5	1.20	1.10	[0.24 ...]
6	0.98	0.99	[0.59 ...]
7	0.82	0.91	[-0.64 ...]
8	0.00	0.02	[0.00 ...]
9	0.41	0.20	[-0.01 ...]
10	0.13	0.36	[0.01 ...]
11	0.30	0.55	[-0.03 ...]
12	0.33	0.58	[-0.07 ...]
13	0.46	0.68	[0.06 ...]
14	0.60	0.78	[0.00 ...]

Diversity index species
Table C.5 shows the results of the principal component analysis for the biological soil indicator diversity index for species.

Table C.5: The results from the principal component analysis for the biological soil indicator diversity index of species. λ_i is the eigenvalue, σ_i is the spread and v_i is the eigenvector.

i	λ_i	σ_i	v_i
1	3.40	1.84	[0.05 ...]
2	2.73	1.65	[0.02 ...]
3	1.69	1.30	[0.07 ...]
4	1.42	1.19	[-0.38 ...]
5	1.23	1.11	[0.45 ...]
6	1.11	1.05	[-0.52 ...]
7	0.79	0.89	[0.43 ...]
8	0.53	0.73	[-0.36 ...]
9	0.46	0.66	[0.10 ...]
10	0.33	0.57	[-0.10 ...]
11	0.28	0.53	[-0.18 ...]
12	0.13	0.35	[-0.06 ...]
13	0.00	0.02	[-0.00 ...]
14	0.04	0.20	[-0.00 ...]

C.2 PCR Spain

Bacterial ratio
Table C.6 shows the results of the principal component analysis for the biological soil indicator of the fraction of $gram+/gram-$ bacteria.

Table C.6: The results from the principal component analysis for the biological soil indicator of the fraction of $gram+/gram-$ bacteria. λ_i is the eigenvalue, σ_i is the spread and v_i is the eigenvector.

i	λ_i	σ_i	v_i
1	3.86	1.96	[-0.02 ...]
2	2.79	1.67	[-0.10 ...]
3	1.37	1.17	[-0.03 ...]
4	0.00	0.02	[-0.01 ...]
5	0.06	0.24	[-0.01 ...]
6	1.42	0.38	[0.03 ...]
7	0.24	0.49	[0.03 ...]
8	0.42	0.65	[-0.09 ...]
9	0.52	0.73	[-0.06 ...]
10	1.12	1.06	[-0.75 ...]
11	1.06	1.03	[0.22 ...]
12	0.90	0.95	[-0.23 ...]
13	0.83	0.91	[-0.24 ...]
14	0.78	0.88	[0.51 ...]

Actinomycetes
Table C.7 shows the results of the principal component analysis for the biological soil indicator of the presence of actinomycetes bacteria.

Table C.7: The results from the principal component analysis for the biological soil indicator of the presence of actinomycetes bacteria. λ_i is the eigenvalue, σ_i is the spread and v_i is the eigenvector.

i	λ_i	σ_i	v_i
1	3.85	1.96	[0.03 ...]
2	2.79	1.67	[0.10 ...]
3	1.38	1.17	[0.16 ...]
4	1.08	1.04	[0.67 ...]
5	1.06	1.03	[0.40 ...]
6	1.00	0.02	[0.00 ...]
7	0.06	0.24	[0.01 ...]
8	0.14	0.38	[0.03 ...]
9	0.24	0.49	[0.06 ...]
10	0.91	0.95	[0.28 ...]
11	0.84	0.92	[0.27 ...]
12	0.78	0.88	[-0.44 ...]
13	0.42	0.65	[0.09 ...]
14	0.52	0.72	[0.01 ...]

Saprotrophs
Table C.8 shows the results of the principal component analysis for the biological soil indicator of the presence of saprotrophic fungi.

Table C.8: The results from the principal component analysis for the biological soil indicator of the presence of saprotrophic fungi. λ_i is the eigenvalue, σ_i is the spread and v_i is the eigenvector.

i	λ_i	σ_i	v_i
1	3.39	1.84	[-0.00 ...]
2	2.78	1.66	[0.03 ...]
3	1.37	1.31	[0.03 ...]
4	1.16	1.17	[-0.79 ...]
5	1.06	1.12	[0.13 ...]
6	0.93	0.95	[-0.11 ...]
7	0.83	0.93	[-0.10 ...]
8	0.73	0.02	[-0.56 ...]
9	0.52	0.20	[-0.10 ...]
10	0.42	0.35	[-0.10 ...]
11	0.40	0.77	[-0.00 ...]
12	0.58	0.68	[0.04 ...]
13	0.24	0.54	[0.01 ...]
14	0.14	0.57	[-0.01 ...]

Mycorrhiza
Table C.9 shows the results of the principal component analysis for the biological soil indicator of the presence of mycorrhiza forming fungi.

Table C.9: The results from the principal component analysis for the biological soil indicator of the presence of mycorrhiza forming fungi. λ_i is the eigenvalue, σ_i is the spread and v_i is the eigenvector.

i	λ_i	σ_i	v_i
1	3.86	1.96	[0.00 ...]
2	2.77	1.67	[0.01 ...]
3	1.39	1.18	[0.20 ...]
4	0.06	0.24	[0.06 ...]
5	0.00	0.02	[0.00 ...]
6	0.14	0.38	[-0.03 ...]
7	0.24	0.48	[0.08 ...]
8	0.43	0.65	[-0.06 ...]
9	0.52	0.72	[-0.11 ...]
10	0.11	1.07	[-0.74 ...]
11	1.07	1.03	[0.08 ...]
12	0.87	0.93	[-0.10 ...]
13	0.80	0.90	[0.60 ...]
14	0.82	0.90	[0.03 ...]

Diversity index species
Table C.10 shows the results of the principal component analysis for the biological soil indicator diversity index for species.

Table C.10: The results from the principal component analysis for the biological soil indicator diversity index of species. λ_i is the eigenvalue, σ_i is the spread and v_i is the eigenvector.

i	λ_i	σ_i	v_i
1	3.87	1.97	[-0.07 ...]
2	2.77	1.67	[0.02 ...]
3	1.37	1.17	[-0.03 ...]
4	1.09	1.04	[-0.71 ...]
5	1.06	1.03	[0.42 ...]
6	0.90	0.95	[0.37 ...]
7	0.86	0.93	[-0.25 ...]
8	0.80	0.89	[0.29 ...]
9	0.52	0.72	[-0.03 ...]
10	0.41	0.64	[-0.17 ...]
11	0.24	0.49	[-0.05 ...]
12	0.14	0.37	[-0.07 ...]
13	0.06	0.24	[0.00 ...]
14	0.00	0.02	[0.00 ...]

DEPARTMENT OF TECHNOLOGY MANAGEMENT AND ECONOMICS
DIVISION OF ENVIRONMENTAL SYSTEMS ANALYSIS
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY