MSc Computer Science
Final Project

From multidimensional soil properties to reduced summary scoring curves for soil health

Ruxandra Simioniuc

Supervisor: Dr. Doina Bucur
Committee member: Dr. Ir. Pieter-Tjerk de Boer

October, 2024

Department of Computer Science
Faculty of Electrical Engineering,
Mathematics and Computer Science,
University of Twente

UNIVERSITY OF TWENTE.

Contents

Abstract

The world's soils are one of the most important base resources, providing food and supporting most life forms. Lately, the degradation of the pedosphere has been a focus point for restoration plans and legislation, which aim to improve the health of the soils, a concept still evolving as more research is done. The study of the soil is a multidisciplinary field, spanning domains such as ecology, biology and chemistry. Some frameworks for evaluating soil health have already been proposed. However, most approaches use basic statistical techniques, are not completely analytical as they rely on field knowledge, are hard to generalise or do not provide a start-to-end pipeline. We propose a framework with clear steps and new approaches, which has the potential to be generally applicable. The most frequent soil health indicators for chemical, physical, and biological aspects are selected from relevant literature and aligned with the available dataset. The samples are split based on soil texture type and processed separately. Redundant features are discarded using correlation analysis and the remaining subsets are fed into sparse principal component analysis (SPCA). This dimensionality reduction technique has not been utilised in other related research, leading to a 90% reduction in the number of indicators. **Scoring curves** for individual indicators are obtained from literature, adapted to the data, and combined into **summary indicators**, using the weights and components obtained through SPCA. Summary scoring curves are computed, reducing the evaluation of each pillar to only 2 dimensions, while retaining over 75% of the explained variability for the biological and physical pillars. The final step is providing a simple, intuitive and interpretable visualisation of the results, so stakeholders can easily interpret the health of a soil sample.

Keywords: soil health, dimensionality reduction techniques, Sparse PCA, scoring functions, summary indicators

Chapter 1

Introduction

What is soil and why does a resource that seems to be in abundance need a close-up analysis? With few exceptions, the world's land biomes are covered with a layer of soil essential for many life forms and the completion of environmental functions. Although it might appear there is enough to spare, just one centimetre of soil can take hundreds of years to form [1] and can disappear in a matter of days or even hours due to erosion or heavy rainfall. Not only is soil the most important medium for growing food, but it also harbours the most biodiverse and complex ecosystem and protects against climate change.

1.1 Research context

The experts at the Netherlands Institute of Ecology (NIOO-KNAW) study various aspects of the soil, including its composition and how it affects plant growth, among other factors. They highlight the need to understand and evaluate the state of soils to preserve the natural lands and habitats and provide relevant information to farmers for promoting sustainable agricultural practices. Their expertise is crucial in this research, as they offer valuable guidance in understanding the concepts, factors and associated connections, while also providing the data for analysis.

The current state of soils

The European Union's Green Deal from 2020 [2] heavily accentuates the need to restore the continent's soils, therefore the EU Soil Strategy for 2030 [3] was presented in 2021, giving an extensive context and arguments for the immediate need to preserve this resource. As a direct successor to this strategy, in July 2023 the proposal for a Soil Monitoring Law [4] was adopted as a part of attaining the European Green Deal's goals. This law outlines the importance of conserving Europe's soils, as their advanced state of degradation threatens the agriculture industry as well as the related ecosystems. It also imposes certain criteria and offers guidance for member states of the European Union on how to assess the degradation of soil.

The high demand for resources causes the cultivated land to constantly be under agricultural use and, implicitly, be subjected to practices meant to improve the overall production, such as tilling, ploughing, irrigation and drainage [5]. All of these interventions on the soil are known to cause changes in its chemical and physical structure [5], which can lead to a decrease in the soil's ability to perform its functions. The Soil Monitoring Law [4] estimates that, in the European Union, 60 to 70% of soils are not healthy. A similar percentage is "affected by erosion, loss of organic carbon, nutrient exceedances,

compaction or secondary salinisation (or a combination of these threats)", whereas a fifth of EU's agricultural lands exceeds the limit for cadmium concentrations. Over 20% of the EU's land has a high degree of subsoil density, which is a good indicator for compaction and a quarter of the area in Southern, Central and Eastern Europe was at high or very high risk of desertification in 2017 [1].

The concept of soil health

The EU's Soil Monitoring Law [4] defines healthy soil as *the physical, chemical and biological condition of the soil determining its capacity to function as a vital living system and to provide ecosystem services.* There is no standard definition across literature, but the main points are the same, with slight variations in terminology or how explicit the authors choose to be — sometimes mentioning the services ("sustain plant and animal productivity, maintain or enhance water and air quality, and promote plant and animal health" [6]) or focusing slightly more on the usage of land "key to ecosystem function within land-use" [7].

The concept of "soil health" is not a novel trend, but throughout the years it has been used interchangeably with terms such as "soil quality" or "soil fertility". Lehmann et al. [8] make a clear distinction between the concepts, from the narrowest to the broadest: fertility only considers crop production, and quality goes further, taking into account not only agriculture but also the environmental services that usually are directly connected to humans, but both terms predominantly focus on the impact on agroindustry. The authors believe that soil health extends to a planetary level, beyond humanity's needs and also consider the most general term, *soil security*, which regards soil as a common good and access to its services as important as any other human right.

The existence of subtle differences in definition also translates into the fact that the assessment methods also vary and shift focus according to the authors' interest in specific problems. Therefore, by also considering the high complexity of the soil and the influence it has on countless aspects of the environment, relevant literature has yet to reach a consensus on a standard set of features and limits that can indicate the soil state beyond any doubt while also considering all of its functions. Generally, soil health is assessed by measuring various physical, chemical and biological features, although the latter is usually comprised of aggregated measurements and does not consider individual microorganisms that reside in the soil.

1.2 Research goals

Soils are vital in most ecosystems and host a considerable proportion of life forms. They have direct and indirect effects it has on human life and food security and, if in good condition, can help achieve several of UN's Sustainable Development Goals [9]. Given these processes and the standing legislation, preserving the health of the soils and their ability to fulfil essential functions is imperative. However, the concept of *soil health* and establishing criteria and limits for evaluation have not yet been universally set and are still fluctuating or relative to many factors. Being such a complex structure, one major impediment is the intrinsic interdisciplinarity of the study of soil [10], as it requires knowledge of biology, chemistry, geology and physics, among others.

A framework which tackles the problem of soil health should be able to quantify the healthiness of a soil sample, as proposed legislation and plans require [3, 4]. A soil sample is characterized by many properties and various individual indicators, different subsets of

them being listed in the relevant literature (most common are organic matter, pH, some chemical elements, and soil structure; microbial biomass and magnesium are less frequent). This makes the concept of soil health highly multidimensional, with a significant amount of features to consider, which need to be condensed into summary indicators. Each relevant indicator needs to be assigned a score or a grade which can accurately indicate the health of the soil. Obtaining a score should be based on a scoring curve rather than a hard threshold, as health is more of a spectrum rather than a "healthy / not healthy" problem, comparable to human health (having a mild cold does not mean a person is immediately unhealthy, just that there are actions to take for their state to get better). These scoring curves are based on mathematical models which are specific to each indicator, according to how they influence certain aspects of the soil.

Another factor which is to be considered when computing the score of an indicator is the context of the soil sample. Climate, location, texture type and other factors influence the shape of the curve and, implicitly, the final score. Two separate samples could have the same measurements of indicator X, but if any of the other constraints tied to this specific measurement differ, the resulting scores will also be distinct. To incorporate these details, the result should be a flexible, parametric framework, which can handle all of the different characteristics of a sample. These points make the development of summary indicators for soil health a complex problem.

Evaluating and quantifying soil health requires an approach that considers the numerous quantitative soil properties and obtains an individual scoring curve for each property. The dimensionality issue should also be attenuated, by removing redundant features and combining the remaining ones based on real data in real contexts. Ultimately, the framework should provide an accessible and interpretable visualisation of the summary scoring curves, which can express the state of the soil.

Research on this issue goes back decades, with considerably different approaches, from soil health cards which require only simple tools and the farmers' input to using sophisticated machine learning methods. Considering the previously stated requirements for an accurate and robust framework, the ones that come the closest are the SMAF (Soil Management Assessment Framework) [11] and SHAPE (Soil Health Assessment and Evaluation Protocol) [12] frameworks. They both use scoring curves and have common researchers, the latter being more recent. SMAF provides individual mathematical models for many indicators. SHAPE uses a model trained on data from the USA to assess the relevant soil characteristics and construct the scoring curve for organic carbon (the only indicator implemented at the time of writing). Both methods lack the starting (dimensionality reduction) and finishing (creating summary indicators and easy-to-use visualisation) steps, focusing solely on transforming measurements to score. Other researchers perform some basic feature selection (usually with the help of principal component analysis – explained briefly in section 2.2) and then build on top of SMAF. Some papers attempt to create summary indicators, usually by assigning equal weights to indicators, while no research was found that tried to construct a visualisation of the summary indicators.

This project aims to further advance the science of soil health by analysing existing methods and using data science techniques to analyse the soil's properties. Answering the question **How to consistently and robustly quantify soil health?**, along with implementing a widely applicable assessment framework are the main goals of this research. The research question is split into five subproblems which tackle more specific aspects:

RQ0. What soil properties are relevant to be used as individual indicators for soil health?

- This question is literature-based, requiring a thorough analysis of related papers that approach the subject of soil health or adjacent concepts. The result should reflect the measurements most frequently mentioned in relevant research publications. The extracted indicators are then matched with the available data, resulting in a subset of features that accurately describe the health of a soil sample and can be used for further analysis.

RQ1. How can related soil properties be grouped to reduce the problem of multidimensionality?

- There are various measurements which can be taken when analysing a soil sample, which leads to an extensive dataset with a high number of dimensions for each observation. The features that describe the soil's state are part of the three main pillars: physical, chemical and biological. These features are inherently linked within each category, but strong interconnections exist across pillars. Identifying the relationships and correlations across characteristics can then assist in assembling indicators that consider multiple features at once. This question can be answered using dimensionality reduction techniques or feature selection methods. Limited similar research was found in specialised literature, possibly due to the minimal overlap between soil health research and data science, or the widespread use of standalone indicators. The papers that attempt to use these techniques do not go into depth and consider only basic algorithms and methods.

RQ2. How can summary indicators be scored in a manner which expresses a soil sample's health?

- Evaluating a soil measurement is an inherently complex task as it is highly contextual, depending on factors such as the climate, location, history and scope. Some papers provide scoring curves for a limited number of individual soil properties. The aim is to adapt these mathematical models to the current measurements and aggregate the scores for the previously obtained summary indicators for each pillar.

RQ3. What is the most suitable manner of expressing the results of the assessment?

- This refers to the format in which the results should be presented to the stakeholders. Assessment frameworks are either scored based on a *healthiness* scale or labelled in a binary or progressive way (intermediate labels between *healthy* and *not healthy*). This aspect of the project depends heavily on the stakeholders' needs and opinions, ease of interpretability and relevance. Both scale and label-based methods already exist, but they are not consistent across the literature, nor do rigorous explanations back them. This is mostly due to the different forms of the assessments and their specific purpose.

RQ4. How does soil texture type influence the indicators and the scoring?

- The values of soil properties are dependent on the soil's composition (percentages of silt, clay and sand) and experts often make a distinction between different texture types. Some frameworks do consider these factors when looking at the values of indicators. The relationship between context and indicators will be analysed by separating the dataset samples based on soil texture type. The behaviour at different stages of the research will be observed and compared, potentially revealing how indicators relate to a change in soil type.

Chapter 2

Background

2.1 Ecology concepts and terminology

Given this is a multidisciplinary project, some notions and aspects must be briefly explained for the content to be easily understandable. Soil is made of three main **pillars**: physical, chemical and biological. Variations in the properties of soil affect its behaviour and qualities, but management practices or extreme weather events also have an impact on these pillars.

Soil **property** refers to any measurement which describes the behaviour and contents of a soil sample, while the term **indicator** is used for those properties which are relevant in the context of soil health.

Physical pillar

Soil can be made of differently-sized particles, usually in combinations of clay, silt, sand or stones. Based on the percentages of these particles, the soil has different characteristics and behaves in particular ways [13]. Directly connected to these particles is the concept of **aggregation** or **aggregate stability**, which measures how much the soil particles bind together and how well they resist external forces. A commonly used indicator of soil aggregate stability is the **Mean Weight Diameter** (MWD), which expresses the average size of soil aggregates. Water holding capacity and water infiltration values are determined by soil texture, high percentages of sand, as it is a bigger particle, is more permeable than clay.

Chemical pillar

The chemical properties of the pedosphere aggregate information like the **pH** and concentrations of various chemical elements and compounds (phosphorus, potassium, heavy metals etc.). The analysis of this pillar gives a detailed overview of the amount of available **nutrients** in the soil (which can be in deficit, between normal limits or in excess), as well as information about the **pollutants** present in the land.

Biological pillar

The soil organic matter (SOM) is a critical element when it comes to the soil's function of producing biomass. It is a feature that is difficult to measure, therefore the soil organic carbon (**SOC** or **OC**) is usually considered when assessing the organic composition. A

high level of SOM will prevent pollutants such as heavy metals and organic toxins from infiltrating the groundwater [8].

Management characteristics

The management features refer to the ways humans interact with soil during farming or related activities. Important aspects include the types of crops planted, if and how deep tillage was performed, the fertilisers applied, the diversity and intensity of crop cultivation, and the presence of cover crops. Cover crops, in particular, play a key role in protecting the soil by managing erosion and enhancing soil quality.

Soil texture

As mentioned in the description of the physical pillar, soil can be made up of different types of particles. Conventionally, these particles are sand, silt and clay and, depending on the percentages for each of them, the soil can be classified as having a certain texture. Figure 2.1 illustrates how soil texture classes are determined. Texture influences the majority of the other soil properties, from the capacity to hold water to optimal levels of micronutrients and microorganisms.

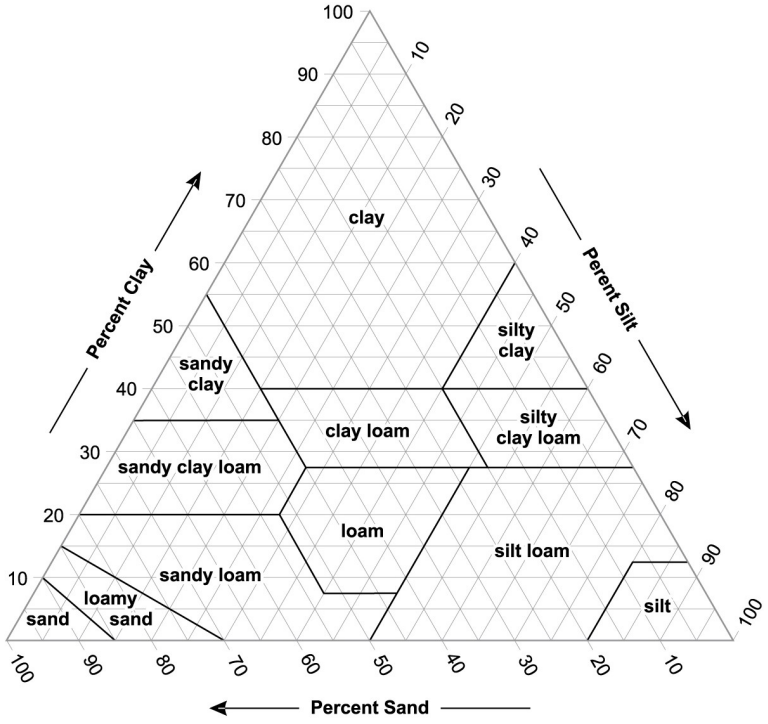


FIGURE 2.1: Soil texture triangle

Soil functions

Soil functions are defined as collections of soil processes which catalyse ecosystem services. Table 2.1 contains the 5 important soil functions mentioned in the European Environment Agency's report on soil monitoring from 2023 [14].

Soil Function	Description
Production	Capacity to produce biomass
Water storage and quality	Capacity to store precipitation water and filter for soil pollutants
Carbon storage	Capacity to store and stabilise SOC
Nutrient cycling	i. Capacity to provide nutrients from mineral and organic soil resources in available form(nutrient mobilisation capacity) ii. The capacity to store mobile nutrients within the root zone to avoid losses by leaching and gaseous emissions (nutrient buffering capacity)
Habitat for biological activity	Provision of a species (gene) pool that can buffer ecosystem functions against species extinction (assumption: loss of soil function is more likely with low species diversity in each functional group)

TABLE 2.1: Main soil functions. Table retrieved from the *Soil monitoring in Europe — Indicators and thresholds for soil health assessments* report [14].

To better understand the importance of the soil as a whole and build more context regarding its roles and current legislation, some more extensive information is provided in the following subsections.

Soil as a resource

The continuous population growth requires a proportional increase in resource production, which entails improving the crop yield and the rate at which it can be obtained. At a global level, the average wheat harvest per hectare has more than tripled from 1961 to 2021, while the total world population grew 2.5 times. The global land used for agricultural practices is estimated to have claimed approximately 330 million more hectares in the same 60-year period, which amounts to slightly more than the area of India. Still, land use by area peaked in the early 2000s, and ever since a negative trend has followed[1]. This tendency to reduce agricultural land, as opposed to increasing yield and population, suggests that numerous effective agricultural advancements have been made and are still being developed. The crops are genetically superior to their predecessors from years ago [15], and various enhancers such as synthetic fertilisers, pesticides, fungicides, and antibiotics for livestock are also heavily utilised. These agrochemicals, along with the antibiotics which ultimately end up in the ground, alter the structure of the soil, leading to homogenisation [16], infiltrate the groundwater and can have significant effects on the soil biota, which can lead to a decline in biodiversity [5].

The Food and Agriculture Organization of the United Nations estimates that 95% of the food is produced directly or indirectly on soils [17]. Roughly 45% of the Earth's habitable land is allocated for agriculture, an area which spans more than the combined surface of both Americas and which supplies animal products, feed, food and non-food crops such as biofuel and cotton [18].

[1]Data on crop yield, population and land use was summarised from https://ourworldindata.org/. Accessed in February 2024.

Soil as a biodiversity pool

More than 40% of terrestrial organisms are directly associated with soils at some point in their life cycle [19]. The Earth's soils amount to one of the most biologically diverse habitats, with one teaspoon of soil holding more living organisms than the total human population [19]. As a quantitative reference, 1 hectare of land can contain up to 500 kg of bacterial carbon and just 1 gram can reach 1 kilometre of fungi networks [14]. All organisms that reside in soil, along with the structural and chemical composition of the soil influence its capacity to perform vital ecosystem functions, such as absorbing and filtering water, storing carbon, serving as an environment for plant growth (also known as biomass) and cycling nutrients [3]. All of these processes are the foundation of the economy, society and environment [4].

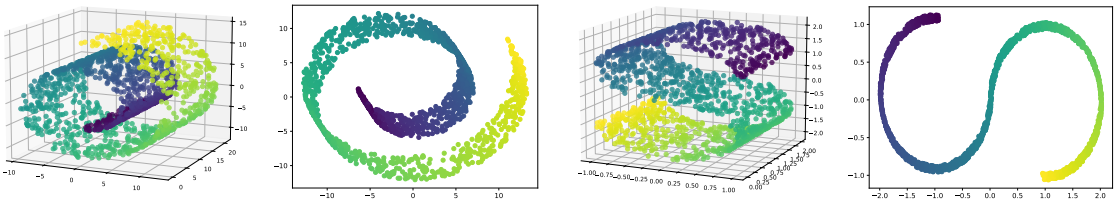
Soil as a climate change regulator

Healthy soils play a crucial role in maintaining a stable natural environment and can help combat extreme weather phenomena, as they offer protection against droughts and floods and resilience to climate change [4]. Certain strains of bacteria (corroborating the importance of biodiversity) can help plants be more tolerant to drought and some management practices can lead to more CO_2 being stored in the soil, which reduces the gas's greenhouse effect [4]. On the other hand, sealed soils cannot absorb as many pollutants as their healthy counterparts and can maintain high temperatures for longer in the case of heat waves [14].

The ongoing climate change affects the planet on various levels, threatening not only the stability of the world's ecosystems but also the prosperity of human societies and the health of humans themselves. According to the State of Global Climate report from the World Meteorological Organisation, the year 2023 was the warmest since 1850, going approximately 1.45 °C over the average temperature of the pre-industrial era [20]. Given the urgency to preserve and, at least to some degree, restore the environment we live in, strategies, legislation and plans have been put into place in the past years at international, regional and local levels alike. The Paris Agreement signed in 2015 by 196 parties [21], the European Green Deal from 2020 [2], and the USA's Federal Sustainability Plan from 2021 [22] are some of the most extensive plans for stopping the drastic temperature rise and reducing carbon emissions. Other industry-rich states have also set targets to achieve carbon neutrality (China, Japan) or net-zero emissions (Canada, South Korea). Following through with these action plans is extremely relevant for the soils of the world, as the biodiversity is affected by global climate changes [16] and can be radically reduced due to higher concentrations of CO_2 in the air [5]. Moreover, changes in precipitation and temperatures can lead to desertification, which, in turn, implies severe degradation of the soil.

2.2 Data Science methods

Dimensionality reduction techniques (DRT) are used extensively in the machine learning field. The aim is to "find lower dimensional data representations that preserve their key properties for a given problem" [23]. Essentially, the dataset is assessed and transformed to compress the number of features that describe it, without losing a considerable amount of the initial information. Figure 2.2 exemplifies how a 3D representation of two point groups is converted to only two dimensions. While some information is lost (the depth axis), the general shape of both datasets is retained.



(A) Swiss roll (B) Swiss roll reduced to 2D (C) S-curve (D) S-curve reduced to 2D

FIGURE 2.2: Two examples of reducing the representation of data points from 3 dimensions to only 2 using PCA. The Swiss Roll and S-curve datasets from the `scikit-learn` library are used.

One of the most popular DRTs is **Principal Component Analysis** (PCA), an unsupervised learning algorithm. PCA uses linear algebra and statistics methods to reduce the dimensionality of large datasets. It transforms the data into orthogonal principal components that capture the most variance. These components are linear combinations of the original features, with weights indicating the contribution of each feature (higher weight = more information). PCA flattens the data while retaining the most critical information by focusing on the components that explain the most variance. The weights associated with each feature are called **loading vectors** or **loadings** and a **principal component** (PC) is the new dimension and consists of the linear combination of the original features, based on the loading vectors. Each PC is a weighted sum of the original features.

Let X be the original data in matrix form of n x p dimensions, with n observations of p variables. To reduce the data from p to k dimensions, with $k < p$, the PCA algorithm has following steps:

- Scale and centre the data – each variable has $\mu = 0$ and $\sigma = 1$ by subtracting its mean
- Compute the covariance matrix
- Extract eigenvectors and eigenvalues
- Sort eigenvalues in decreasing order
- Select k eigenvectors with the highest eigenvalues

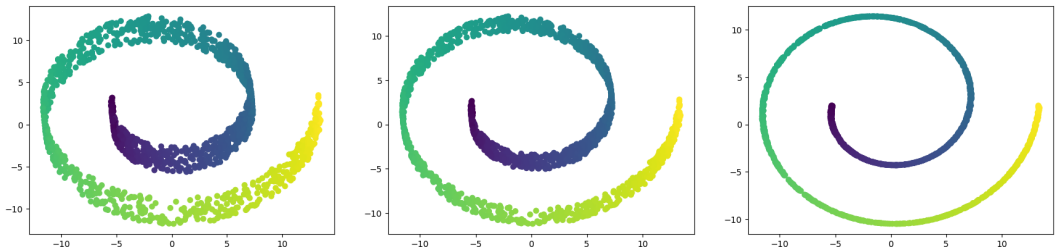
The resulting eigenvectors $v_1, ...v_k$ form the loading matrix V with p x k dimensions. Transforming the data to the new subspace is done according to equation 2.1. $Z = [z_1, ..., z_k]$ is the principal component matrix with a shape of n x k. Essentially, Z is the final product of the algorithm which compresses the initial dataset to a lower dimensional space.

$$Z = XV^T \tag{2.1}$$

Another method for computing PCA exists, which consists of computing the singular value decomposition (SVD) of the data as $X = UDV^T$. In this case, V is still the loading matrix and the principal components are computed from $Z = UD$.

Multiple variants extending the classic PCA have been developed. One alternative to PCA that is used in this research is the **Sparse Principal Component Analysis** (SPCA). The classic PCA assigns nonzero weights to all features in all principal components, whereas in SPCA some of the loadings are equal to zero. This leads to a considerably simpler representation of the compressed data and it is easier to deduce which initial attributes influence the differences between samples [24]. **Kernel Principal Component Analysis** (KPCA) is a non-linear variant of PCA which can be more effective in detecting and separating clusters. A general downside to this method is choosing the right kernel, which needs a certain level of understanding of the shape and structure of the data, besides multiple test cases. Also, KPCA does not provide the combinations for each resulting principal component, as it is mainly used for visualisation purposes.

Sparse PCA is extensively used in this research due to its ability to produce simpler principal components. This increases the interpretability of the variables, an important aspect for the stakeholders, while also minimising the number of features which make up a PC, drastically reducing the complexity of the assessment. By increasing the sparseness level of the algorithm, the PCs depend on fewer variables, but the overall explained variance is diminished.



(A) $\alpha = 1$ (B) $\alpha = 3$ (C) $\alpha = 5$

FIGURE 2.3: The impact of the sparseness level (α) in SPCA applied on the Swiss Roll dataset. The `scikit-learn` Python library was used to generate this example.

Figure 2.3 illustrates how different sparseness levels affect the algorithm's output. A higher value for α yields a simple representation of the transformed data, easy to interpret and describe. However, certain details which contribute to describing the dataset are lost in the process. Less sparseness ($\alpha = 1, 3$) retains more information but creates more intricate principal components. Finding a trade-off between the explained variance and the sparseness of the low-dimensional data is a necessary step to ensure the results are appropriate to the specific use case.

Chapter 3

Related Work

What makes a good indicator

As a generalised rule which is also endorsed by other scientists ([8]), Doran and Zeiss [6] argue that a soil health indicator should be **sensitive** enough to management practices and climate, well **correlated** with soil functions, **informative** in determining why and how the soil performs its functions and **easy** and **inexpensive** to measure.

Specialised literature tends to be more and more in favour of paying more attention to the **biological pillar** when it comes to soil health. Since the primary use of soil in the world's industry is agricultural, Lehmann et al. [8] note that considerable attention has been directed towards conserving those features that sustain related practices. The authors recognise that most assessment schemes are dominated by chemical indicators and argue that biological characteristics should play a bigger role in quantifying soil health and propose the integration of pathogens, parasites, and bioavailable and mobile toxins in soil health assessment schemes.

Banerjee and van der Heijden [16] also underline the importance of microorganisms by placing the same evaluation in the context of *one health*, tying soil to the well-being of animals, people and the environment as a group. The authors also note that soil health bioindicators usually use only broad criteria, such as microbial mass. Ulea et al. [25] claim that seasonality and land management have a noticeable influence on bacterial density and diversity, which impact soil fertility indirectly. Bruggen and Semenov [26] state that disease suppression is an important factor in soil health, a characteristic promoted by biodiversity, among other factors. The authors mention that microbial diversity is usually quantified as the number of present species and not from a structural and functional point of view and suggest that analysing microbial consortia might offer more insight than looking only at individual microorganisms.

Doran and Zeiss [6] also regard the organisms which reside in the soil as being essential to performing soil and ecosystem functions. They propose adapting already existent frameworks for measuring sustainability in farming by adding features needed for environmental conservation. However, they also note that measuring soil organisms results in a sensitive indicator and correlated with essential soil processes, but is hard and expensive to measure and does not offer immediately relevant information for management practices or how it influences other services. Since it is hard to have indicators which comply with all criteria, it is recommended first to evaluate the objectives of the analysis and then determine the utility of an indicator. De Vries et al. [27] endorse biota's vital role in soil's services and show that a minimum level of biodiversity is required for sufficient carbon and nitrogen cycling.

On the opposite side of the argument, while not undermining the importance of the biota, Coyne et al. [7] stress that our understanding of *soil health* should not be dominated by biological characteristics. The counterargument comes from the inability to understand how the microorganisms behave in situ fully (at least for now), the inherent link between the biota and the chemical and physical factors and that microfauna has a lower impact on plants and animals than the higher-level soil structure does.

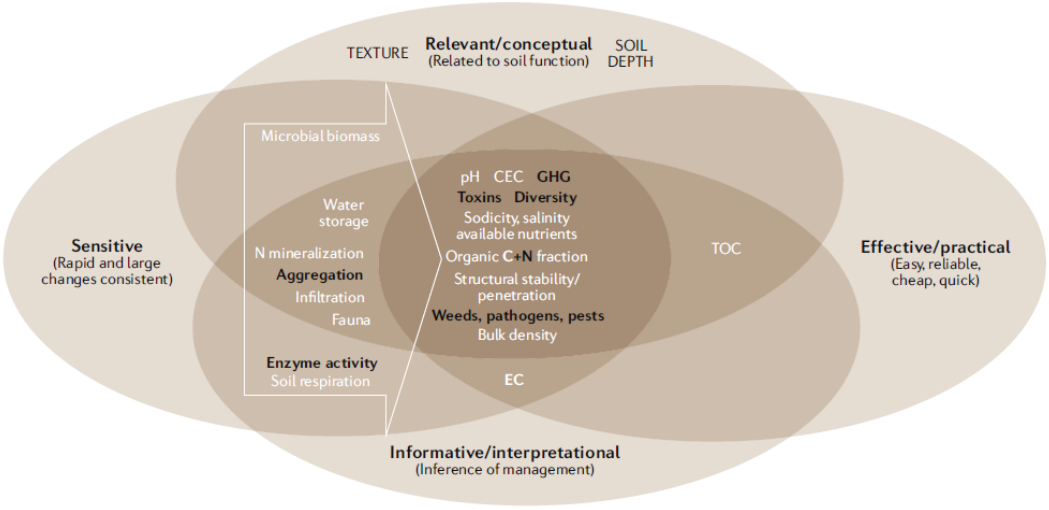


FIGURE 3.1: Soil health indicators and the requirements from [6] with which they comply. Image extracted from [8]. The white arrow contains the indicators which should be developed to be effective and practical. C, carbon; CEC, cation-exchange capacity; EC, electrical conductivity; GHG, greenhouse gases; N, nitrogen; TOC, total organic carbon

Figure 3.1 is extracted from Lehman et al. [8] and illustrates how the soil properties deemed relevant by the authors comply with the base requirements of an indicator.

Existing frameworks for health assessment

One popular practice in assessing the land status is the use of **soil health cards**, which are usually meant to be filled in by farmers based on their experience and knowledge of the environment and the land they manage. The state of Maryland provides a card [28] with 7 indicators, where farmers can score their lands upon a close inspection, without specialised tools, and classify their soil as *excellent, good, fair* or *poor*. The card provided by the Natural Resource Conservation Service of the U.S. [29] requires information about management practices. It is destined to be used by conservation planners and aims to offer insights into the state of degradation of the soils, while also acknowledging that it is not a comprehensive analysis of all soil pillars. The Montana Soil Health Card [30] splits the analysis into 2 tiers, the first being executed on-site in 30 minutes with minimal equipment, whereas the second tier offers more in-depth information by processing soil samples in a laboratory.

With the increased interest in the soil states, **commercial services** that evaluate and score the soil have also appeared on the market. The Solvita Soil Tests [31] is a service that provides kits to evaluate soil and tests performed in their associated laboratories. The samples are scored relatively to the local expectations and use only 5 indicators. The

Haney Soil Test [32] is a laboratory procedure where a multitude of chemical elements are extracted and, by looking at the organic carbon and nitrogen in the soil, the sample is scored from 0 to 50, 7 being good. Cornell University developed the CASH (Comprehensive Assessment of Soil Health) test [13], looking into 12 biological, physical and chemical properties and scoring each from 1 to 100, the average of all being the overall score of soil health. The scores are given relative to other analysed samples from the same region.

On a **European level**, The European Environment Agency (EEA) published a report in early 2023 [14] recommending intervals for certain soil characteristics, such as pH and SOC, while also considering the importance of biodiversity in the cycling process of carbon and nitrogen. EU's Soil Monitoring Law [4] provides an extensive list of criteria which should be satisfied for soils to be assessed. The list contains 14 indicators, physical, chemical and biological, some of which have limits imposed at the Union's level, while some thresholds are left for the member states to set. The law also offers a set of principles that should be applied when considering land management.

Organic matter is an essential component in soils and deterioration quickly follows if SOC levels drop below 1%, while at least 2% concentration is needed for maintaining aggregate stability [14]. Fowler et al. [33] use spatial patterns to reveal that yield stability and volume can be used to determine SOC levels in the ground. Unstable zones and stable and high zones correlate with more organic carbon, while low and stable yield zones have lower SOC. The authors confirmed the assumption that higher-yielding areas are associated with increased SOC levels and higher soil health scores (scores provided by Solvita).

Wade et al. [34] propose a quantitative approach to evaluating soil health. Current frameworks look at how management changes indicators and then derive the soil health score. The authors propose a different method, where management practices change the soil health, which is reflected in the indicators' values. Overall, they propose analysing soil health from the perspective of the soil's scope.

With a more **analytical strategy**, Andrews et al. (2004) [11] implement the *Soil Management Assessment Framework* (SMAF). Their framework selects a minimum dataset (MDS) of indicators based on the site's management goal and soil function then assigns a nonlinear scoring curve to every indicator. Each of these functions is fine-tuned with parameters which depend on other factors (climate, temperature or texture, for example) The last step consists of integrating all individual scores into a final index value, yielding a number between 1 and 10. The individual scoring functions for each indicator are one of three types: more-is-better, less-is-better and mid-point optimum range. Expanded on top of CASH and SMAF, the *Soil Health Assessment Protocol and Evaluation* (SHAPE) was published in 2021 by Nunes et al. [12]. Using over 14.000 observations from the United States to create a cumulative distribution function (CDF), the authors acquire a scoring curve which accounts for the sample's origin (region), texture type, mean annual temperature and precipitation. They focus mainly on scoring the organic carbon content but argue their framework can be scaled and adapted to any other soil property.

Using scoring curves instead of hard limits is also documented in specialised literature, with some researchers applying and further developing SMAF to specific case studies (Brazilian mangroves [35], oxisoils [36], sugarcane fields [37], Italian sites [38]). Where SMAF lacks explicit functions for evaluating a soil property, an alternative used by researchers (or as a stand-alone framework) [38, 39, 40] to obtain these three types of curves is gathering the values for the lower (L) and upper (U) thresholds, as well as the optimum (only for a few indicators, such as the pH).

The idea of reducing the amount of soil properties to analyse (a minimum data set -

MDS) has been approached by other researchers as well. Sparling and Schipper [41] create an MDS after removing a portion of soil properties by computing the correlation between the measurements, using their expertise and evaluating the ease of acquiring an indicator. Their research focuses on the sites of New Zeeland and how land use and soil type influence the measurements. They conclude that 8 indicators explain 87% of the total variability of a sample. As for other methods of obtaining an MDS, the review from Bunemann et al. [42] mentions using principal component analysis, clustering techniques and decision trees in literature to remove redundant features. Another popular method of selecting relevant indicators with an analytical technique is the use of principal component analysis (PCA), where researchers select the indicators with the highest weights in each principal component (or use a threshold for the weights) [43, 44, 45].

The consensus seems to be that creating a widely applicable framework is a highly complex task, especially as multiple advancements and research are still being made. Scientists are still in the process of understanding how the biota acts and reacts and the connections between organisms and the chemistry of the soil are still being studied. Most of the available frameworks and existing methods of scoring soil health do not analyse the diversity of microorganisms, their biological pillar is limited to counting the earthworms and measuring the soil's respiration and its organic matter content.

Most of the available literature does not propose any innovative way of analysing soil health. They argue in favour or against considering certain features, sometimes providing criteria such as "high", "low" or "normal" as to what the values of the indicators should be. **Still, hard limits, consistent approaches or scoring curves for many indicators are hard to come across**. One reason for this is the **high diversity and multidimensionality of soil features**, being **dependent on climate, use, location and history**, which can be also seen in the Solvita, CASH and SHAPE scoring practices. Overall, the field lacks a **start-to-end framework** which is analytical, easy to understand and can easily adapt to specific needs and sites.

Chapter 4

Methodology

4.1 Data

The NIOO-KNAW institute provides the VitalSoils dataset for this project, formatted as `csv` and `xlsx` files [46, 47]. A total of 225 soil samples from farmlands have been collected from different regions of The Netherlands and analysed for relevant features. The samples are classified according to the management practices they underwent, comprising a total of 114 entries labelled as conventional and 111 entries as organic[1]. Another relevant discriminant is the **soil texture**, with 144 samples labelled as **marine clay** and 81 as **sand**.

The dataset is composed of abiotic characteristics associated with each sample. Chemical measurements, physical structures and summarised biological attributes describe each soil sample. The management practices for agricultural lands are also features in the dataset, encompassing crop types and methods used to cultivate the area. Additionally, considerably bigger files describe the biotic component of the samples, containing counts for various types of organisms which reside in the soil (bacteria, fungi, nematodes, protists). These files play no part in this research and do not overlap with the biological pillar (as it is referred to in this study).

The dataset is comprised of roughly 150 columns and contains features that were determined by interviewing the farmers (farming practices, what crops were planted, what and if cover crops were present) or by analysing the soil samples in a laboratory - either one of the institute's or the Eurofins laboratory – and shows information about the pH, concentration of heavy metals, various chemical components, organic matter composition, among others. Some columns may depict the same measurement (`pH` and `pH_Eurofins`) or highly similar ones (`K_Eurofins, K.vrd_Eurofins, K.bez_Eurofins` – all represent various forms of potassium), based on what laboratory performed the measurement, the method or element availability in the soil.

[1] In the EU, organic agriculture prohibits the use of genetically modified organisms, limits the use of artificial fertilisers, herbicides and pesticides and promotes techniques like crop rotation and natural pest control - https://agriculture.ec.europa.eu/farming/organic-farming/organic-production-and-products_en

4.2 Approach

RQ0. What soil properties are relevant to be used as individual indicators for soil health?

There is no consensus over what subset of features can accurately describe the health of a soil sample, therefore the starting point of this research is rather vague. To answer this question, related literature needs to be gathered and analysed to extract the relevant indicators. Indicators that are more frequent across specialised literature are considered meaningful in predicting soil health and will be tried to be matched with the columns of the current dataset. The resulting columns will comprise the working set for further analysis.

The initial step in obtaining the set of indicators is manually selecting the relevant columns from the dataset based on the most frequent indicators in the specialised literature. As mentioned in section 4.1, the samples were analysed in two separate laboratories, therefore some columns may represent the same or highly similar indicators. This selection does not discriminate over this similarity and considers all corresponding columns. After finalising this process, the number of missing values for each column is accounted for and removed if too many `NaNs` are found.

RQ1. How can related soil properties be grouped to reduce the problem of multidimensionality?

The number of available soil samples is rather low. However, each sample has 148 features containing information about the field's origin and management, concentrations of chemical elements and other descriptors. Measurements for individual species residing in the soil are not included, only summaries on certain biological aspects. This research question aims to produce a simple and concise manner of expressing the soil features by converting them to a lower-dimensional space.

The main points in answering this question are eliminating redundant indicators, applying and evaluating dimensionality reduction techniques and obtaining summary indicators for each pillar. We define a **summary indicator** as a weighted sum of individual features.

Feature selection — Obtaining a minimum dataset

After obtaining the set of columns for each indicator, redundant columns can be removed further by computing the Pearson correlation coefficient (r) between the indicators of a certain pillar. Either positive or negative, a high correlation coefficient between two features is considered an indication that they model the same behaviour and it would be ineffective to use both in the following stages of analysis. Additionally, the data is scaled and hierarchically clustered using Ward's method, which assumes an Euclidean distance between points and aims to minimise the within-cluster variance [48], ultimately revealing how closely related the indicators are. The result is plotted as a dendrogram, which aids in validating the positively correlated features.

If two indicators are strongly correlated, one is deemed redundant and removed by manually analysing the meaning, importance, ease of measurement and number of missing values. Cross-pillar correlation is not computed, as feature elimination within a pillar is the least destructive, whereas looking for associations between the three subsets can potentially lead to losing important indicators. Moreover, if two columns within the same pillar are strongly associated, it is highly probable that they are related and not just coincidentally correlated. Even if the pillars are known to influence each other, the extent to which this

effect takes place is not extensively explained and understood, thus features cannot be confidently removed based on cross-pillar associations.

Reducing data dimensionality

After these operations have been completed, the final set of features for each of the three pillars is expected to downsize significantly compared to the initial dataset. Dimensionality reduction techniques (DRT) can now be applied more efficiently to these compact collections of indicators. Until this step, all samples are used to create a minimum dataset (MDS), assuring general applicability. The next stages of analysis will make a distinction based on the type of soil texture (M – marine clay or S – sand). The texture is of high importance when analysing soils from various points of view, as they have different behaviours based on their particle composition. The data is scaled and centred before applying any DRT, as the algorithms require the input to have a mean of 0 and a variance of 1 across a feature column.

Both linear and non-linear approaches are used and compared, first for each separate soil pillar, then combined. The linear methods applied to the data are Principal Component Analysis (PCA) and variations of it. These algorithms are applied using the implementations from the `scikit-learn` library. PCA, Incremental PCA (IPCA) and Sparse PCA (SPCA) provide the coefficients of each feature for each resulting principal component. This aspect is crucial in this research question, as it provides an answer which can be interpreted and explained. Kernel PCA (KPCA) with various kernels (polynomial of degree 2 and 3, RBF, sigmoid and cosine) is the non-linear method used to obtain the principal components from the data. The coefficients cannot be extracted from these algorithms, but they serve as a comparison for visualisation and evaluation metrics.

Evaluation

The evaluation of DRTs is not a straightforward process. It is highly relative to the nature of the data, the objective of the analysis and the domain knowledge of the assessor. In literature, these techniques are evaluated on runtime, visual analysis based on cluster formation and retention after applying the method on a labelled dataset [49, 50, 51] or by comparing how a classifier performs on both low-dimensional and high-dimensional data [52, 53]. In this specific use case, the number of samples is limited enough for the required time and resources to be negligible. The dataset is unlabelled and clustering at this particular step is not the aim, therefore these methods cannot be used for evaluation.

Evaluating, comparing and selecting suitable algorithms to solve this problem of high dimensionality requires other metrics. The main objective is achieving explainability, interpretability, and sparseness, while also retaining as much information from the initial collection. Therefore, the metrics deemed relevant are R^2 and *trustworthiness*. Additionally, visual analysis of the embeddings is important in ensuring the projection is reasonably spread across the plane, verifying cluster formation and tuning hyperparameters.

$$F = \frac{tr(\hat{P}\hat{T}^T\hat{T}\hat{P}^T)}{tr(X^TX)} \tag{4.1}$$

The total fraction of explained variance is a metric which accounts for how much of the original information is retained in the reduced-dimensional space. Using formula 4.1 [54], this measure can be computed for PCA, IPCA and SPCA. In this equation, $\hat{P}$ represents the matrix of loadings, $\hat{T}$ is the transformed data in the low dimensional space (scores) and X is the initial data (with the mention that X is centred). However, since this equation

requires the loadings, it cannot be applied to KPCA. Thus, this measure is replaced by the R^2 score (coefficient of determination), which is conventionally used in evaluating the performance of a regression model. In the current problem context, this metric is applied to the original (scaled) data and the representation obtained from the inverse transform of the embedding. With a maximum value of 1, indicating perfect reconstruction, this function indicates how information can be reproduced after transformation. The R^2 score is also regarded as a measure of how much variance is explained [55] and can be applied to all of the used algorithms.

Trustworthiness and *continuity* [56, 57, 58] are rank-based measures of how well the data structure is retained and can be applied to DRTs regardless of the underlying process. Trustworthiness, as seen in formula 4.2, decreases if the neighbouring points in the projection are further apart in the original data, whereas continuity (formula 4.3) is lower if originally neighbouring points are embedded far away from each other. N is the number of samples, k the number of neighbours, $r(i,j)$ the rank of the distance between point i to point j in the high dimensional space $U_k(i)$, and $\hat{r}(i,j)$ is the ranking between i and j in the lower dimensional space $V_k(i)$. These scores are normalised and have values between 0 and 1.

$$T(k) = 1 - \frac{2}{Nk(2N - 3k - 1)} \sum_{i=1}^{N} \sum_{j \in U_k(i)} (r(i,j) - k) \tag{4.2}$$

$$C(k) = 1 - \frac{2}{Nk(2N - 3k - 1)} \sum_{i=1}^{N} \sum_{j \in V_k(i)} (\hat{r}(i,j) - k) \tag{4.3}$$

These metrics are used to evaluate the performance of a DRT technique on each pillar and also to compare the results between different types of soil texture.

Extracting summary indicators

This step consists of obtaining the linear equations provided by the preferred DRT. For ease of representation and interpretation, only the results for 2 principal components are considered. By using this method, instead of separately analysing each indicator, the data for each pillar is reduced to only two dimensions, which incorporate all individual relevant properties.

Equation 4.4 generalises the structure of a summary indicator (S), with $i \in 1, 2$. Each individual indicator I is multiplied by its respective weight w computed by the DRT algorithm and then summed together. Here, p indicates the pillar (chemical, physical or biological), and n is the number of individual indicators for the pillar, thus transforming the data from n dimensions to only 2.

$$S_{p,i} = \sum_{j=1}^{n} w_{p,j} \cdot I_{p,j} \tag{4.4}$$

RQ2. How can summary indicators be scored in a manner which expresses a soil sample's health?

The approach chosen for evaluating the indicators is using scoring curves. This method is based on SMAF [11], a framework which offers mathematical models for scoring various soil properties used to evaluate soil health. Not all previously selected indicators (according to

RQ0 and RQ1) are found in SMAF. To compensate for the missing scoring curves, other equations are extracted from related literature to score at least one pillar fully.

After each individual indicator is transformed into its corresponding scoring curve, the goal is to obtain a score for the summary indicators. This is done by multiplying the weights obtained from the DRT with the indicators score. This operation yields a new scoring curve, as described by equation 4.5. The variables w, i, S, p and I retain the same meanings as those defined in equation 4.4. An important modification is that the scores are multiplied by the *absolute* value of the weights (which can be < 0 in the PCs), as scoring is always positive. f is the transformation function that converts raw measurements into scores. This function can also accept multiple additional arguments, $c_{1,..,m}$, such as climate or soil texture class. These extra parameters are difficult to generalise, as they are contextual and vary not only according to what indicator is scored but also to the general properties of the sample itself. To ensure the scores remain in the $[0, 1]$ interval, we divide by the sum of the weights and the final score of the summary indicator i for pillar p is denoted as $SS_{p,i}$.

$$SS_{p,i} = f_{S_i}(S_{p,i}) = \sum_{j=1}^{n} |w_{p,j}| \cdot f_{I_{p,j}}(I_{p,j}, c_1, ..., c_m) \Big/ \sum_{j=1}^{n} |w_{p,j}| \tag{4.5}$$

RQ3. What is the most suitable manner of expressing the results of the assessment?

The main goal of this part of the research is to visualise the soil health scores per pillar. The stakeholders (soil ecologists and farmers, mainly) need an easy-to-interpret, straightforward representation method which not only reveals the score of a sample but also places it relative to the summary indicators, therefore being able to detect if and what improvements need to be made in the soil.

The approach consists of constructing a 2-dimensional heatmap plot with the summary indicators as axes. As the previous RQ managed to assign a numerical score to a summary indicator, combining the two indicators is not a complex task. The healthiness of a sample for each pillar can be easily assessed by consulting the position of the sample in the heatmap. The visualisation of the scores on the heatmap is done in two separate ways, by discretising the scoring intervals and by simply plotting a colour gradient. The two versions are compared and analysed to understand which plot better fulfils the requirements.

Each summary indicator explains a certain fraction of the data variance. Therefore, a principal component which describes a high amount of information is of more significance in scoring a soil sample than a PC with a lower information content. For short, the final score of a pillar is computed as the weighted sum of the scores for both summary indicators, as described in equation 4.6.

$$TS_p = \sum_{i=1}^{k} ev_{p,i} \cdot SS_{p,i} \Big/ \sum_{i=1}^{k} ev_{p,i} \tag{4.6}$$

TS_p represents the *total score* for pillar p. k denotes the number of associated summary indicators (or dimensions to which the data was reduced). SS is the *summary indicator score*, as derived by equation 4.5. Finally, ev is the weight of the summary score, corresponding to the *percentage of explained variance*. The equation is normalised by dividing by the sum of these weights to ensure the resulting score lies within the $[0, 1]$ interval.

The algorithm used to obtain the summary indicators is Sparse PCA, a variant of PCA. Given the significant differences between the two methods, the formula for calculating the fractions of explained variance must be adjusted to accurately reflect these values.

Adjusted explained variance for SPCA

In regular PCA, calculating the explained variance ratio for each component is a simple and direct process, as the components are orthogonal, meaning no overlap in the information they capture. However, in Sparse PCA (SPCA), the loadings are not orthogonal, leading to some redundancy across the vectors. Shen and Huang [59] offer a method to compute the variance explained by each component while accounting for both the correlations among principal components and the non-orthogonality of the loadings — challenges absent in standard PCA. While an earlier approach to calculating adjusted variance existed [60], it only addressed the correlation among the principal components without resolving the issue of non-orthogonal loadings.

Instead of using the projection of the data onto each of the loading vectors, the authors consider the projection onto a *k-dimensional subspace* defined by the first k loading vectors. Equation 4.7 describes how the projection is computed. Here, $\mathrm{V}_k = [\mathrm{v}_1, ..., \mathrm{v}_k]$, where v_i is the i^{th} sparse loading vector and X is the high-dimensional data in matrix form.

$$\mathrm{X}_k = \mathrm{X}\mathrm{V}_k(\mathrm{V}_k^T\mathrm{V}_k)^{-1}\mathrm{V}_k^T \tag{4.7}$$

By applying these concepts, equation 4.8 shows how the *total variance explained* (TVE) by the first k principal components is computed. For PCA, this equation simplifies to $tr(\mathrm{U}_k^T\mathrm{U}_k)$, where $\mathrm{U}_k = [\mathrm{u}_1, ..., \mathrm{u}_k]$ is the matrix of the first k principal components. The tr function refers to the *trace of a matrix*, which sums the elements on the main diagonal. The total variance in the original data matrix is computed as $tr(\mathrm{X}^T\mathrm{X})$.

$$TVE_k = tr(\mathrm{X}_k^T\mathrm{X}_k) \tag{4.8}$$

Equation 4.9 shows how the *adjusted variance for the k^{th} principal component (AV_k)* is obtained.

$$AV_k = tr(\mathrm{X}_k^T\mathrm{X}_k) - tr(\mathrm{X}_{k-1}^T\mathrm{X}_{k-1}) \tag{4.9}$$

The authors define the *cummulative percentage of explained variance (CPEV)* by the first k PCs using equation 4.10. The result of this formula is in the $[0, 1]$ interval.

$$CPEV_k = tr(\mathrm{X}_k^T\mathrm{X}_k)/tr(\mathrm{X}^T\mathrm{X}) \tag{4.10}$$

Although not explicitly detailed in the paper, the *adjusted percentage of explained variance for the k^{th} sparse PC (APEV)* can be derived from equations 4.9 and 4.10. These particular values are required in this research to assign a numerical weight to each principal component (also referred to as a summary indicator) in the score computation of a soil sample. The formula 4.11a computes the fraction of explained variance for each principal component. Since it is a recursive formula, the first value (associated with the first principal component, which explains the most variance) is determined by equation 4.11b. The $APEV$ formulas are used to calculate the ev parameters from formula 4.6.

$$APEV_k = [tr(\mathrm{X}_k^T\mathrm{X}_k) - tr(\mathrm{X}_{k-1}^T\mathrm{X}_{k-1})]/tr(\mathrm{X}^T\mathrm{X}) \tag{4.11a}$$

$$APEV_1 = tr(\mathrm{X}_1^T\mathrm{X}_1)/tr(\mathrm{X}^T\mathrm{X}) \tag{4.11b}$$

Automated pillar scoring tool

Another objective of this research is to develop a prototype for an easy-to-use tool that can evaluate and score soil samples by scoring each of the three pillars.

The `solara` framework is utilised to create a demo web application that allows users to input values for relevant soil health indicators. These inputs are then analysed, and the application generates the pillar's final score. Additionally, it displays a scoring heatmap with the current sample highlighted, facilitating visual analysis.

Ideally, this tool should be easily extendable, ultimately calculating a comprehensive *soil health score* as either an average or weighted sum of the individual pillar scores.

RQ4 - How does soil texture type influence the indicators and the scoring?

The column `texture` from the dataset indicates whether a sample is classified as being marine clay (M) or sand (S). The samples are split according to this attribute before applying any processing. The aim is to compare the results obtained by each soil texture type and observe if any significant differences appear. Experts consider different soil textures to be completely different mediums, therefore discrepancies in behaviour are expected to occur. The two soil textures are compared according to how they react and change across various steps of the research.

Besides the general interpretation of the results obtained for previous research questions, each sample of the dataset can be scored. Basic statistic metrics can be extracted for each soil type and compared, potentially providing insight into how scoring varies with textural context.

Chapter 5

Results

5.1 RQ0 - Finding relevant soil health indicators

After collecting and screening roughly 30 papers, books, articles and law proposals, approximately 80 different measurements were extracted. Each of these properties is proposed as a soil health indicator. However, it is not feasible nor effective to even attempt to match all of them to the dataset. To perform an initial reduction, each indicator was attributed a frequency, keeping track of how many times it shows up in literature.

Table 5.1 compiles some of the most used indicators used to evaluate soil health. The indicators are classified based on soil pillars (physical, chemical, biological) and are associated with the number of references in which they have occurred as relevant. Full citations for each indicator can be viewed in Appendix A. The chemical pillar has 12 indicators, the physical one has 8 and the biological pillar has only half of the chemical one, with 6 measurements, with a total of 26.

Certain contextual differences exist and the names of the features vary slightly, for example "x in soil", "extractable x", and "plant available x". All these variants are counted towards the same element x. Aggregate stability is used to describe more. In another case, surface hardness and penetration resistance are different measures, although both indicate the degree of soil compaction, so they are under the name "compaction".

Finding thresholds for each indicator was not possible, but the values are under the Criteria column if they were available. Some of the limits are also summarised: for each of the 4 references, the pH had different values and the current interval shows the lower and upper limits and the heavy metals have slightly different recommended values for each metal. The thresholds are only present in this table as proof that they are hard to find and not constant. Most, if not all of these indicators are highly contextual, depending on other factors (climate, location, crop etc.), which is not reflected in the current criteria column.

The references included are articles, research papers, reviews of existing methods, legislation, and commercial tests. Compared to the other two pillars, the number of biological indicators appearing in more than two publications is lower, suggesting a lack of focus on bioindicators. Another reason for the relative absence of more diverse biological indicators is the feasibility of obtaining such measurements; the number of existent species is very large, thus difficult and time-consuming to measure, interpret, and even summarise.

Dataset Exploration

The individual indicators from table 5.1 were matched with the columns from the dataset, trying to obtain as many correspondents as possible. The final subset of columns for each

	Indicator	Criteria	# ref
Biological	Respiration *		18
	Organic Carbon (SOC) *	≥ 20 g/kg [14]	15
	Organic Matter (SOM) *		15
	Earthworms		10
	Microbial biomass *		9
	Active Carbon *		5
Physical	Bulk density *		22
	Aggregate stability		20
	Water holding capacity *		14
	Texture *		10
	Water infiltration		9
	Porosity		7
	Mean weight diameter *		6
	Compaction		5
	Erosion	<2 t/ha/y [4]	4
Chemical	Nitrogen (N) *		21
	pH *	∈[5, 8] [61, 42, 14, 30]	19
	Electrical conductivity	<4 dS/m [4]	17
	Phosphorus (P) *	<30-50 mg/kg [4]	17
	Potassium (K) *		14
	CEC *		9
	Iron (Fe) *		9
	Heavy metals (Zn, Co, Cu etc.) *	≤ 200-300 mg/kg [62, 63]	8
	Manganese (Mn) *		8
	Calcium (Ca) *		6
	Magnesium (Mg) *		5
	Aluminum (Al) *		4

TABLE 5.1: Most used indicators grouped by soil component with available limits and the number of publications which mention them as relevant to soil health. 27 sources were used for this table. Indicators marked with an asterisk (*) show that a correspondent in the dataset exists. Appendix A shows the complete set of sources for each indicator.

pillar can be seen in table 5.2. Especially for the chemical pillar, most elements have more than one measurement present in the dataset, which can indicate specific forms in which it can be found in the soil (soil stock or plant available element, for example) or the method used for extraction, most of which have different units of measurement (g/kg, mol/kg, %

etc). Selecting all related columns would result in a high number of columns and lead to a lack of consistency, as not all relevant elements have the same extra measurements. For these reasons, the most straightforward columns were selected for each indicator, with the same unit of measurement where possible (mg/kg). In the case of the water-holding capacity of the physical pillar, only one column is selected, although 3 other hydrological measurements are present in the dataset. They were not included as they are either vaguely explained or do not directly indicate this precise property. Compaction can be described through multiple measurements, such as penetration resistance, surface and subsurface hardness and bulk density. The latter is expressed as a standalone feature as it is the most used physical indicator in literature.

The number of missing values for each column was analysed, resulting in excluding some measurements of nitrogen and carbon (`TN_WUR` (18 `NaNs`), `CN_WUR` (18), `coarseCN` (50) and `fineCN` (17)). `Bact.mass` and `Fung.mass` both have 18 `NaNs` but were not removed due to the already limited set of features associated with the biological pillar. After this initial analysis and feature selection, the subset on which further analysis will be performed is comprised of 13 chemical columns, 3 physical and 8 biological, with a total of 24 features. Including all related columns would have led to over 40 separate measurements. Certain indicators are assigned multiple columns from the dataset (the indicator for heavy metals has 3 measurements for different elements, and organic carbon has 3 columns which express the same measurement).

Texture is considered an important physical indicator and the percentages of silt, sand and clay are present as measurements in the dataset. They are not included in table 5.2 as they are considered contextual parameters, which play a role in evaluating other selected indicators for soil health.

Some of the column names in table 5.2 may not be entirely suggestive of what they represent, therefore they will be briefly explained. Some abbreviations are in Dutch, while some are in English. The `Eurofins` addition indicates the laboratory which extracted a significant amount of measurement from the samples, whereas `WUR` stands for Wageningen University and Research. For some chemical elements, the name also reflects how the element is found in the soil (`PAE` or `besch`[1] — plant available element) or, in the case of phosphorus, the method used for extraction (`Olsen`). `CEC` stands for *cation exchange capacity*, a measure for soil fertility. In the biological pillar, the `Respiration` measurement accounts for how much CO_2 a soil sample can produce, indicating microbial activity. `OS`[2] — measures the organic matter and `RE6` stands for *Rock-Eval 6* and is an extraction method for organic carbon (`OC`).

5.2 RQ1 - Grouping individual indicators

After selecting relevant columns for each indicator, the samples are split into two subsets, based on the texture type: sand (S) or marine clay (referenced as M or M Clay). Further processing and analysis are performed on each subset separately, as soil has an inherently different behaviour according to its fundamental composition. Samples have a dedicated `texture` column (which indicates either S or M) and, for this binary split, the percentages of clay, silt and sand were not used.

[1] NL: *beschikbaar*
[2] NL: *organisch stof*

	Indicator	Column Name	Unit
Biological	Respiration	`Respiration` ▲●	µg/g/day
	Organic Carbon	`OC_RE6` ▲●, `OC_WUR`	g/kg
		`C.org_Eurofins`	%
	Organic Matter	`OS_Eurofins`	%
	Microbial biomass	`Bact.mass` ▲●, `Fung.mass`	µg/g
	Active Carbon	`ActiveC_soil`	g/kg
Physical	Bulk density	`BD` ▲●	g/cm^3
	Water holding capacity	`moisture` ▲●	%
	Mean weight diameter	`MWD` ▲●	mm
Chemical	Nitrogen (N)	`N.Tot_Eurofins` ▲●	mg/kg
	pH	`pH` ▲●	-
	Phosphorus (P)	`P.Olsen` ▲●	mg/kg
	Potassium (K)	`K_Eurofins` ▲●	mg/kg
	CEC	`CEC_Eurofins` ●	mmol/kg
	Iron (Fe)	`Fe_Eurofins` ▲●	µg/kg
	Heavy metals	`Cu.PAE_Eurofins` ▲●, `Co.PAE_Eurofins` ▲●	mg/kg
		`Zn.PAE_Eurofins` ▲	
	Manganese (Mn)	`Mn_Eurofins`	mg/kg
	Calcium (Ca)	`Ca_besch_Eurofins` ▲●	kg/ha
	Magnesium (Mg)	`Mg_Eurofins` ▲●	mmol/kg
	Aluminum (Al)	`Al_ox` ●	g/kg

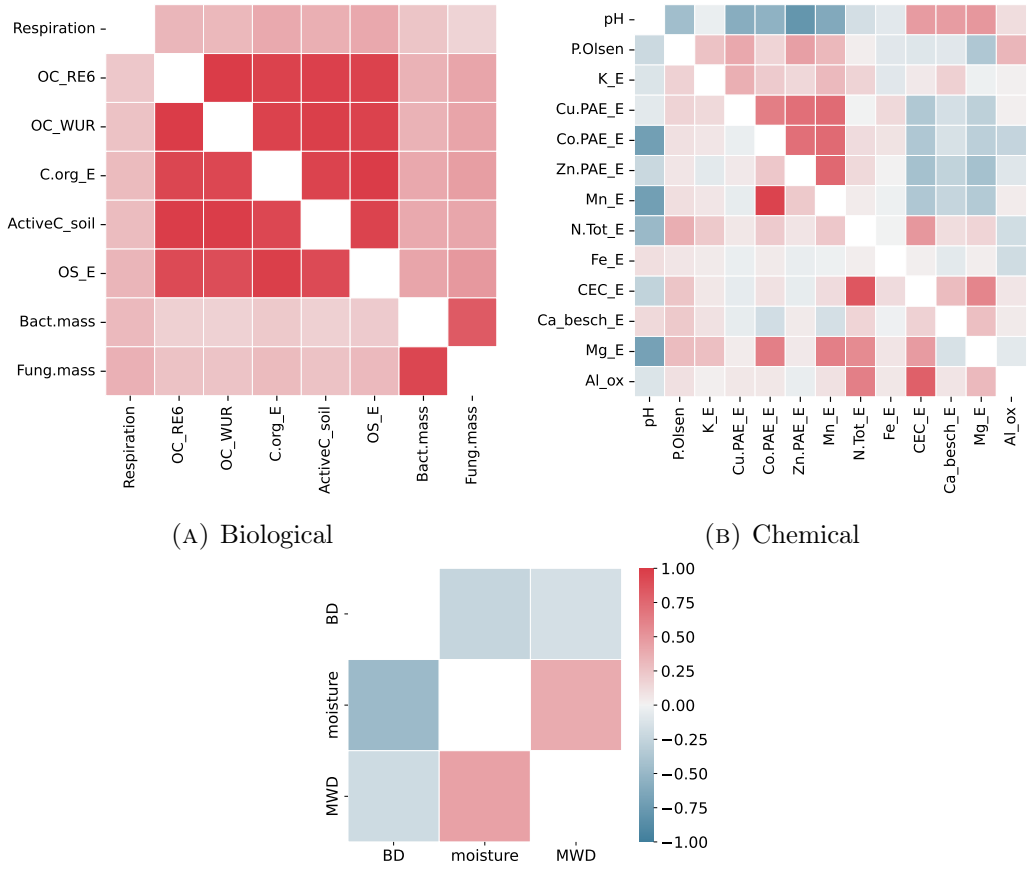
TABLE 5.2: All columns that align with the indicators from table 5.1, split by pillar type. After removing redundant columns, purple triangles ▲ indicate features retained in the marine clay subset, and yellow bullets ● - the sand subset.

Removing redundant features

The first step in reducing the dimensionality of the data consists of selecting non-correlated individual indicators. This translates to computing the Pearson correlation coefficient (r) between each feature in a pillar. We consider two features to be strongly correlated if $abs(r) \geq 0.75$, a threshold which ensures that connections between measurements are detected, especially where it would be expected (measurements assigned to the same indicator, for example), while also being high enough that potentially important features are retained.

The correlation matrices for the biological (figure 5.1a) and physical pillars (figure 5.1c) indicate the absence of any strong negatively correlated features in either type of soil texture. The connections between closely related indicators are straightforward for these two pillars, especially given their already limited number of features. For both marine clay and sand subsets, the intra-pillar associations are almost identical (visually,

they are mirrored versions of each other), the only distinction being in the magnitude of the correlation coefficients. In the case of the physical pillar, no two indicators correlate strongly enough to pass the threshold in either the sand or marine clay subset, thus no features are eliminated in this step for this pillar.



(A) Biological

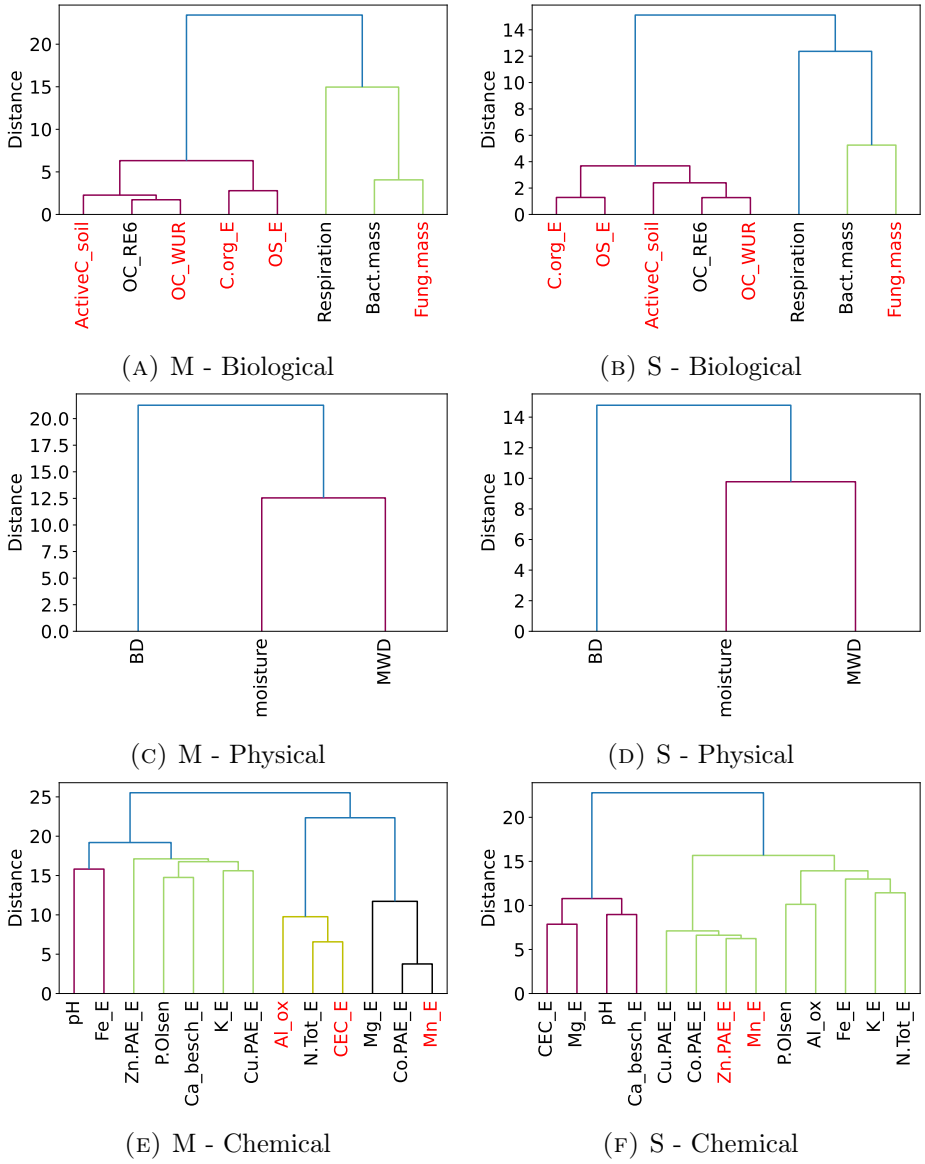
(B) Chemical

(C) Physical

FIGURE 5.1: Feature correlation matrices for all 3 pillars. The lower triangular half is for the marine clay subset, and the upper half is for sand. *Eurofins* was replaced with *E* for convenience. Full correlation values are shown in Appendix B.

In the case of the biological pillar, all organic carbon measurements (`OC_RE6, OC_WUR, C.org_Eurofins`), along with the active carbon (`ActiveC_soil`) and the organic matter (`OS_Eurofins`) model the same behaviour. All these 5 features have $r \geq 0.95$ in the sand subset and at least 0.9 for the marine clay one, indicating a meaningful positive correlation. Another strong connection is established between the fungi (`Fung.mass`) and bacteria (`Bact.mass`) quantities, potentially indicating, to some extent, the co-dependence of these two kingdoms in the environment. `Respiration` is described in the related literature as a good indicator for microorganisms residing in the soil, but has low correlation scores with the rest of the selected columns, regardless of soil texture. As it does not exhibit a similar behaviour as the others, the `Respiration` column is retained, along with `OC_RE6` and `Bact.mass` from previous categories.

The associated dendrograms for both the biological (figures 5.2a and 5.2b) and physical pillars (figures 5.2c and 5.2d) reiterate the associations between the features, revealing negligible distances between the organic carbon/matter values and clustering together the



(A) M - Biological (B) S - Biological

(C) M - Physical (D) S - Physical

(E) M - Chemical (F) S - Chemical

FIGURE 5.2: Ward linkage dendrograms for all 3 pillars for both textures (M - marine clay, S - sand). *Eurofins* was replaced with *E* for convenience. The columns coloured in red are removed after the correlation analysis.

columns which express similar soil characteristics. The clusters form identically for either soil texture type, with small distance-value discrepancies.

The chemical pillar does not offer associations as intuitive as its counterparts. The coefficients also differ more drastically between the sand and the marine clay types, not only in magnitude but also in polarity, as it can be seen in figure 5.1b (the r for `Mg_Eurofins` and `pH` is positive for sandy soils with $r = 0.49$ and negative for samples made out of marine clay with $r = -0.69$). This significant discrepancy supports analysing the two soil types separately, as their characteristics differ too much to yield substantial and relevant results if examined together. The dendrograms in figures 5.2e and 5.2f show the clusters form very differently depending on texture type. Heavy metals (Cu, Co, Zn) cluster together in the sand subset but are more dispersed in the clay soils, where the distance between copper and zinc almost doubles. No clear pattern or connection between features can be

retrieved from these dendrograms. According to the correlation values, `Mn_Eurofins` is removed from both texture types, `CEC_Eurofins` and `Al_ox` from the marine clay only and the sand subset loses `Zn.PAE_Eurofins`.

	MARINE CLAY	SAND
B	Respiration, OC_RE6, Bact.mass	
P	BD, moisture, MWD	
C	pH, P.Olsen, K_Eurofins, N.Tot_Eurofins, Fe_Eurofins Cu.PAE_Eurofins, Co.PAE_Eurofins, Ca_besch_Eurofins Mg_Eurofins	
	Zn.PAE_Eurofins	CEC_Eurofins, Al_ox

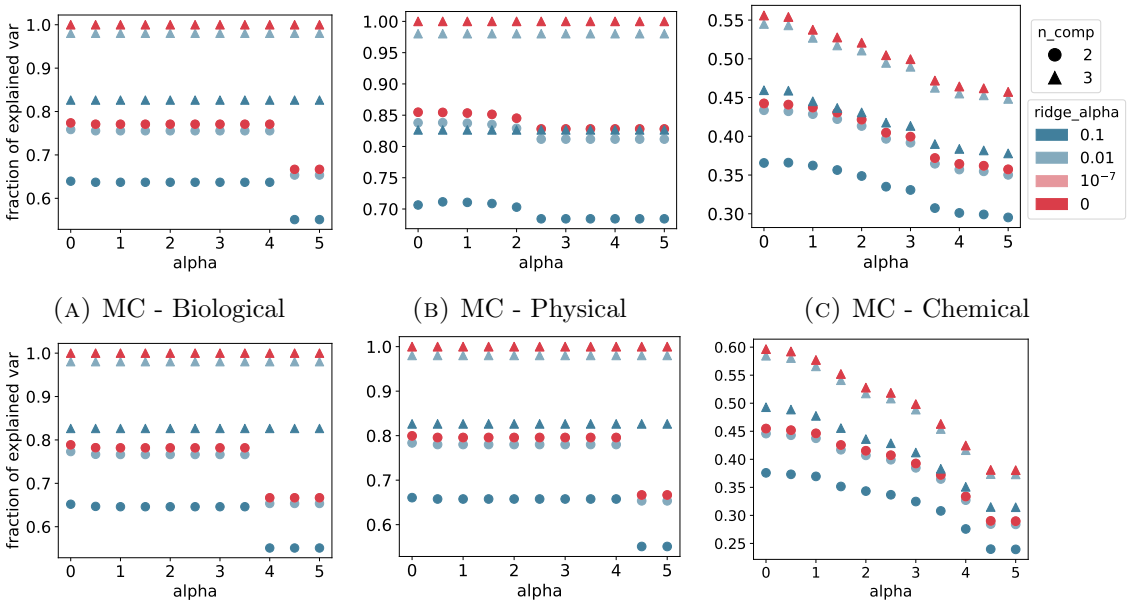
TABLE 5.3: Selected indicators per pillar (denoted only by initial) and soil texture type after correlation analysis

After discarding redundant features, out of 24 initial columns, 16 remain for the marine clay texture and 17 for the sand one. For both textures, the biological pillar loses the same columns, with a count of 3 final features, while the physical pillar remains unchanged with the initial 3 columns. There are 9 chemical features shared by both textures. Specific column names and indicators left for each soil type are marked in table 5.2 with a superscript symbol: triangle - marine clay, bullet - sand. They are also shown in table 5.3.

Finding the trade-off in Sparse PCA parameter tuning

After obtaining each pillar's final set of features for both texture types, the dimensionality reduction algorithms can be adjusted, applied and compared. The implementation of Sparse PCA (SPCA) can be tuned to maximise the components' sparseness while retaining sufficient information. This requires finding a trade-off between the `alpha` and `ridge_alpha` hyperparameters and the explained variance of the model (computed with formula 4.1). The higher `alpha` is, the sparser the components, whereas `ridge_alpha` is a regularisation parameter. Sparsity is considered an important factor in the context of this research as it greatly simplifies the resulting principal components (more coefficients are set to 0) without losing a considerable portion of the explained variance.

Experiments are run for `alpha` between 0 and 5, with a step of 0.5 and `ridge_alpha` $\in \{10^{-i} | i = 1..7\} \cup \{0\}$. Lower values for the latter parameter yield overlapping points, therefore only four of its values are plotted. In all graphs `ridge_alpha` $= 10^{-7}$ is obscured by the points when this parameter is set to 0. Figure 5.3 illustrates how the explained variance fluctuates for each soil type for the three data pillars (biological, physical and chemical) relative to the hyperparameters. The goal is to maximise sparseness while, at the same time, preserving as much of the explained variance as possible. It is immediately visible that increasing the value of `alpha` leads to a decline in the variance score, especially in the case of the chemical pillar in figures 5.3c and 5.3f. Similarly, `ridge_alpha` has a perceptible impact on the score, a lower value for this parameter leading to better results. From these plots, the adequate values for `alpha` for each pillar are extracted and can be observed in table 5.4. The choice of `alpha` is based mostly on the results associated with 2 principal components (`n_comp`), given that the corresponding explained variance decreases



(A) MC - Biological (B) MC - Physical (C) MC - Chemical

(D) S - Biological (E) S - Physical (F) S - Chemical

FIGURE 5.3: Values for total fraction of variance explained plotted against `alpha` for Sparse PCA. Experiments were run for 2 and 3 principal components (n_comp) and four values for `ridge_alpha`. The points corresponding to `ridge_alpha`=10^{-7} overlap with the 0 values, being identical. The first row shows results after running for the marine clay samples (MC) and the second row is for the sand subset (S).

faster and further analysis will be done mainly for 2 PCs. No hard threshold was set for extracting `alpha`, the parameter was chosen in all 6 cases as the last value before a drop in explained variance. The curve is smoother in the case of the chemical pillar, where the aim was to extract `alpha` by retaining at least 0.4 in explained variance. The `ridge_alpha` parameter is set to 0 in all cases, as it provides the best performance.

	BIO	PHYS	CHEM	ALL
M CLAY	4	5	2	4
SAND	3.5	4	2	3

TABLE 5.4: Selected `alpha` values for Sparse PCA for each pillar and soil texture. The All column combines the three pillars; its associated graphs for the SPCA trade-off can be found in Appendix B.

Comparing evaluation metrics

Further, the dimensionality reduction techniques are applied to the marine clay and sand samples separately and evaluated with three metrics: R^2, trustworthiness and continuity, as described in section 4.2. Continuity shows an almost identical trend in scoring algorithms as the trustworthiness metric and, also due to space limitations, it is not included in the table 5.5. Since neither of the kernel PCA results provide explicit values for the composition of the principal components (coefficients for every raw indicator), these algorithms are not

included in this section of results. With few exceptions, all kernels perform poorer than their linear counterparts. The full tables with all metrics and algorithms can be consulted in Appendix C.

		MARINE CLAY				SAND			
		n = 2		n = 3		n = 2		n = 3	
		R^2	Trust	R^2	Trust	R^2	Trust	R^2	Trust
BIO	PCA	0.77	0.90	1	1	0.79	0.89	1	1
	SPCA	0.77	0.90	1	1	0.78	0.91	1	1
	IPCA	0.77	0.90	1	1	0.79	0.90	1	1
PHYS	PCA	0.85	0.94	1	1	0.80	0.92	1	1
	SPCA	0.83	0.92	1	1	0.80	0.92	1	1
	IPCA	0.85	0.94	1	1	0.80	0.92	1	1
CHEM	PCA	0.44	0.79	0.56	0.87	0.45	0.85	0.60	0.93
	SPCA	0.42	0.79	0.52	0.79	0.42	0.83	0.53	0.91
	IPCA	0.44	0.79	0.55	0.86	0.44	0.86	0.59	0.93
ALL	PCA	0.40	0.82	0.51	0.87	0.44	0.86	0.54	0.93
	SPCA	0.31	0.79	0.41	0.81	0.38	0.84	0.45	0.90
	IPCA	0.39	0.83	0.50	0.88	0.44	0.86	0.54	0.93

TABLE 5.5: Evaluation scores for PCA. SPCA. and IPCA applied to individual pillars, and all three combined, for 2 and 3 resulting principal components (n). Results for marine clay and sand subsets.

The differences **between pillars**, regardless of soil texture, metric or number of principal components, are quite obvious. The more features (columns) a pillar has, the lower the results. The biological and chemical pillars have only 3 final columns as input for the DRT, which causes the perfect scores of 1 for both metrics in the case of 3 PCs (there is no dimensionality reduction in this case, the data does not transform). The physical pillar performs slightly better than the biological one for the marine clay subset, the highest difference in R^2 being 0.08 and 0.04 for trustworthiness, with negligible discrepancies for the sand texture (less than 0.03 difference regardless of metric or algorithm). The chemical pillar receives significantly lower values due to having nearly twice as many features as the other two pillars combined. Its explained variance is nearly half that of the physical pillar, while trustworthiness decreases by approximately 0.15 for 2 PCs in marine clay and 0.09 in sand.

There are little differences between **soil textures**, any two correspondent values being less than 0.11 apart for both evaluation metrics. The physical pillar shows better results in the case of the marine clay subset. In contrast, the biological and chemical pillars (also the combined pillars) have higher metric values for the sand texture.

Concerning the discrepancies between the **dimensionality reduction techniques**, SPCA is generally outperformed by PCA and IPCA. However, the differences are not considerably high, peaking at 0.08 for trustworthiness and 0.07 for R^2 among the three

pillars, regardless of the number of principal components. The 'all' section shows similar trends, with 0.06 being the biggest difference in trustworthiness and 0.09 for R^2. As for comparing the results between **the number of principal components**, trustworthiness and R^2 are both higher when the number of PC is higher for any algorithm, pillar or soil texture.

Overall, the evaluation metrics show that some variations exist across soil texture, number of principal components, pillar and what algorithm is applied. The negligible differences between the three techniques indicate that all of them can be used with similar results. However, given SPCA's other advantages (sparseness, mainly), even if it mostly scores the lowest in trustworthiness and explained variance, this algorithm is a top choice for further analysis.

Extracting summary indicators

Concerning the composition of the resulting principal components, PCA and IPCA are complete, each feature having a coefficient $\neq 0$. This leads to intricate linear equations, the only discriminant being the magnitude of the associated factor and its polarity. In Sparse PCA, many coefficients are reduced to 0 while also minimising the loss in explained variance (the equivalent of R^2). The exact features that are nullified and in which principal component this occurs vary for each pillar according to the subset used (either marine clay or sand) and how many principal components are extracted (n). Thus, sparser components offer a better alternative than the complete ones for evaluating how and if the indicators in each component relate to each other and if this type of clustering is consistent with the relationships that occur in nature.

					n = 2		n = 3		
		P1	P2	P3	S1	S2	S1	S2	S3
M CLAY	Respiration	0.62	-0.19	0.76	0.71	0	0	0	1
	Bact.mass	0.59	-0.53	-0.61	0.71	0	1	0	0
	OC_RE6	0.52	0.83	-0.22	0	1	0	1	0
SAND	Respiration	0.55	0.75	-0.36	0	1	0	1	0
	Bact.mass	0.57	-0.65	-0.50	0.70	0	-1	0	0
	OC_RE6	0.61	-0.07	0.79	0.71	0	0	0	1

TABLE 5.6: The coefficients of 2 and 3 principal components (n) for PCA (P) and SPCA (S) applied to the **biological** pillar. Extracting 3 PCs with SPCA does not reduce dimensionality in this case. it only provides a way of comparing the results of 2 PCs.

Two and three principal components (n) are extracted from the data subsets (the two soil textures) using PCA, IPCA, and SPCA. This step's main objective is to examine SPCA comparatively, as this algorithm provides the most relevant results for addressing the current research question. Therefore, for ease of visualisation and analysis, the coefficient tables (tables 5.6, 5.7 and 5.8) present only the results for PCA and SPCA, with full tables in Appendix D. P1, P2, and P3 are the names for PCA's first, second, and third components, respectively, while S1, S2, and S3 are the equivalent components of SPCA. Only SPCA is split according to the number of components (n), as the resulting coefficients

					n = 2		n = 3		
		P1	P2	P3	S1	S2	S1	S2	S3
M CLAY	BD	-0.55	0.67	-0.50	-0.71	0	-1	0	0
	moisture	0.65	-0.03	-0.76	0.71	0	0	0	1
	MWD	0.52	0.74	0.42	0	1	0	1	0
SAND	BD	0.47	0.86	0.19	0	-1	0	-1	0
	moisture	-0.64	0.19	0.74	0.71	0	0	0	1
	MWD	-0.60	0.47	-0.64	0.70	0	1	0	0

TABLE 5.7: The coefficients of 2 and 3 principal components (n) for PCA (P) and SPCA (S) applied to the **physical** pillar.

change with n. PCA extracts the same values for principal components regardless of n, therefore the results for this technique are not separated. The components are in descending order based on how much variance they explain.

The most significant difference between the two DRTs in all cases (whether comparing pillars, soil types, or the number of components) is the number of zeroes produced by SPCA. In contrast, regular PCA yields only non-zero coefficients (the 0.00 coefficient for `Co.PAE_Eurofins` in the chemical pillar is due to the number of decimals used for representation; the weight is very small and round to 0). Typically, although not an explicit rule, SPCA nullifies the lowest absolute weights from its corresponding PCA component.

Due to varying sparsity levels (changes in `alpha` as seen in table 5.4) or inherent disparities in how indicators interact and relate to each other in different soil textures, different coefficients are zeroed out when changing the soil texture type when applying SPCA. When transitioning from marine clay to sand samples, the distribution of biological and physical indicators across components changes noticeably, showing minimal overlap and no identifiable pattern.

For the **biological pillar** (table 5.6), for $n = 3$, each feature is given a coefficient of 1, as the pillar has only 3 columns. No transformation is done on the data (the negative coefficient in the sand subset can be reversed) when applying SPCA. However, PCA assigns various coefficients to all three indicators, as it does not consider the number of dimensions needed. The first principal component computed by PCA in the sand subset assigns similar weights to all features, whereas SPCA's first component contains only the bacterial mass for $n = 3$.

When looking at the **physical pillar** in table 5.7, the same behaviour as the previous pillar is displayed, and the absolute values of the weights are equal. What is noticeable, however, is the polarity of the bulk density (`BD`) weights, as they are negative regardless of texture or number of principal components. It suggests this indicator has an inverse relationship with the other two. Being grouped in the same SPCA principal component as `moisture` in the marine clay subset can be interpreted as "denser soil can hold less water".

Applying Sparse PCA on the **chemical pillar** completely eliminates the `Fe_Eurofins` column, regardless of texture type or number of components. PCA attributes quite low weights to the iron measurement in all cases, with the highest appearing in the third component, contributing the least to the total explained variance. This may signify this feature's low impact in modelling a soil sample. `K_Eurofins` is discarded completely by SPCA in the sand samples and, for the marine clay subset, it only comes into play in

					n = 2		n = 3		
		P1	P2	P3	S1	S2	S1	S2	S3
M CLAY	pH	-0.51	-0.21	-0.10	-0.58	0	-0.58	0	0
	P.Olsen	0.26	-0.48	0.13	0	0.66	0	0.66	0
	K_Eurofins	0.19	-0.37	-0.33	0	0.27	0	0.27	0
	Cu.PAE_Eurofins	0.07	-0.21	0.05	0	0	0	0	0
	Co.PAE_Eurofins	0.44	0.36	0.00	0.53	0	0.53	0	0
	Zn.PAE_Eurofins	0.14	0.15	0.65	0.01	0	0	0	1
	N.Tot_Eurofins	0.40	-0.30	0.02	0.23	0.44	0.23	0.44	0
	Fe_Eurofins	0.02	-0.11	-0.50	0	0	0	0	0
	Ca_besch_Eurofins	-0.04	-0.55	0.38	-0.14	0.54	-0.14	0.54	0
	Mg_Eurofins	0.51	0.05	-0.20	0.56	0	0.56	0	0
SAND	pH	-0.47	-0.06	-0.13	-0.55	0	-0.55	0	0
	P.Olsen	0.29	0.16	-0.45	0.26	0	0.20	0	0.52
	K_Eurofins	0.12	0.49	-0.24	0	0	0	0	0
	Cu.PAE_Eurofins	0.43	0.26	-0.05	0.47	0	0.47	0	0
	Co.PAE_Eurofins	0.39	0.23	0.23	0.43	0	0.45	0	0
	N.Tot_Eurofins	-0.05	0.56	0.21	0	0.84	0	0.85	0
	Fe_Eurofins	0.04	0.01	0.40	0	0	0	0	0
	CEC_Eurofins	-0.38	0.40	-0.01	-0.27	0.54	-0.29	0.53	0
	Ca_besch_Eurofins	-0.25	0.28	-0.19	-0.16	0	-0.16	0	0
	Mg_Eurofins	-0.38	0.23	0.12	-0.36	0.04	-0.35	0.03	0
	Al_ox	-0.03	-0.08	-0.64	0	0	0	0	0.85

TABLE 5.8: The coefficients of 2 and 3 principal components (n) for PCA (P) and SPCA (S) applied to the **chemical** pillar. `Eurofins` was replaced with `E` in this table. The red zeroes indicate the complete absence of an indicator after running SPCA for either 2 or 3 PCs.

S2. However, potassium dominates the second principal component of PCA (P2) in both textures, having the highest absolute weight, while also being the second most important factor in the P3 of the sand subset. This can suggest that the first two sparse components sufficiently describe the dataset behaviour. Similarly, in the **sand** subset the loadings remain very close in value for S1 and S2 regardless of n. Aluminium (`Al_ox`) is only present in the third sparse principal component. S3 has two indicators (as opposed to just one for the other soil texture), with rather high weights, suggesting that they explain a non-negligible portion of variance. Regarding the polarity of the weights, a noticeable change occurs in magnesium (`Mg_Eurofins`), which flips its sign across textures. The distribution of indicators across sparse principal components differs between the two soil textures, emphasizing the underlying variation when texture changes.

To avoid overcomplicating the result and for ease of representation and interpretation, further research will be performed using the results for SPCA with 2 principal components. As the algorithm deems some chemical indicators redundant, they are also discarded. Therefore, the new count for features per pillar is: biological - 3, physical - 3, chemical marine clay - 8 and chemical sand - 8, totalling 15 separate measurements for both textures.

5.3 RQ2 - Scoring summary indicators

SMAF [11] is from 2004, a rather old framework, but has still been extensively used by other researchers. The mathematical equations and additional parameter computations are not described in the paper but in a separate Excel file. This file is not available online, but Dr. Kristen S. Veum, a research soil scientist at USDA-ARS[3] (not an author of SMAF, but one of SHAPE [12], its successor), was able to provide an updated copy of the file (from 2014) and also explain in detail how it is used and its limitations. SHAPE and SMAF are explained in more detail in Chapter 3.

Establishing the contextual parameters

Besides the actual values for the indicators, the framework requires some additional parameters, some of which are not explicitly included in the dataset. Where possible, they have been deduced logically or with the help of online resources.

Given that this framework is built to be applied in the United States, a country which spans multiple time zones, biomes and climates, with soil types which do not even exist in the region of The Netherlands, some contextual parameters can be set as constants for our research. The **climate** class does not change across the current samples, as it is determined based on annual precipitation and degree days, which are inside the limits of one class for the whole country. Given that most of The Netherlands is flat, the chosen **slope** class is the one with the lowest inclination. The **mineral class** of the soil is set as *other*, given that the other two options were smectitic and glassy (both of volcanic origin). The **season** when the samples were collected is not specified, so spring was chosen, as it was also associated with the least impacting value on the score for this parameter. The **organic matter** class is also hardcoded when applying the framework to the current samples. The four different classes are set according to the soil suborder, which is not known, and the content of organic matter (the highest content gets class 1, and the lowest is assigned class 4). We consider the soils in the dataset to be high in organic matter, therefore all are classified as class 1. Numerical tables with associations between classes and real values for site-specific parameters are extracted from the framework and can be consulted in Appendix F.

In SMAF, **texture classes** from 1 to 5 are assigned to the samples, following the logic in figure 5.4. So far, the current research focused on a binary classification — marine clay or sand. Now, the new class for each sample is determined according to its percentages of sand (`Zand_Eurofins`), silt (`Silt_Eurofins`) and clay (`Lutum_Eurofins`). In the dataset, these three measurements do not add up to exactly 100, so they are normalised to ensure correct classification.

Each sample fits into a soil class according to the texture columns. The soil class and texture counts are in table 5.9. Five entries have the class 0, as they are missing all texture columns. Ideally, the sand and marine clay subsets have no common texture classes. However, there is an overlap for class 2, the sand subset having two entries classified as such. These entries are treated as outliers and disregarded. Also, it is noticeable that both

[3] U.S. Department of Agriculture – Agricultural Research Service

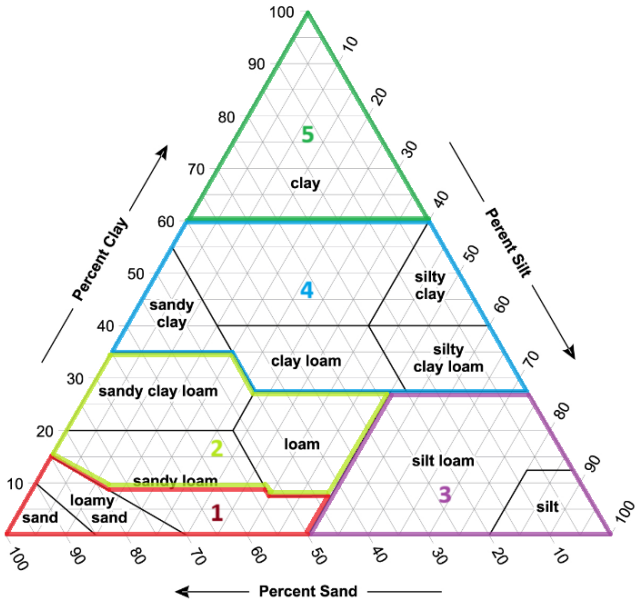


FIGURE 5.4: Soil texture classes distribution according to SMAF [11]

textures have one very dominant class: class 2 with 109 samples for marine clay and class 1 with 76 entries for sand. The sand subset is left with only one class after removing the outliers and a simple comparison between sand and clay is still the most straightforward way of analysing the results. Therefore, it seems reasonable to pick only the most populous class from each texture. With this method, the number of curves and results which need to be plotted and interpreted is reduced.

	M				S		
texture class	0	2	3	4	0	1	2
count	2	109	4	29	3	76	2

TABLE 5.9: Counts for all combinations of texture and texture classes

Extracting scoring curves

As for the scoring curves, related papers do not express the equations in as much detail as SMAF. Another alternative to having different mathematical models for each indicator is using a more general formula, which normalises the scores [38]. Equation 5.1 is used by other researchers when SMAF curves are not available for their choice of indicators, where x is the indicator value, L is the lower threshold, B is the baseline value where the score is 0.5 and S is the slope at the baseline. L, B and S are specific to each type of indicator. Three general shapes for the curves can be generated with this formula: more-is-better, less-is-better and optimum (a range of values yields a maximum score) [64].

$$y = \frac{1}{1 + [(B - L)/(x - L)]^{[2S \cdot (B + x - 2L)]}} \tag{5.1}$$

After reviewing the SMAF file, the relevant indicators' equations were extracted and listed in table 5.10. Due to space constraints, not all details are included; more precise calculations for contextual parameters are available in the framework file and detailed in Appendix F. Contextual parameters are computed using combinations of values from lookup tables for each factor (e.g., climate, texture).

	Indicator	Score Equation	a	b	c	d
Chemical	pH	$y = a \cdot \exp[(-(x-b)^2)/(2 \cdot c^2)]$	1	●	●	-
	Nitrogen	$y = a/[1 + b \cdot \exp(-c \cdot x)]$	1	161.32	■▲♦	-
	Potassium	$y = a \cdot [1 - \exp(-b \cdot x)]$	▲	▲	-	-
	Phosphorus	$y = (a \cdot b + c \cdot x^d)/(b + x^d)$	> 0	●■▲	1	3.06
		$y = a - b \cdot \exp(-c \cdot x^d)$	1	4.5	●■▲♥	-2
			L	**B**	**S**	**Ref**
	CEC	Equation 5.1	10	32.2	0.0507	[65]
Physical	MWD	Equation 5.1	0.1	0.8	2.18	[65]
			a	**b**	**c**	**d**
	BD	$y = a - b \cdot \exp(-c \cdot x^d)$	0.994	▲	▲	▲
	AWC	$y = a + b \cdot \cos(c \cdot x + d)$	0.477	0.526	6.877	▲
Biological	OC	$y = a/[1 + b \cdot \exp(-c \cdot x)]$	1	50.1	■▲♦	-
	MBC	$y = a/[1 + b \cdot \exp(-c \cdot x)]$	1	40.748	■▲♦★	-
			L	**B**	**S**	**Ref**
	Respiration	Equation 5.1	2	6.6	0.311	[65]

TABLE 5.10: Scoring equations and associated parameters for some indicators. Each coloured symbol shows that the parameter is a function of one or more contextual factors: crop type ●, texture ▲, climate ♦, organic matter content ■, slope ♥and season ★. CEC – cation exchange capacity, MWD - mean weight diameter, BD – bulk density, AWC – available water capacity, OC – organic carbon, MBC – microbial biomass carbon. CEC, MWD and Respiration are scored using formula 5.1, while the rest of the equations are extracted from the SMAF Excel file. Appendix F offers more details about how each parameter is computed.

Adapting the chemical pillar

Despite thorough literature research, not all selected MDS indicators for the chemical pillar could be matched with corresponding scoring curves or general thresholds which can be applied in equation 5.1. While partial scoring for this pillar can be conducted for individual indicators, the lack of a complete set of mathematical models makes it impossible

to generate summary scoring curves. For proof of concept and for the sake of gaining more insight into soil health, pillars and for creating a widely applicable framework, the chemical pillar will be reduced to the indicators which have available mathematical models, as depicted in table 5.10: pH (`pH`), nitrogen (`N.Tot_Eurofins`), potassium (`K_Eurofins`), phosphorus(`P.Olsen`) and cation exchange capacity (`CEC_Eurofins`).

To ensure the correctness and validity of the results, some previous steps are re-executed for this smaller subset of features. The results of the correlation analysis do not change in this case, with only CEC being redundant in the marine clay subset. Further, `alpha` trade-off in SPCA is performed, resulting in an optimum of 3.5 for both textures. The associated plots can be consulted in Appendix B. The next essential step is extracting the sparse principal components, which are displayed in table 5.11. Extended coefficients for PCA, SPCA and IPCA are noted in Appendix D and evaluation values in Appendix C.

	M CLAY		**SAND**	
	S1	S2	S1	S2
`pH`	-0,63	0	-0,71	0
`P.Olsen`	0,31	0	0,71	0
`K_Eurofins`	0	1	0	0
`N.Tot_Eurofins`	0,71	0	0	0,71
`CEC_Eurofins`	-	-	0	0,71

TABLE 5.11: New sparse principal components for the chemical pillar

When extracting 2 principal components with SPCA, potassium (`K_Eurofins`) is discarded in the sand subset and in the marine clay one it is the only one in the second component, confirming its reduced impact in describing the samples. By removing it from the sand texture (as it has no weights $\neq 0$), both soil types are left with a total of 4 indicators. These principal components are used for the chemical pillar for the rest of this study.

Scoring individual indicators

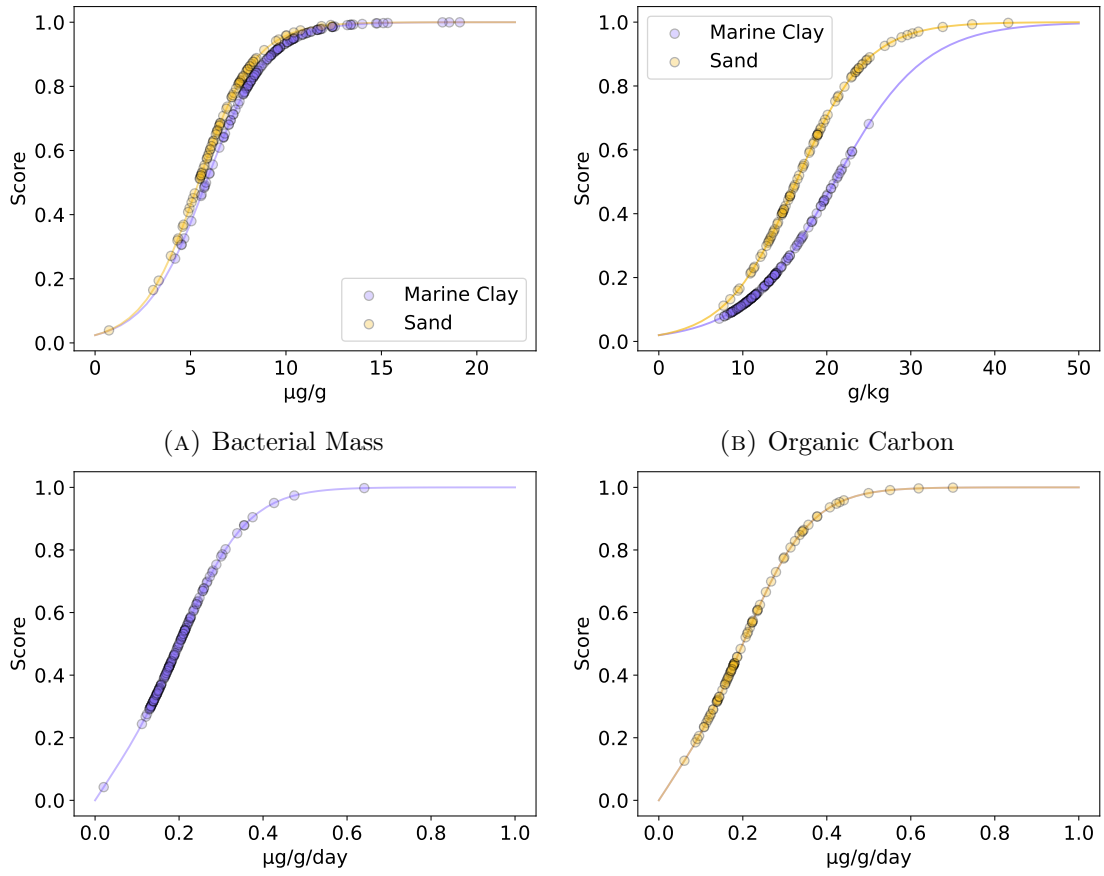
Every score equation is applied to the dataset and plotted separately for each soil texture type (marine clay and sand). To ensure all ranges of scores or possible values are covered, a theoretical curve is also plotted for each soil health indicator.

- **Biological**

Before applying the scoring equations, certain adjustments need to be made. The methods of measuring respiration vary across studies, as do the range of values this indicator can take. The thresholds for this indicator are detailed in table 5.10. The `Respiration` column in the dataset needs to be multiplied by 10 before applying formula 5.1 and the thresholds are also modified to extract a curve, keeping only the slope value unchanged. The lower threshold, L, is set to 0 and the baseline, B, to 2.5. The equation does not support scoring a sample with a value equal to the lower threshold. To be able to perform most tests and avoid division by 0, the program assigns a score of 0 when these cases occur. The baseline limit is set after experimenting with various values. The aim was to have the points reasonably dispersed on the scoring curve and obtain an equation that could span all possible

scores. More research and insight should be gained by specialists who can deduce a more proper value for the baseline parameter.

Figure 5.5 shows how each biological indicator is scored. **Bacterial mass** (figure 5.5a) shows almost no difference between marine clay and sand soils as a scoring trend. Still, marine clay exhibits higher values for this parameter, overall. The sand subset has rather evenly distributed points across the scoring range, while the clay one scores higher on average. In figure 5.5b, the curves for **organic carbon** are very clearly separated based on different soil textures. The marine clay subset seems to have quite low values for this indicator, most of which are concentrated in the 0 - 0.3 scoring interval. The highest score for this texture type is less than 0.6. For the sand texture, the values are more evenly distributed across the curve, indicating a higher diversity in the samples.



(A) Bacterial Mass

(B) Organic Carbon

(C) Respiration, marine clay

(D) Respiration, sand

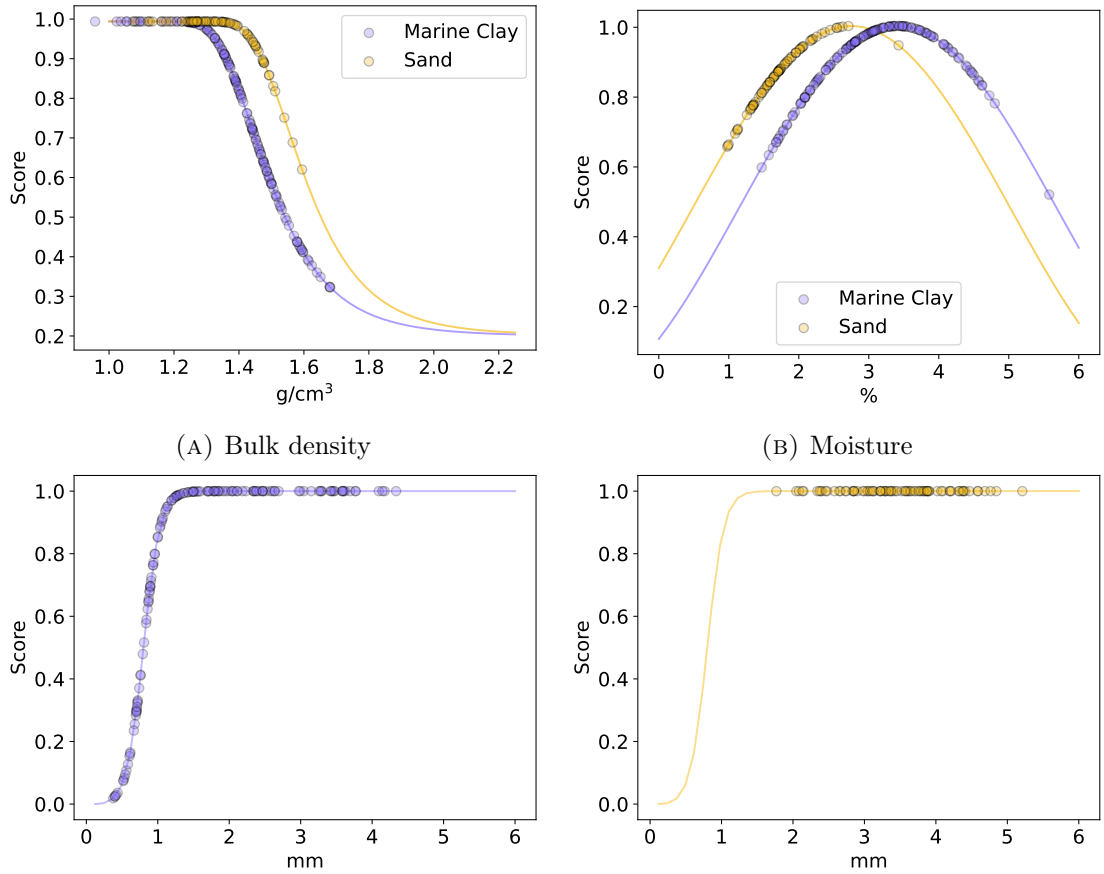
FIGURE 5.5: Scoring curves for the biological indicators by soil texture type

In the case of **respiration**, the curves are identical for both soil types, as no distinction is made in the thresholds. The sand samples in figure 5.5d are more often scored over 0.8 than their marine clay counterparts, with the latter (figure 5.5c) being concentrated in the 0.2 to 0.7 interval.

All three curves are the same type, more-is-better, attaching higher scores for higher values for each measurement. Even if the respiration indicator has a different base formula for scoring, the resulting curve has a very similar shape to the other two features.

- **Physical**

Figure 5.6 contains the scoring curves for each physical indicator. In contrast to the biological pillar, this one has all three curve types: less-is-better (bulk density), optimum (moisture) and more-is-better (mean weight diameter).



(A) Bulk density

(B) Moisture

(C) Mean weight diameter - Marine Clay

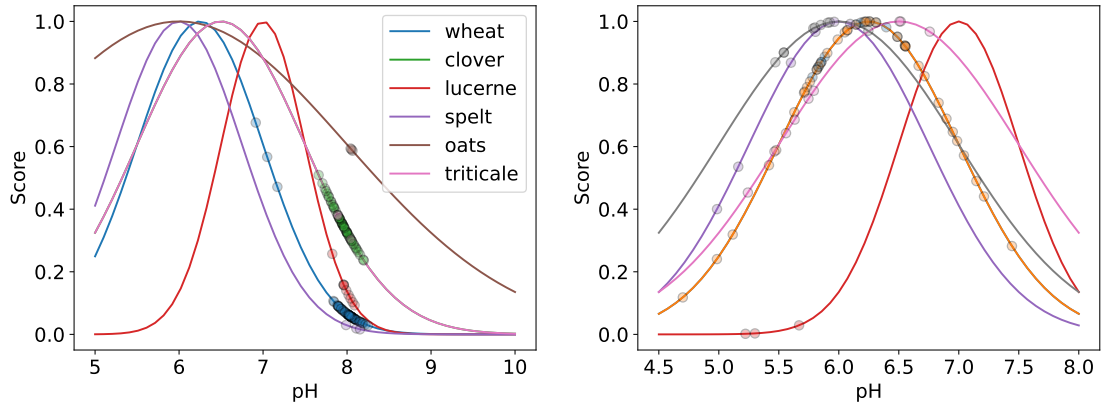
(D) Mean weight diameter - Sand

FIGURE 5.6: Scoring curves for the physical pillar by soil texture type

The sand scores for **bulk density** (figure 5.6a) are all greater than 0.6, with a big majority above 0.9. The marine clay subset is distributed uniformly across its associated curve, ranging from a perfect score to 0.3. This particular curve equation does not reach a score of 0 and converges to 0.2. The marine clay function is more sensitive to higher values for bulk density, reaching lower values faster than the sand one. As the **moisture** (figure 5.6b) is scored according to a cosine curve, it has an optimum: of around 3.5% for marine clay and 2.7% for sand. All scores are above 0.6, with only some samples exceeding the optimum percentage. For the **mean weight diameter** (figures 5.6c and 5.6d), both curves are the same. Every value higher than 1.5mm is assigned a perfect score, as seen in the case of the sand texture. The marine clay offers a broader range of values, with samples populating the entire slope.

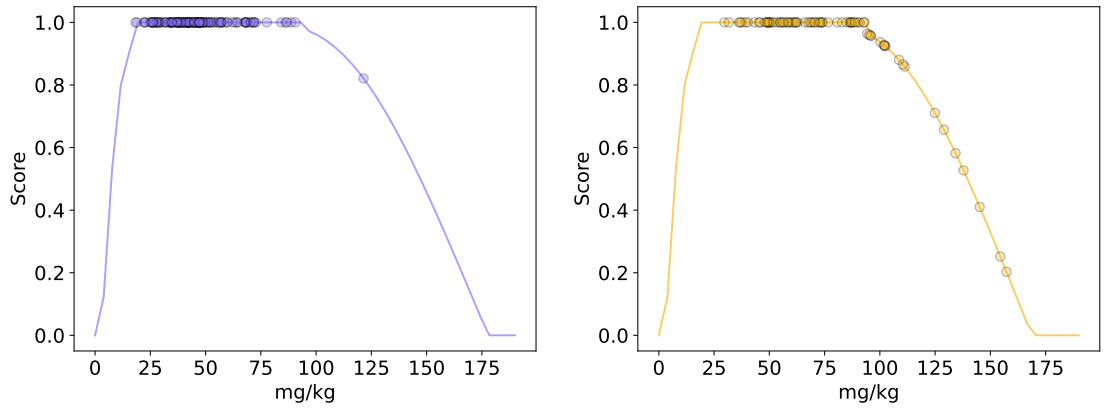
- **Chemical**

For this pillar, 5 of the 12 indicators have available equations and can be scored. The **pH** curves are shown in figure 5.7, where the distinction is made based on crop type, not texture. The graphs are separated based on soil type to emphasise the differences in the scores' distribution, but the shape of the curves is constant across textures. These curves are only for the 8 types of crops found in the dataset, which were then matched with their respective parameters using the SMAF file or other resources if needed (more details in Appendix F). Some crops have the same equation parameters, resulting in an overlap (clover and triticale in marine clay, for example). Each bell curve has an optimum value ranging from a pH of 6 to 7. The width of the bell is denoted by the crop's ability to tolerate and thrive when the acidity varies. A narrower shape would indicate a more sensitive crop, which can properly develop only in a specific interval (e.g. lucerne). In contrast, a crop with a wider bell (e.g. oats) can grow with fewer constraints. For marine clay, all but three pH values are grouped around a pH of 8, relatively clustered when compared to the sand values, which are more acidic and span from 4.5 to 7.5.



(A) Marine clay (B) Sand

FIGURE 5.7: pH scoring curves for both soil texture types

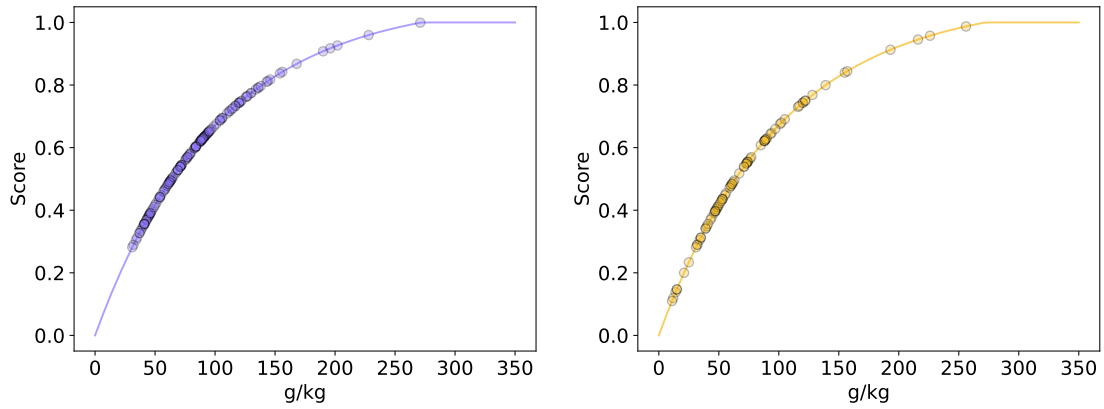


(A) Marine clay (B) Sand

FIGURE 5.8: Phosphorus scoring curves for both soil texture types

Phosphorus (figure 5.8) ascends quickly from a score of 0 to 1, where it plateaus until a certain threshold is reached, decreasing again to 0. The two textures do not have exactly

the same curve shape, but they are too similar to be easily decipherable from the same graph. With only one outlier, the marine clay samples are all awarded a perfect score, whereas some of the sandy samples have an excess of phosphorus. This indicator is also dependent on the crop type; this variable is not included in this analysis of phosphorus scores and the values associated with *wheat* are used. The **potassium** (figure 5.9) equations are identical for the two soil types. Both have a similar distribution of samples across the curves, with the sand being slightly lower in scores. Also with the same equations for both sand and marine clay, the **cation exchange capacity** (figure 5.10) assigns a score of 1 to values over 50 mmol/kg. Marine clay samples are all over 100, while the sand subset has much lower values, with only a fraction of the total samples scoring the maximum. For **nitrogen** (figure 5.11), the textures have separate curves, but all samples are scattered across the 0.2 - 1 interval.



(A) Marine clay (B) Sand

FIGURE 5.9: Potassium scoring curves for both soil texture types

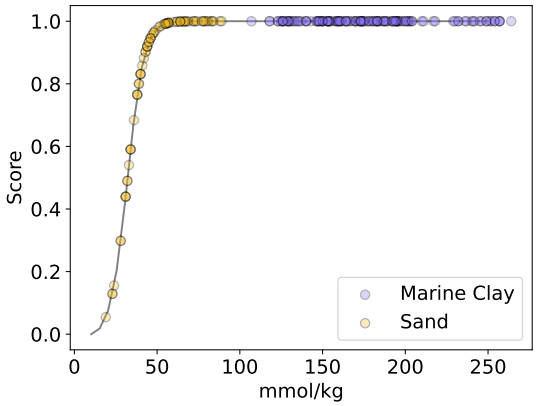


FIGURE 5.10: Cation Exchange Capacity scoring (CEC) curves

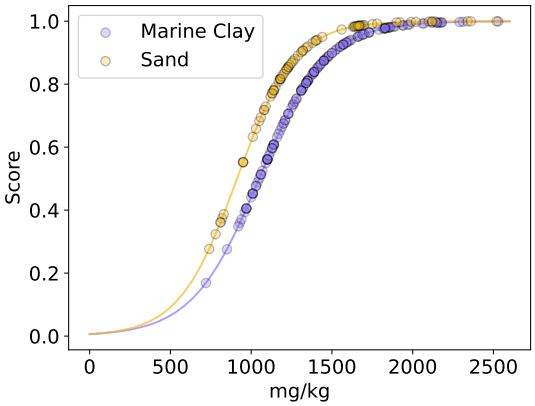


FIGURE 5.11: Nitrogen scoring curves

Scoring summary indicators

After attributing scores to all individual indicators, the next step is evaluating the score of a principal component. Each extracted PC is expressed as a linear combination of the selected features. Some PCs consist of just one individual indicator (most of the second

principal components), in which case computing the score is straightforward and does not require additional processing. The majority of the principal components are, however, made out of 2 or 3 individual measurements (up to 6 if referring to the original chemical pillar), which need to be analysed more in-depth.

As a side note, when applying SPCA, the data is first scaled and centred to compensate for the differences between the measurements (e.g. `Respiration`'s values are < 1, while `OC_RE6` approaches 50). This ensures the resulting PCs describe the data accurately and are not biased by units of measurement and scales. However, raw indicator values are used when plotting, as well as when computing the scores. The value of the PC itself is of little interest, as it does not provide any meaningful information on the sample score. The actual equation of each PC is the crucial element in scoring summary indicators.

Due to different value-to-score functions, the scoring of summary indicators cannot be represented by a single, smooth curve. This is because various combinations of indicator values can yield the same X-axis values while producing different associated scores. Appendix E details the attempt to obtain a single scoring equation for a summary indicator using a curve-fitting algorithm. The solution is not appropriate for the scope of the project, as it requires additional user input and specific knowledge.

All but one individual measurement can be treated as constants to eliminate the need for additional computation, analysis, and user input while also obtaining singular scoring curves. The new scores are calculated using the linear equation of the PC, substituting the indicator's raw value with its corresponding score. The weights do not add up to 1, so the result is then divided by the sum of the weights (using formula 4.5).

- **Biological Pillar**

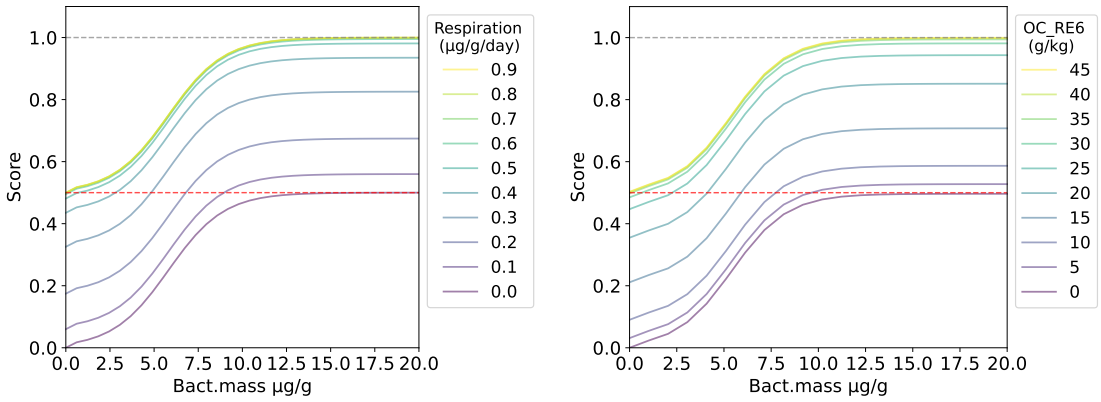
Table 5.12 shows the equations used for computing the scores for each principal component. As the second principal component (PC2) is composed of only one indicator for both texture types, its evaluation is straightforward and requires no additional analysis.

M	$s(PC1) = [0.71 \cdot s(\texttt{Bact.mass}) + 0.71 \cdot s(\texttt{Respiration})] \,/\, (0.71 + 0.71)$ $s(PC2) = 1 \cdot s(\texttt{OC\_RE6})$
S	$s(PC1) = [0.70 \cdot s(\texttt{Bact.mass}) + 0.71 \cdot s(\texttt{OC\_RE6})] \,/\, (0.70 + 0.71)$ $s(PC2) = 1 \cdot s(\texttt{Respiration})$

TABLE 5.12: Formulas for scoring the principal components for the biological pillar by soil texture type (M = marine clay, S = sand). The s() function transforms the value of an indicator into a score.

Since the first PCs only have 2 indicators for the two soil textures, only one measurement is discretised. The first marine clay and sand subsets' first PC has the `Bact.mass` indicator in common, which will not be sampled.

Figure 5.12 illustrates how the scoring curves develop according to the values of the indicators. The evolution of the scores when only the `Bact.mass` measurement varies is easier to assess and assigning a certain score for the summary indicator is possible. For both first principal components, 10 samples are extracted from the `Respiration` (figure 5.12a) and `OC_RE6` (figure 5.12b) indicators to demonstrate their impact on the summary curves.



(A) PC1 of Marine Clay

(B) PC1 of Sand

FIGURE 5.12: Summary scoring curves for the first principal component of both soil textures for the **biological pillar**. Plotting is done relative to the `Bact.mass` indicator, which is common in both principal components. The second indicator in each component is discretised. The Score axis is a function of `Bact.mass` and the second indicator. The dashed red line marks the score of 0.5.

For both textures, all scoring curves reach over the 0.5 score threshold. Even the lowest of the curves (when either `Respiration` or `OC_RE6` are set 0) converge to 0.5, even when the second indicator of PC1 is at its minimum. This happens because the two indicators in both first PCs have almost equal weights. Thus, when `Bact.mass` reaches the maximum score, the final summary score is halved by the second indicator's minimum score (0).

For the marine clay principal component, the lowest five `Respiration` values (0.0 to 0.4) produce curves approximately 0.1 apart in score. Higher values of this parameter are reaching maximum scores and are hard to distinguish from each other, as they overlap almost completely. All curves converge when the `Bact.mass` value is close to 12.5.

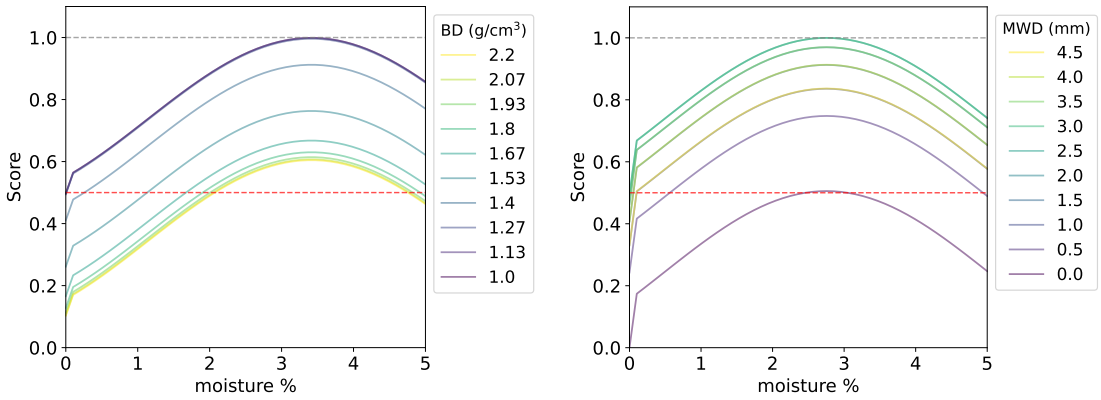
In the case of the sand texture in figure 5.12b, each curve reaches its specific maximum also around 12.5 µg/g of bacterial mass, similar to its counterpart. This is to be expected, as the individual scoring curve for `Bact.mass` converges to 1 at this value (figure 5.5a).

- **Physical Pillar**

Table 5.13 displays the physical PCs in the form of linear equations and figure 5.13 shows the shape of the physical summary scoring curves. With only 3 total indicators, the same approach as for the biological pillar applies. `moisture` is the common element in the first PC of both texture types, so this measurement is maintained as a variable on the x-axis.

M	$s(PC1) = [0.71 \cdot s(\texttt{moisture}) + 0.71 \cdot s(\texttt{BD})] \,/\, (0.71 + 0.71)$ $s(PC2) = 1 \cdot s(\texttt{MWD})$
S	$s(PC1) = [0.71 \cdot s(\texttt{moisture}) + 0.70 \cdot s(\texttt{MWD})] \,/\, (0.71 + 0.70)$ $s(PC2) = 1 \cdot s(\texttt{BD})$

TABLE 5.13: Formulas for scoring the principal components for the physical pillar by soil texture type. When scoring, all weights are taken with absolute value.



(A) PC1 of Marine Clay

(B) PC2 of Sand

FIGURE 5.13: Summary scoring curves for the first principal component of both soil textures for the **physical pillar**. Plotting is done relative to the `moisture` indicator, which is common in both principal components. The second indicator in each component is discretised. The Score axis is a function of `moisture` and the second indicator. The dashed red line marks the score of 0.5.

Unlike the previous pillar, the physical one scores its individual indicators using three distinct function types: more-is-better (the only type in the biological pillar), less-is-better and optimum, resulting in a considerably distinct summary scoring curve shape. Since `moisture` is maintained as a variable and its associated equation is an optimum value type, its general shape is retained for the summary indicator.

One immediate observation when comparing the two soil textures is that the colours of the lines are in opposite order. This is caused by the scoring function of bulk density (`BD`), which yields better scores for lower values of the parameter. Moreover, the worst-performing curve in the sand subset reaches a maximum score of 0.5, as expected when all weights of the PC are almost equal. Its equivalent in the marine clay texture goes up to 0.6 due to the `BD` scoring curve not converging to 0.

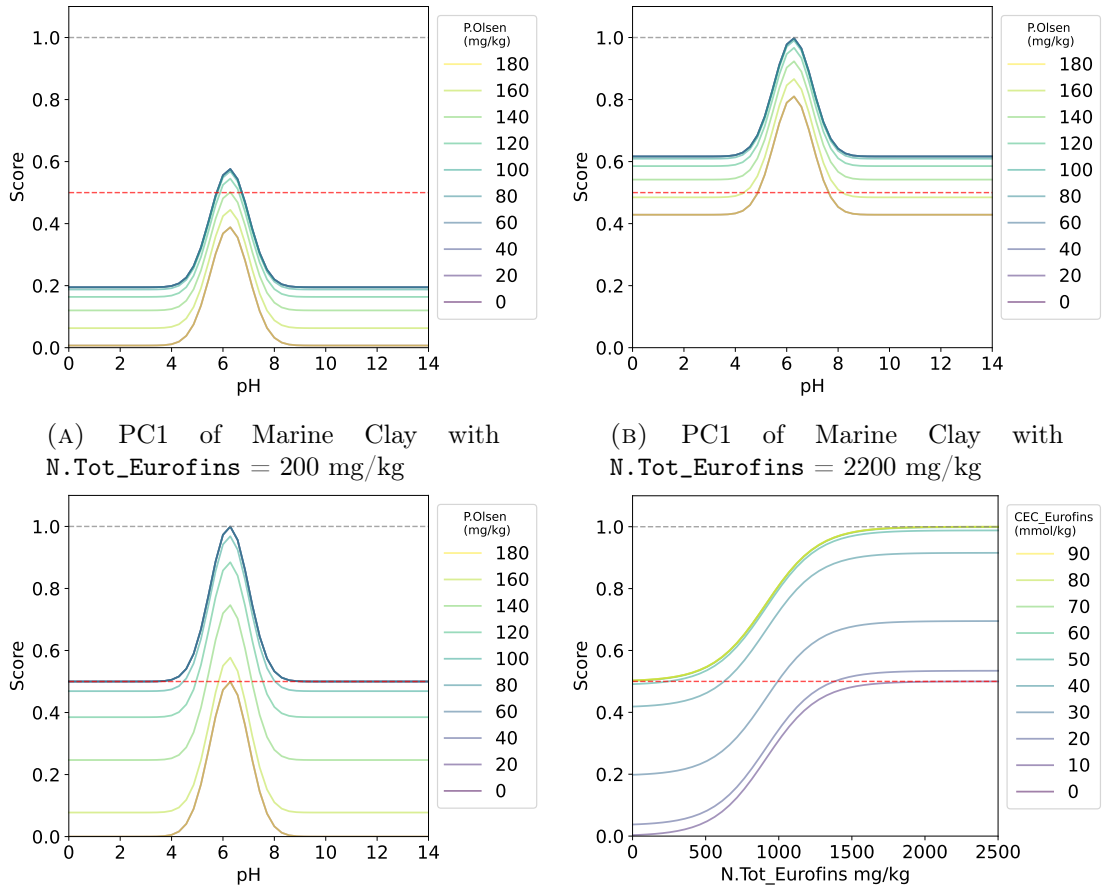
- **Chemical Pillar**

The equations of the chemical principal components are laid out in table 5.14 and figure 5.14 exemplifies how the summary scoring curves develop. The plots are generated only for *wheat* (crop is relevant when scoring the pH). Appendix G shows additional graphs for other crop types.

M	$s(\text{PC1}) = 0.63 \cdot s(\texttt{pH}) + 0.31 \cdot s(\texttt{P.Olsen}) + 0.71 \cdot s(\texttt{N.Tot\_E}) \;/\; (0.63 + 0.31 + 0.71)$ $s(\text{PC2}) = 1 \cdot s(\texttt{K\_E})$
S	$s(\text{PC1}) = 0.71 \cdot s(\texttt{pH}) + 0.71 \cdot s(\texttt{P.Olsen}) \;/\; (0.71 + 0.71)$ $s(\text{PC2}) = 0.71 \cdot s(\texttt{N.Tot\_E}) + 0.71 \cdot s(\texttt{CEC\_E}) \;/\; (0.71 + 0.71)$

TABLE 5.14: Formulas for scoring the principal components for the chemical pillar by soil texture type. `Eurofins` is replaced by `E` for convenience.

When analysing the **sand** texture, both principal components look very similar to the PC1s of the previous pillars. Equal absolute weights indicate all features contribute



(A) PC1 of Marine Clay with `N.Tot_Eurofins` = 200 mg/kg

(B) PC1 of Marine Clay with `N.Tot_Eurofins` = 2200 mg/kg

(C) PC1 of Sand

(D) PC2 of Sand

FIGURE 5.14: Summary scoring curves for the **chemical pillar**

equally to the final component score. For the first PC, the second indicator is phosphorus, an optimum value type of scoring function, like the pH. This can be seen in the overlap of the first (`P.olsen` = 0 mg/kg — scored 0) and last line plots (`P.Olsen` = 180 mg/kg – scored 0) in figure 5.14c, both being the worst-performing curves. The second PC (figure 5.14d) looks very similar to the biological summary curves, as both `N.Tot_Eurofins` and `CEC_Eurofins` have the same equation type (more-is-better).

The first principal component for the **marine clay** texture is different than all other curves analysed until now, as it is composed of three individual indicators. To be able to produce the associated plot, both `P.Olsen` and `N.Tot_Eurofins` are discretised, with the latter being able to take just one value per plot. The three indicators also have different quite different weights, unlike the other PCs where each indicator contributes equally to the score. This imbalance leads to scores surpassing the 0.5 line when two of the indicators are 0, or going under it when 2 features have maximum scores.

5.4 RQ3 - Visualising scores

As previously described in section 4.2, the adjusted explained variance is computed for all sparse principal components using formulas 4.11a and 4.11b. The results are displayed in table 5.15, separated according to texture type and summary indicator. When computing the final pillar score and the scoring heatmaps, these values serve as weights for the principal

components.

	BIO		PHYS		CHEM	
	PC1	PC2	PC1	PC2	PC1	PC2
M	0.437	0.333	0.495	0.333	0.422	0.250
S	0.448	0.333	0.462	0.333	0.290	0.300

TABLE 5.15: Adjusted percentages of explained variance for each summary indicator. Results are for all pillars and texture types (M = marine clay, S = sand).

Similar to the curve plots in the previous section, the heatmaps are visualised with only one variable per axis. The first principal component in every pillar will keep one measurement as a variable. The other indicator(s) will be sampled. The second summary indicator is made up of just one feature for most cases, therefore no adjustments are needed. The chemical PC2 for sand is the exception and has one feature sampled.

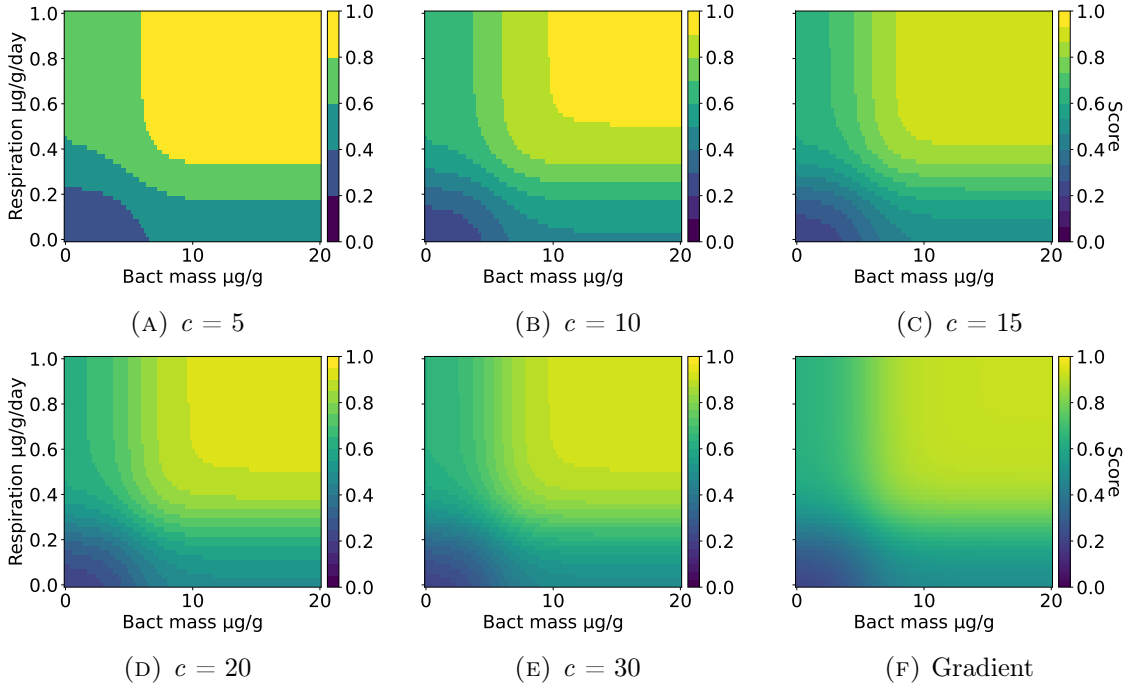
Adjusting heatmap granularity

Interpretability and ease of use are key factors in plotting a score heatmap. Human perception of a heatmap can be influenced by the number of colours, their intensity, and whether hard borders are in place or only gradual transitions between distinct scores are present. The same data is plotted by varying the number of colour divisions, to better compare the different ways of visually representing the score map (figure 5.15). By increasing the granularity of the colour-coded heatmaps, the accuracy with which a soil sample can be scored also improves. Figures 5.15a and 5.15b have sharp transitions, clearly differentiating between the score intervals and providing the user with a general idea of how the soil sample performs. While clarity is high, the abrupt changes and the reduced number of divisions can lead to an oversimplified representation of more sensitive scores.

More accurate scoring can be extracted from figures 5.15c and 5.15d, where each 0.1 score interval is represented through 2 to 3 shades. Borders are still relatively clear, with smoother transitions. More details on how the scores develop across the heatmap are visible, especially for scores below 0.6. These plots can potentially offer better insights and a more precise idea of how close the soil sample is to reaching a higher score step. The lower contrast between regions might alter the perception of scores, as the distinction between score milestones is not as emphasised compared to previous graphs.

Increasing the total amount of colour divisions to 30 (figure 5.15e) blurs the boundaries further, while a complete gradient (figure 5.15f) offers only a general idea of the score range. These heatmaps are better fitted for a more research-oriented scope, such as analysing how the health score is altered with small differences between samples. However, for an average end-user, these plots are not the best visual representation for interpreting soil health.

The threshold for classifying a soil sample as *healthy* or *not healthy* is not set, as it requires specific field knowledge and understanding. Considering the pillar score and visualising its position relative to possible improvements is currently the only criterion in the framework for evaluating the state of a soil sample. Due to this aspect, along with the scope of the assessment and the general interest of the stakeholders, the choice of granularity for the heatmaps is $c = 10$. This ensures every score level of 0.1 can be clearly distinguished, facilitating quick decision-making and providing a good overview of what changes could be made to increase the soil's health. While this heatmap has a lower accuracy than those with a higher c, it is better suited for quick evaluations and is easier to



(A) $c = 5$ (B) $c = 10$ (C) $c = 15$

(D) $c = 20$ (E) $c = 30$ (F) Gradient

FIGURE 5.15: Varying the number of colour divisions (c) when plotting scoring heatmaps. Data is from the biological pillar, sand texture, `OC_RE6` = 20 g/kg.

interpret. Further analysis and results are presented using this representation. However, the figures 5.15c or 5.15d can be more fitting for users who possess a deeper understanding of the soil and its underlying processes.

Heatmap transformation based on data values

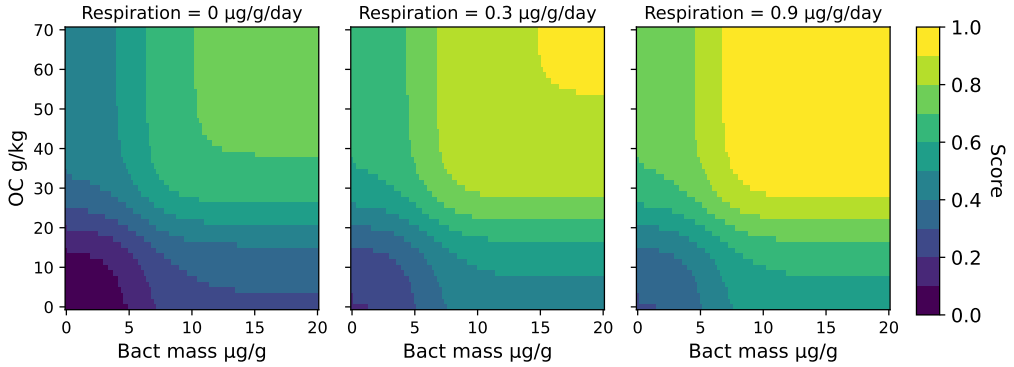
After deciding on the preferred granularity, scoring heatmaps for all pillars are generated and analysed. Tables 5.12, 5.13 and 5.14 have the formulas for scoring the three pillars.

- **Biological Pillar**

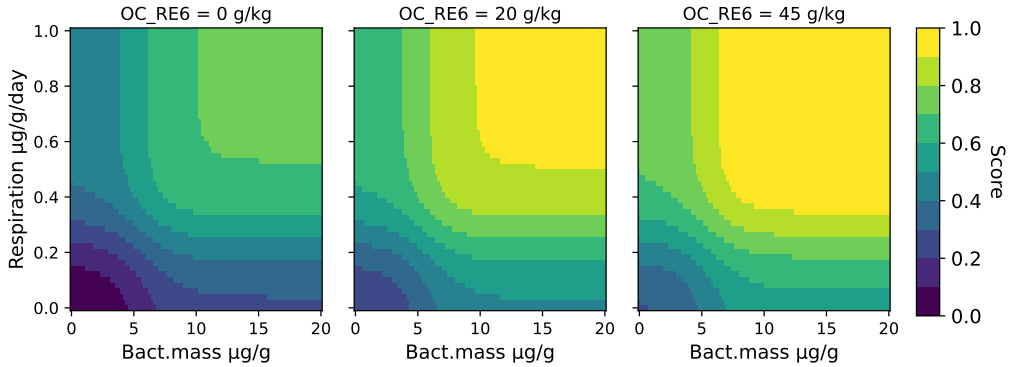
Figure 5.16 illustrates how the biological pillar score varies based on different values of two variables: `Respiration` for the marine clay texture (first row) and `OC_RE6` for the sand one (second row). As all biological indicators have a *more-is-better* type of scoring function, the score distribution is straightforward (the lower right corner has the lowest scores and the upper right corner, has the highest).

The discrepancies between textures are not immediately visible. The shapes of the separate score intervals are nearly identical, with only slight variations in the width of the intervals. Multiple factors produce similarly shaped curves, including the chosen granularity and the relatively equal SPCA weights for both PC1 and the indicator score equations.

When either of the second indicators — `Respiration` for marine clay and `OC_RE6` for sand — reach high values relative to when they reach their maximum score (as illustrated in figure 5.5), approximately half of the heatmap area shows a score above 0.9. Achieving a score below 0.2 is virtually impossible for medium values for the second indicator, even if the other two measurements are approaching 0. Attaching a moderate score to a severely unhealthy sample can potentially be a problem in scoring soil. However, in the case of the biological pillar, it is highly unlikely, if not impossible, for either of the indicators to



(A) Marine Clay texture



(B) Sand texture

FIGURE 5.16: Heatmaps for scoring the biological pillar for marine clay and sand. The second indicator in PC1 (`Respiration` for marine clay and `OC` for sand) is set as a constant in each first principal component. `OC` = Organic Carbon.

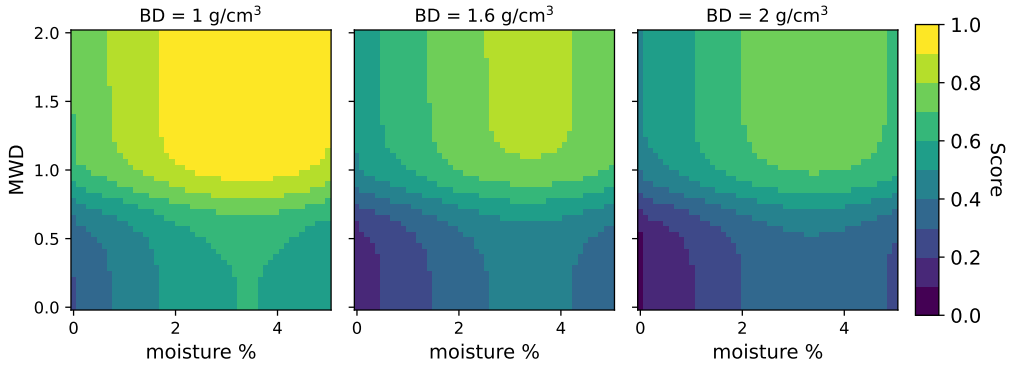
have a medium to high value. In contrast, the others are extremely low, as the indicators have closely related meanings (improbable to have high respiration but almost no organic carbon or bacteria in the soil).

- **Physical Pillar**

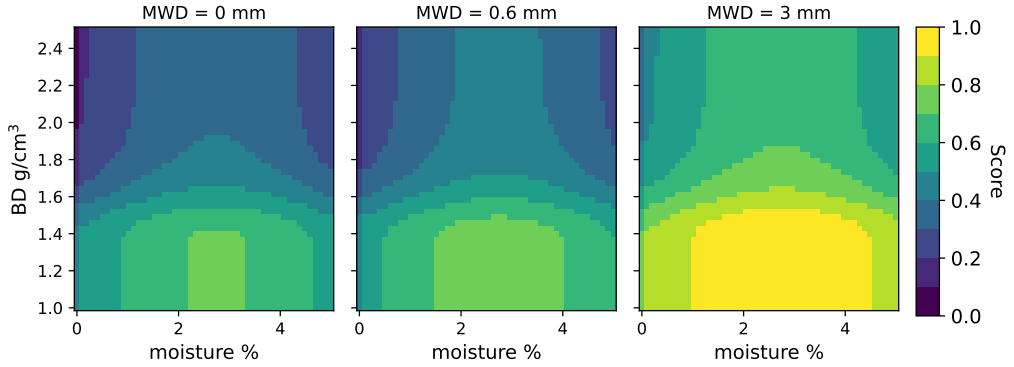
The heatmaps for the physical pillar in figure 5.17 show how scores develop with the change of the indicators. For both texture types, the `moisture` indicator is kept variable across the first principal component (the x-axis), while the bulk density (marine clay) or mean weight diameter (sand) are set as constants.

The shapes of the scoring areas are more complex than those of the biological pillar due to a higher variety in the equations used to extract the score. Increasing the values of the measurements does not necessarily lead to a better score, as `moisture` achieves perfect scores when reaching its contextual optimum, while lower values of `BD` are attributed to healthier soils.

For marine clay, the possible scores decrease as the bulk density of the soil increases. This is visible in the sand subsets as well in figure 5.17b, as the score value drops when bulk density is higher. For the sand subset, increasing the mean weight diameter of soil particles improves the resulting scores, suggesting that better aggregation of soil particles (larger MWD) improves the physical properties of sandy soils. The heatmaps for the two textures have the same are shaped the same, with the exception that they are vertically



(A) Marine Clay texture



(B) Sand texture

FIGURE 5.17: Scoring heatmaps for the physical pillar. Results are for both soil texture types. `BD` = bulk density and `MWD` = mean weight diameter.

flipped. This happens due to bulk density being on the y-axis in the sand textures, causing the map to be upside down since BD is scored *less-is-better*.

- **Chemical Pillar**

This pillar is more intricate than the biological and physical ones, as it has more than three indicators, with one of them (`pH`) being dependent on crop type instead of soil texture. To provide a concise visualisation, the plots in figures 5.18 and 5.19 are generated using the scoring parameters for **wheat**. Appendix G contains more examples of heatmaps generated for crops found in the VitalSoils dataset.

In the case of marine clay, the chemical pillar's first principal component is a linear combination of three indicators. Figure 5.18 illustrates how the possible scores for a soil sample change as nitrogen (N) and phosphorus (P) concentrations fluctuate. Higher nitrogen concentrations typically correspond to broader regions with higher scores, indicating that more nitrogen generally improves the health of the soil. In contrast, phosphorus follows an optimum value scoring pattern, where medium concentrations (50 mg/kg) perform best, leading to wider regions with higher scores compared to the other two phosphorus levels.

The sand soil texture has two indicators per principal component, meaning both axes have a variable indicator (pH for PC1 and nitrogen for PC2) and a constant one (phosphorus for PC1 and CEC for PC2). Figure 5.19 depicts how the scoring possibilities change with different value combinations of individual indicators. Like nitrogen in the previous

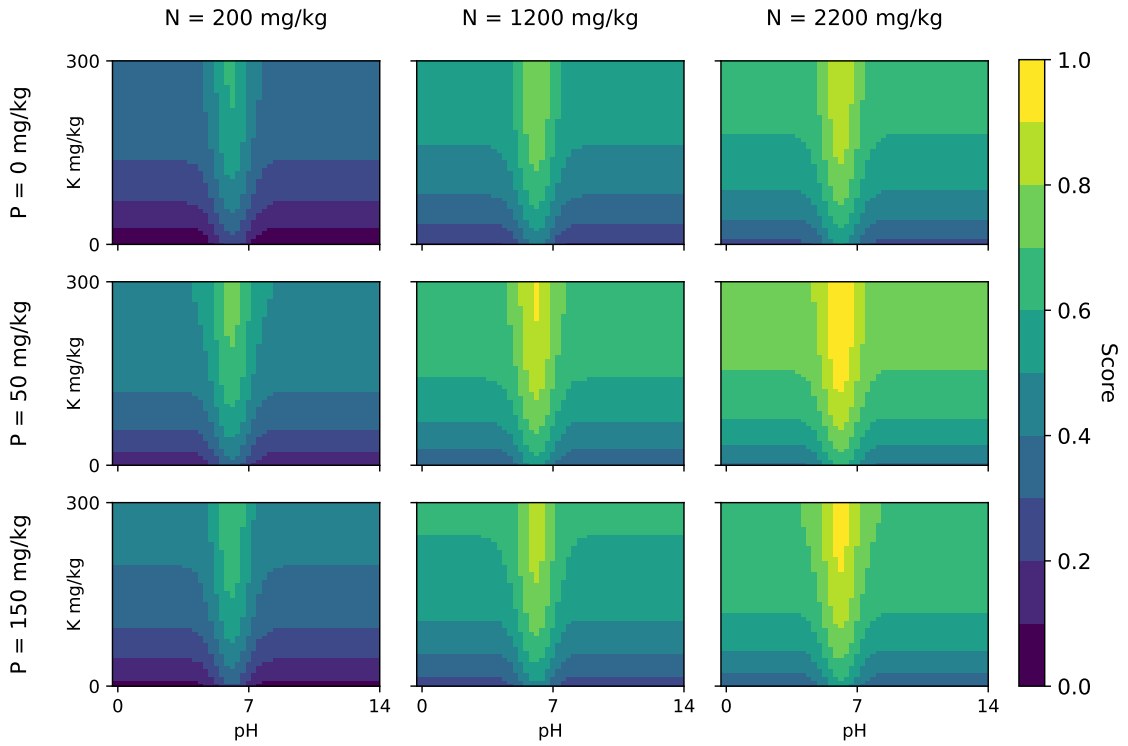


FIGURE 5.18: Scoring heatmaps for the chemical pillar, marine clay texture. N = Nitrogen (`N.Tot_Eurofins`), P = Phosphorus (`P.Olsen`), K = Potassium (`K_Eurofins`). `pH` is scored for wheat only.

texture type, higher values of CEC get better scores. Therefore, the heatmaps for sandy soils develop similarly.

The heatmap for the lowest values of CEC (10 mmol/kg) and Phosphorus (0 mg/kg) has the lowest possible scores from all pillars and textures, peaking at the 0.3 - 0.4 interval. The bottom half of the plot is dominated by a deep blue corresponding to the lowest score interval. For the other pillar-texture combinations, the second principal component is made out of just one indicator and plotting its full range produces all possible scores. This implies higher scores occur, which in combination with lower ones from PC1 would still yield medium values. In the current case, as the secondary indicators for both principal components are set to their values which produce minimum scores, the combination of PCs is also scored very poorly.

Due to how the pillar scores are computed, no plot displays the full range of scoring intervals in a single map. Medium values for bulk density get quite close, as seen in figure 5.17a when `BD` = 1.6 g/cm^3, having all but two intervals plotted (the highest and the lowest). This "lack of representation" of all intervals in just one plot is neither a positive nor negative aspect, as it greatly depends on the granularity of the scoring intervals and the accuracy needed for a specific analysis. These heatmaps aim to offer an overview of a soil sample's health by understanding what indicator(s) ought to be changed to improve the soil.

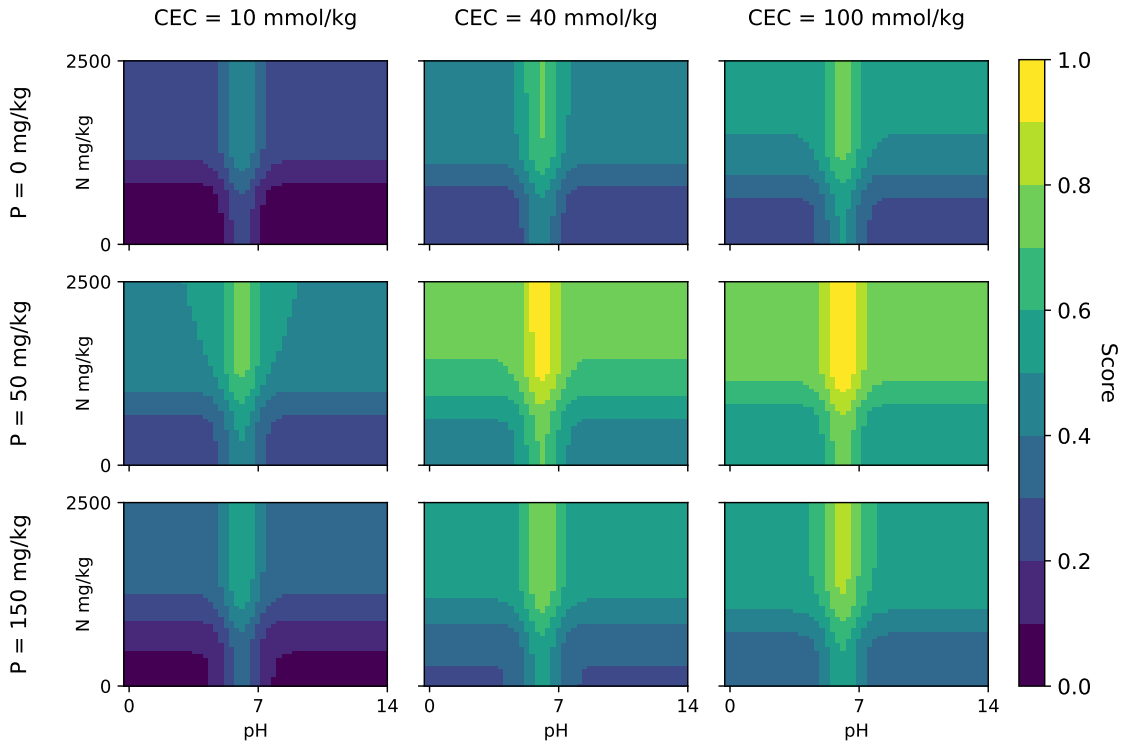


FIGURE 5.19: Scoring heatmaps for the chemical pillar, sand texture. CEC = Cation Exchange Capacity (`CEC_Eurofins`), P = Phosphorus (`P.Olsen`), N = Nitrogen (`N.Tot_Eurofins`). `pH` is scored for wheat only.

Automated scoring tool

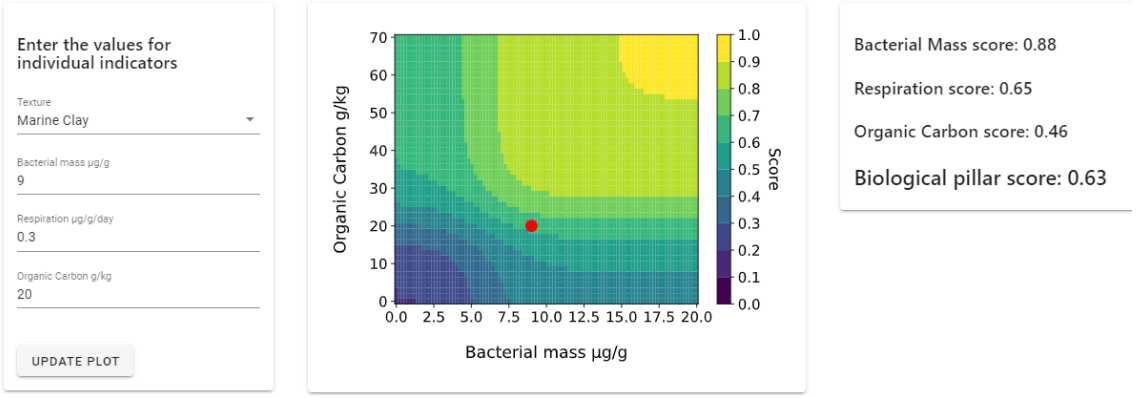


FIGURE 5.20: Demo of the scoring tool for the biological pillar.

Visualisation and testing should be possible outside of the development environment. A minimal web application is built with the help of the `solara` Python framework. The app can handle user input for the soil texture type and values for individual indicators. According to this information and the selected pillar, the program generates and displays the appropriate heatmap plot, with a red bullet indicating the score associated with the input. A separate widget displays the individual scores for the three indicators and the total pillar score. Figure 5.20 is a screenshot of the application interface for the biological

pillar. It assigns a pillar score of 0.63 to a marine clay sample, with 9 µg/g bacterial mass, respiration of 0.3 µg/g/day and 20 g/kg organic carbon.

5.5 RQ4 - Analysing the differences between soil textures

The discrepancies in behaviour between the marine clay and sand textures have been also discussed in the previous research questions and will be briefly summarised in this section, while also expanding on certain aspects.

After computing the correlation coefficients for each pillar, most indicator pairs exhibit similar values across textures. There are also cases when polarity changes across different texture types, suggesting a significant shift in how soil properties interact in these environments. The removal of redundant features differs only for the chemical pillar: the marine clay subset retains the zinc indicator, while the sand subset retains cation exchange capacity and aluminium measurements.

During SPCA tuning, the different values selected for the `alpha` parameter highlight another way texture influences the indicators. Marine clay samples have higher values across all pillars, resulting in sparser components. `alpha` was chosen to maximize information retention while ensuring the components remain as sparse as possible. This suggests that the selected indicators may better describe the behaviour of marine clay than sandy soils. However, a potential alternative explanation is the difference in sample sizes, with marine clay having approximately 78% more samples than sand. The evaluation metrics reveal mostly negligible differences between the two textures, with sand performing slightly better for all pillars except the physical one.

The extraction of principal components varies strongly from one texture to another. For the biological and physical pillars, the resulting weights are nearly identical from one texture to another. The distribution of indicators in PCs has almost no overlap for all pillars, making it difficult to make a consistent comparison. The chemical pillar eliminates some indicators when assigning weights with the most noticeable difference being that the sand subset completely disregards potassium (`K_Eurofins`), whereas marine clay assigns 0 weights to copper (`Cu_Eurofins`). Both textures deem iron (`Fe_Eurofins`) irrelevant in their principal components. As opposed to the other pillars, the choice of SPCA's `alpha` is the same for the two textures, thus both have the same degree of sparsity. This suggests different indicators better describe a each texture.

Overall, the two soil textures seem to exhibit slightly distinct but related behaviours, but the discrepancy in the number of samples attributed to each texture type most probably plays an important role in the different results.

Further analysis of cross-texture variability

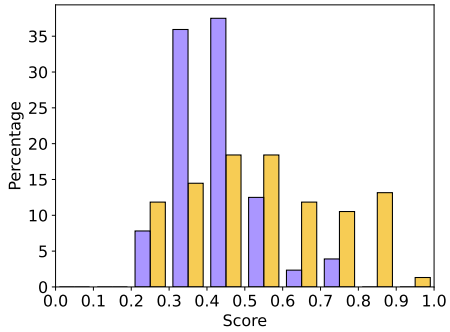
Each soil sample in the VitalSoils dataset is scored per pillar according to their texture type. Various statistic metrics are applied to the dataset and displayed in table 5.16. For all pillars, the marine clay samples reach a minimum lower than sandy soils, while sand gets higher maximum scores (the exception is the physical pillar, where both soil types reach perfect scores). This phenomenon is also reflected in the mean scores, especially within the chemical pillar, where sandy soils have almost 0.3 score points over marine clay. The standard deviation (σ) does not follow a consistent trend across different texture types. The most considerable difference is 0.11 points in the physical pillar, where the marine clay samples are more dispersed.

The histograms in figure 5.21 describe how the scores for each pillar are distributed

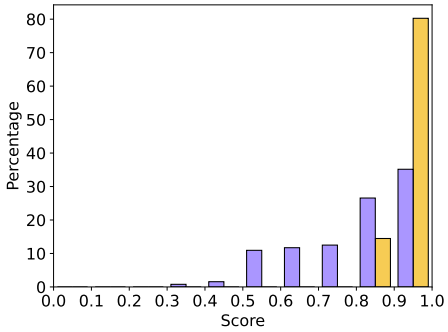
		min	max	mean	σ	var
BIO	**M**	0.22	0.77	0.43	0.11	0.01
	S	0.23	0.97	0.54	0.19	0.03
PHYS	**M**	0.40	1	0.82	0.15	0.02
	S	0.81	1	0.95	0.04	0.001
CHEM	**M**	0.32	0.73	0.52	0.09	0.008
	S	0.37	0.99	0.85	0.13	0.01

TABLE 5.16: Statistical metrics applied to the resulting pillar scores for both soil texture types

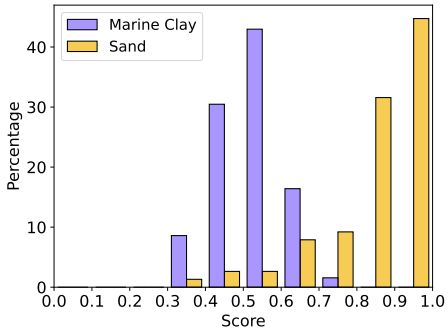
in percentages. The percentages are relative to the total number of samples of a certain texture. The marine clay texture scores preponderantly between 0.3 and 0.5 for the biological pillar, while sandy samples are more uniformly spread out across the scoring intervals. The physical pillar shows that both textures have most of the samples in the highest bin, with approximately 80% of sand scoring above 0.9. The marine clay scores in the chemical pillar are condensed between 0.3 and 0.7, whereas the sand has almost 75% of its samples in the two highest categories. In general, the sand samples score significantly better than their marine clay counterparts, a fact also confirmed by the mean scores from table 5.16.



(A) Biological Pillar



(B) Physical Pillar

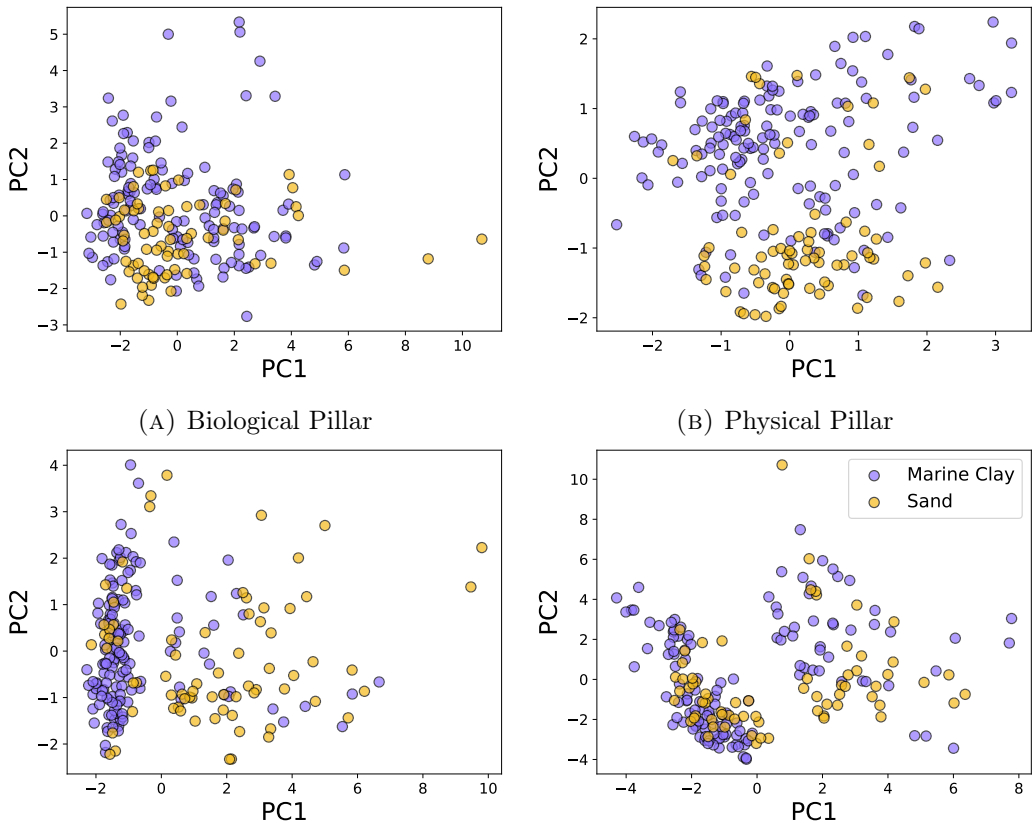


(C) Chemical Pillar

FIGURE 5.21: Distribution of scores per soil pillar in percentages split by texture type. The percentages are relative to the number of samples for a specific texture type.

The first two score bins are completely unpopulated regardless of the pillar, indicating no collected samples are in an extreme state of degradation.

In figure 5.22, SPCA with 2 components is applied to each complete pillar, without dividing the dataset into texture types first. `alpha` is not tuned and has the default value of 1. Each sample is then plotted with the newly extracted sparse components and coloured according to their texture type.



(A) Biological Pillar

(B) Physical Pillar

(C) Chemical Pillar

(D) All pillars combined

FIGURE 5.22: Principal components obtained through SPCA for each individual pillar and all of them combined. The points are split based on texture type (marine clay or sand).

A distinction between the two soil textures can be observed in the physical pillar in figure 5.22b. Due to the outliers, drawing a clear border between the textures is not feasible. However, the second principal component appears to make the distinction between the two soil types, with higher values being attributed to the marine clay texture, whereas the sand samples are mainly located below the 0 value of PC2. The chemical pillar (figure 5.22c) exhibits a similar behaviour, with the first PC being the deciding factor for the textures. Marine clay samples are condensed in a narrow vertical space between -2 and 0 for PC1, with only some points being spread out across the plot. The sand texture is more dispersed, but most associated points have the PC1 value > 0.

The biological pillar(fig. 5.22a) and the merged pillars (fig. 5.22d) prove hard to make a distinction between the soil textures. Most points are randomly dispersed across the plots and textures are intertwined. Combining all pillars seems to produce a cluster of marine clay samples. However, it is also populated by around half of the sandy points, therefore inconclusive if a real distinction can be made between textures.

Chapter 6

Discussion

RQ0. What soil properties are relevant to be used as individual indicators for soil health?

Indicators used for the evaluation of soil health differ greatly from paper to paper, with some researchers relying solely on a couple of measurements they already have, while others simply select a small subset they deem significant based on their domain knowledge. Other publications construct their minimum dataset from other (limited) literature pieces, but there is no general agreement over what set of indicators is the best for expressing soil health. Publications are not always consistent with the names of the indicators, and subtle differences exist also in the meaning of apparently equivalent measurements.

In this research, the indicators were chosen objectively based on the frequency with which they occurred in over 20 publications. The measurements were also aggregated under the same name and frequency count if their meaning was similar. This approach may be improved by requesting more specialised input and guidance, as some aggregated indicators could potentially be important as individuals, even if strongly connected. The choice of indicators is still contextual, depending strongly on specific soil properties (texture) and its environment (climate, annual precipitation, location).

The literature which provided the relevant indicators varies in terms of year of publishing, offering a mix of foundation knowledge and newer studies. However, not a lot of consistent variation exists across many of these papers, with many of them being based on common older research (such as the work of Doran and Parkin from 1997 [66]). This can lead to a lack of integration of newer methods and the acceptance of the same old (and perhaps redundant) indicators. As a tangible example, the latest version of SHAPE [67] (2024) argues in favour of using the POXC[1] measurement instead of other traditional methods of extracting carbon. The reasoning behind this choice is that not only is it faster and cheaper to measure than other biological indicators, but it is also more sensitive to changes in soil and environment than other carbon fractions. Although the VitalSoils dataset contains this particular measurement, it was not used as it was not frequently found in the literature. Correlation coefficients were computed for it and the other biological indicators, resulting in over 0.9 correlation with the other organic carbon columns.

[1] permanganate-oxidizable carbon, a measurement of active carbon

RQ1. How can related soil properties be grouped to reduce the problem of multidimensionality?

A soil sample is initially characterised by over 140 abiotic indicators, most of which need to be measured individually or gathered by interviewing farmers. By extracting relevant attributes from specialised literature, splitting the dataset by soil texture (marine clay or sand), removing strongly correlated features and applying dimensionality reduction techniques, a minimum dataset (MDS) was obtained in the form of *summary indicators* - linear combinations of columns for each pillar. Ultimately, the number of measurements needed in the summary indicators is reduced to 14, **10% of the initial measurements.**

Excluding redundant features based on strong association or applying factor analysis are not novelty approaches in the field of soil health (or quality), the most similar method being presented by Sparling and Schipper [41]. The authors start with only 14 initial indicators, a tenth of the current research's dataset. When assessing correlations, they consider a coefficient of 0.45 strong, no clear threshold is set and some indicators are removed based on the findings of other studies or by using criteria specific to their research scope. Moreover, their minimum dataset is composed of 7 indicators, half of their initial data, which explains 87% of the variance when 4 principal components (n) are extracted with PCA, but only 72% for 3 components and 56% for $n = 2$. Lima et al [68] retain 8 indicators for their MDS from a total of 29, a reduction of 73%. The authors examine the correlations between indicators but do not use strong association as a dismissal criterion. Instead, the dataset is minimised by selecting attributes with high weights in principal components (after applying PCA). Other studies use a similar approach, with the mention that they remove highly correlated features after discriminating based on PCA weights, with a threshold of 0.5 for Li et al. [69], whereas Zang et al. [70] set it to 0.7.

In the current research, the solution to achieving a minimum dataset follows a more objective and analytical trajectory, as attributes are dismissed based almost exclusively on numerical results, with more strict thresholds ($abs(r) \geq 0.75$). PCA is used in this context only for comparison purposes when evaluating sparse PCA (SPCA), an algorithm which does not seem to be used in any related research. Resulted principal components do not serve as a mean of removing even more features but as a way of reducing the dimensionality of a soil sample by condensing each pillar into two summary indicators. The number of remaining standalone measurements is considerably higher than the results in other studies (equal size to the initial dataset in the case of [41]), but they include more biological indicators and have the highest reduction rate compared to the starting dataset, achieving a removal rate of 90% regardless of soil texture.

RQ2. How can summary indicators be scored in a manner which expresses a soil sample's health?

Applying the SMAF scoring curves is not a novel approach, as it has been used in many other case studies. The addition to the field brought by answering this research question consists of computing the summary scoring curves for the summary indicators. Not only are the summary indicators sparser and easier to obtain due to the use of SPCA instead of normal PCA, but the resulting curves provide a general overview of the healthiness of a soil pillar. This is an important aspect for the stakeholders (ecologists and farmers, for example) as it greatly reduces the number of indicators they need to evaluate and individually score and analyse. Moreover, the weights associated with each indicator in a PC automatically scale their importance in the final score. Consulting summary indicators means looking at the state of a pillar as a whole, not only at the individual results, creat-

ing a better outline of the soil's condition and which pillar could potentially be improved. However, this logic is best applied to the biological and physical pillars, where most indicators are interconnected. For their chemical counterpart, it can be argued that looking at standalone scores for each indicator might be more beneficial, as the values of some elements (nitrogen, or phosphorus, for example) or the pH can be changed individually by enriching the soil with certain supplements.

The current scoring curves are not yet final, as some alterations were made to the `Respiration` thresholds. These changes were done according to the dataset values of the indicators. Moreover, the thresholds for the measurements scored by the normalisation equation 5.1 are specific to the case studies from the papers they were extracted from. Some uncertainties are also present for the available water capacity indicator, as the SMAF file denotes this measurement as being *too transitory to be useful* in the context of soil health.

Computing a score based on a weighted sum of individual scores is not a novelty, whether it relates to the pillar or a soil function. Usually, the associated weights are just equally distributed among the components. The current framework constructs the summary indicators based on SPCA, which automatically determines each factor's importance in describing a pillar.

RQ3. What is the most suitable manner of expressing the results of the assessment?

The chosen approach for visualising the scores is as a heatmap. While unable to represent all possible scores of a pillar in one static plot, by discretising all but one variable per axis, the heatmaps are an easy-to-understand and interpretable representation of soil health. The choice of which indicator remains variable on a certain axis influences the visualisation, as the main shape of the scoring intervals is dictated by that of the indicator's scoring curve. The chemical pillar prioritises the pH on the x-axis, resulting in a bell-shaped colourisation (reversed shape due to the negative sign of the pH in the principal components). If the pH is replaced by phosphorus, for example, the heatmap will resemble the shape of its scoring curve, as seen in figure G.2.

CASH [13] also uses a heatmap palette (red to green) to differentiate between scores. Similarly to this research, the step is a tenth of the maximum score and each interval has its own shade. The main difference is what is scored, as CASH displays this heatmap as the background of individual scoring curves. The colours are distributed as stripes corresponding to each interval, only aiding in attaching a score to a specific portion of the curve. The heatmaps created in the current study are designed to evaluate an entire pillar and to provide the user with insight into what soil property needs improvement and how a pillar's components influence the healthiness of a sample.

The proposed representation of scores has not been tested by stakeholders. No user study was done on the granularity degree of the heatmaps, however, the web app can be easily updated to allow a user to modify the number of shades per score interval. As for the concept of heatmaps, although relatively intuitive and easy to read, farmers or ecologists might have different needs or interests which are not accurately represented through this type of plot. We deem the current solution accurate from a *proof of concept* approach. Moreover, combining as many soil properties into a single indicator and providing the option to visually analyse the results is unconventional in the soil health science field, as no known papers offer similar results.

RQ4. How does soil texture type influence the indicators and the scoring?

After analysing each research step from the point of view of texture type, it is obvious that certain differences occur between marine clay and sand samples. The relevance of these differences is hard to quantify, as marine clay has considerably more available samples in the dataset.

When comparing pillar scores, sandy soils perform significantly better. This may be caused by multiple factors, such as certain biases in choosing the fields for sampling, incorrectly labelling the textures in the dataset or perhaps the sand soils are inherently healthier in The Netherlands area.

Dimensionality reduction techniques also fail to make a very clear distinction between the two textures, with just a slight differentiation of sand and marine clay visible in the physical and chemical pillars.

Limitations

The summary indicators obtained through this approach are characteristic of the area of The Netherlands, as all samples have been collected on this territory. Generality is not yet achievable with such a limited selection of samples and also due to the highly contextual parameters. The VitalSoils dataset contains a relatively small number of samples, which limits the general applicability.

From all indicators considered relevant in the literature (approximately 30, as some encapsulate various measurements), 21 could be matched with the dataset. Further, the chemical pillar was reduced from over 10 properties to only 5, due to a lack of scoring curves.

Applying the SMAF equations requires additional contextual parameters, such as climate, soil suborders or mineralogy. These factors are not explicitly stated in the dataset, and while most of them can be logically deduced, it does not offer an analytically sound way of accurately computing the scores. As an example, the soil suborder is needed to determine the organic matter class in the framework. There is no related information in the dataset other than texture classes, so the class was determined based on what it symbolises: high or low organic matter.

After applying DRTs, the results prove to be highly volatile. Across different textures, the principal components have little in common. The weights attributed to the same indicator vary greatly (mainly in the chemical pillar), the distribution of the measurements into principal components has almost no overlap, and some properties might be retained in one texture and removed in the other. The change of `alpha` also results in considerable changes in the structure of the PCs, not just in the total explained variance. All these aspects make the results subjective to the dataset, with a high chance of very different weights occurring when applying the same method to different sets of samples.

Chapter 7

Conclusion

The proposed framework offers an improved method for determining a minimum dataset of indicators which can accurately describe the state of a soil sample. Extensive literature review and the use of adequate statistical methods and correlation thresholds provide a more robust approach to eliminating redundant indicators, rather than relying solely on domain knowledge or just the available features. If basic dimensionality reduction techniques are used in most other studies for creating a minimum dataset, Sparse PCA is utilised here to compact each pillar into two interpretable dimensions while retaining 40% to 80% of variability, depending on the pillar. Based on these dimensions, each of the three pillars is evaluated by integrating the individual scoring equations for each corresponding measurement. The proposed framework also accommodates and provides an interpretable visualisation for the stakeholders in the form of a score heatmap. This particular plotting option can depict the state of a soil sample and possible directions of improvement based on raw values for individual indicators. While some differences occur between different soil textures at all steps of the pipeline, they also fluctuate considerably. The marine clay samples scored, on average, much lower than their sandy counterparts. However, the analysis is rather inconclusive for the extent to which they alter the importance of indicators or scoring.

Overall, this research proposes a new framework for assessing soil health in an analytically sound manner, which can be easily extended to be adequate for more soil types or different datasets.

Future work

Originally from the '90, SMAF is quite an old framework. While innovative for its time, it was created based on a limited amount of data, using soil samples only from the North-West of the USA. SHAPE, its successor, offers a more actual approach to determining scoring curves. Its models are trained on more data and it allows users to see the potential of the soil for a specific measurement. Currently, SHAPE can evaluate only a limited number of indicators. However, with the addition of more measurements, the framework can be implemented into the current research, potentially improving the performance and accuracy.

As previously mentioned, the contextual parameters required by SMAF (and SHAPE) are not determined analytically. Specialised input from specialists in ecology is needed to confirm (or correct) the choice of values of the parameters or to offer a better method for determining them.

To further validate and adjust the results and the visualisation, input from the stakeholders needs to be gathered and implemented. This requires interviewing farmers and

requesting the opinion of ecologists on what is working, what should be changed and what is unclear in the current framework.

Ultimately, the development of a fast, reliable and user-friendly web application can aid in the usability of the framework. Thus, farmers, lawmakers or scientists from various fields can simply input the soil measurements and receive the score of a sample, being able to evaluate its health. Moreover, with the help of specialists, the application can be extended to offer possible ways of enhancing each pillar of the soil, if needed.

Bibliography

[1] European Commission, Directorate-General for Research, Innovation, C Veerman, T Pinto Correia, C Bastioli, B Biro, J Bouma, E Cienciala, B Emmett, E Frison, A Grand, L Hristov, Z Kriaučiūnienė, M Pogrzeba, J Soussana, C Vela, and R Wittkowski. *Caring for soil is caring for life*. Publications Office, 2020. doi:10.2777/821504.

[2] European Commission, Directorate-General for Research, and Innovation. *European Green Deal – Research & innovation call*. Publications Office of the European Union, 2021. doi:10.2777/33415.

[3] European Commission and Directorate-General for Environment. *EU Soil Strategy for 2030: Reaping the benefits of healthy soils for people, food, nature and climate*. Publications Office, 2021. URL: https://environment.ec.europa.eu/publications/eu-soil-strategy-2030_en.

[4] European Commission and Directorate-General for Environment. *Proposal for a Directive on Soil Monitoring and Resilience*. 2023. URL: https://environment.ec.europa.eu/topics/soil-and-land/soil-health_en.

[5] Richard Bardgett. *The Biology of Soil: A Community and Ecosystem Approach*. Oxford University Press, 2005. doi:10.1093/acprof:oso/9780198525035.001.0001.

[6] John W. Doran and Michael R. Zeiss. "Soil health and sustainability: managing the biotic component of soil quality". *Applied Soil Ecology*, 15(1):3–11, 2000. Special issue: Managing the Biotic component of Soil Quality. doi:10.1016/S0929-1393(00)00067-6.

[7] M.S. Coyne, E.M. Pena-Yewtukhiw, J.H. Grove, A.C. Sant'Anna, and Domingo Mata-Padrino. "Soil Health – It's Not All Biology". *Soil Security*, 6:100051, 02 2022. doi:10.1016/j.soisec.2022.100051.

[8] Johannes Lehmann, Deborah Bossio, Ingrid Kögel-Knabner, and Matthias Rillig. "The concept and future prospects of soil health". *Nature Reviews Earth & Environment*, 1:1–10, 08 2020. doi:10.1038/s43017-020-0080-8.

[9] C. T. Kraamwinkel, A. Beaulieu, and T. et al. Dias. "Planetary limits to soil degradation". *Communications Earth & Environment*, 2:249, 2021. doi:10.1038/s43247-021-00323-3.

[10] E. C. Brevik, A. Cerdà, J. Mataix-Solera, L. Pereg, J. N. Quinton, J. Six, and K. Van Oost. "The interdisciplinary nature of SOIL". *SOIL*, 1(1):117–129, 2015. doi:10.5194/soil-1-117-2015.

[11] Susan S. Andrews, Douglas L. Karlen, and Cynthia A. Cambardella. "The Soil Management Assessment Framework". *Soil Science Society of America Journal*, 68(6):1945–1962, 2004. doi:10.2136/sssaj2004.1945.

[12] Márcio R Nunes, Kristen S Veum, Paul A Parker, Scott H Holan, Douglas L Karlen, Joseph P Amsili, Harold M van Es, Skye A Wills, Cathy A Seybold, and Thomas B Moorman. "The soil health assessment protocol and evaluation applied to soil organic carbon". *Soil Science Society of America Journal*, 85(4):1196–1213, 2021. doi:10.1002/saj2.20244.

[13] B.N. Moebius-Clune, D.J. Moebius-Clune, B.K. Gugino, O.J. Idowu, R.R. Schindelbeck, A.J. Ristow, H.M. van Es, J.E. Thies, H.A. Shayler, M.B. McBride, K.S.M. Kurtz, D.W. Wolfe, and G.S. Abawi. *Comprehensive Assessment of Soil Health – The Cornell Framework, Edition 3.2*. Geneva, NY, 2016.

[14] European Environment Agency. *Soil monitoring in Europe – Indicators and thresholds for soil quality assessments*. Publications Office of the European Union, 2023. URL: https://www.eea.europa.eu/publications/soil-monitoring-in-europe.

[15] Abhijit Datta. "Genetic engineering for improving quality and productivity of crops". *Agriculture & Food Security*, 2(15), 2013. doi:10.1186/2048-7010-2-15.

[16] Soumik Banerjee and Marcel G.A. van der Heijden. "Soil microbiomes and one health". *Nat Rev Microbiol*, 21:6–20, 2023. doi:10.1038/s41579-022-00779-w.

[17] Food and Agriculture Organization (FAO). "Healthy Soils are the Basis for Healthy Food Production", 2015. URL: https://www.fao.org/documents/card/en/c/645883cd-ba28-4b16-a7b8-34babbb3c505/.

[18] Hannah Ritchie and Max Roser. "Half of the world's habitable land is used for agriculture". *Our World in Data*, 2019. https://ourworldindata.org/global-land-for-agriculture.

[19] FAO, ITPS, GSBI, SCBD, and EC. "State of Knowledge of Soil Biodiversity - Status, Challenges and Potentialities, Report 2020". Technical report, FAO, 2020. doi:10.4060/cb1928en.

[20] World Meteorological Organization (WMO). *State of the Global Climate 2023*. World Meteorological Organization, 2024. URL: https://library.wmo.int/idurl/4/68835.

[21] UNFCCC. "Policies and measures". In *Paris Climate Change Conference - November 2015*, November 2018. URL: https://unfccc.int/process-and-meetings/the-paris-agreement.

[22] United States Federal Government. *Federal Sustainability Plan*. Office of the Federal Chief Sustainability Officer (CSO), 2021. URL: https://www.sustainability.gov/federalsustainabilityplan/index.html.

[23] Max Garzon, Ching-Chi Yang, Deepak Venugopal, Nirman Kumar, Kalidas Jana, and Lih-Yuan Deng. *Dimensionality Reduction in Data Science*, volume 766. Springer, 2022. doi:10.1007/978-3-031-05371-9.

[24] Scikit-learn. "Sparse principal components analysis". Accessed: 2024-09-09. URL: https://scikit-learn.org/stable/modules/decomposition.html#sparsepca.

[25] Eugen Ulea, Florin Lipşa, Florea Andreea, Filipov Feodor, and Evelina Morari. "Diversity of soil bacteria as indicator of soil pollution in Moldavia region, Romania". *Environmental engineering and management journal*, 16:879–889, 07 2017. doi:10.30638/eemj.2017.089.

[26] A.H.C. van Bruggen and A.M. Semenov. "In search of biological indicators for soil health and disease suppression". *Applied Soil Ecology*, 15(1):13–24, 2000. Special issue: Managing the Biotic component of Soil Quality. doi:10.1016/S0929-1393(00)00068-8.

[27] F.T. De Vries, E. Thébault, M. Liiri, K. Birkhofer, M.A. Tsiafouli, L. Bjørnlund, H. Bracht Jørgensen, M.V. Brady, S. Christensen, P.C. De Ruiter, et al. "Soil food web properties explain ecosystem services across European land use systems". *Proceedings of the National Academy of Sciences*, 110(35):14296–14301, 2013. doi:10.1073/pnas.1305198110.

[28] Natural Resources Conservation Service (NRCS). "Maryland Soil Health Card". Online, 2018. Accessed on: Feb 2024. URL: https://mda.maryland.gov/resource_conservation/counties/SoilHealthCard.pdf.

[29] Natural Resources Conservation Service (NRCS). "Cropland in-field soil health assessment guide". Online, January 2021. URL: https://www.nrcs.usda.gov/sites/default/files/2022-10/Cropland_InField_Soil_Health_Assessment_Guide.pdf.

[30] Natural Resources Conservation Service (NRCS). "Montana Cropland Soil Health Assessment Card", 2022. URL: https://www.nrcs.usda.gov/sites/default/files/2022-11/Montana-Cropland-Soil-Health%20Assessment-Card.pdf.

[31] Woods End Laboratories. "Solvita soil health tests". Woods End Laboratories, Inc.: Mount Vernon, ME, USA, 2021. Accessed on: Feb 2024. URL: https://solvita.com/soil/.

[32] Ward Laboratories. "Haney test". Ward Laboratories, Inc.: Kearney, NE, USA, 2021. Accessed on: Feb 2024. URL: https://www.nrcs.usda.gov/sites/default/files/2022-09/HaneyTest.pdf.

[33] A. Fowler, B. Basso, F. Maureira, N. Millar, R. Ulbrich, and W. F. Brinton. "Spatial patterns of historical crop yields reveal soil health attributes in US Midwest fields". *Scientific Reports*, 14(1):465, 2024. doi:10.1038/s41598-024-51155-y.

[34] Jordon Wade, Steve W. Culman, Caley K. Gasch, Cristina Lazcano, Gabriel Maltais-Landry, Andrew J. Margenot, Tvisha K. Martin, Teal S. Potter, Wayne R. Roper, Matthew D. Ruark, Christine D. Sprunger, and Matthew D. Wallenstein. "Rigorous, empirical, and quantitative: a proposed pipeline for soil health assessments". *Soil Biology and Biochemistry*, 170:108710, 2022. doi:10.1016/j.soilbio.2022.108710.

[35] Laís Coutinho Zayas Jimenez, Hermano Melo Queiroz, Maurício Roberto Cherubin, and Tiago Osório Ferreira. "Applying the soil management assessment framework (SMAF) to assess mangrove soil quality". *Sustainability*, 14(5):3085, 2022. doi:10.3390/su14053085.

[36] Maurício Roberto Cherubin, Cássio Antônio Tormena, and Douglas L. Karlen. "Soil quality evaluation using the soil management assessment framework (SMAF) in Brazilian oxisols with contrasting texture". *Revista Brasileira de Ciencia do Solo*, 41, 2017. doi:10.1590/18069657rbcs20160148.

[37] Maurício Roberto Cherubin, Douglas L. Karlen, André L. C. Franco, Carlos Eduardo Pellegrino Cerri, Cássio Antônio Tormena, and Carlos Clemente Cerri. "A Soil Management Assessment Framework (SMAF) Evaluation of Brazilian Sugarcane Expansion on Soil Quality". *Soil Science Society of America Journal*, 80:215–226, 2016. URL: https://api.semanticscholar.org/CorpusID:3502033.

[38] Ferrarini Andrea, Claudio Bini, and Stefano Amaducci. "Soil and ecosystem services: Current knowledge and evidences from Italian case studies". *Applied Soil Ecology*, 123:693–698, 2018. HUMUSICA 3 - Reviews, Applications, Tools. doi:10.1016/j.apsoil.2017.06.031.

[39] Debarati Bhaduri, Tapan Purakayastha, Ashok Patra, and Debashis Chakraborty. "Evaluating soil quality under a long-term integrated tillage-water-nutrient experiment with intensive rice-wheat rotation in a semi-arid Inceptisol, India". *Environmental monitoring and assessment*, 186:2535–2547, 04 2014. doi:10.1007/s10661-013-3558-8.

[40] ACR Lima, L Brussaard, MR Totola, WB Hoogmoed, and RGM De Goede. "A functional evaluation of three indicator sets for assessing soil quality". *Applied Soil Ecology*, 64:194–200, 2013. doi:10.1016/j.apsoil.2012.12.009.

[41] GP Sparling and LA Schipper. "Soil quality at a national scale in New Zealand". *Journal of environmental quality*, 31(6):1848–1857, 2002. doi:10.2134/jeq2002.1848.

[42] M.A. Arshad and S. Martin. "Identifying critical limits for soil quality indicators in agro-ecosystems". *Agriculture, Ecosystems Environment*, 88(2):153–160, 2002. Soil Health as an Indicator of Sustainable Management. doi:10.1016/S0167-8809(01)00252-3.

[43] Ana Cláudia Rodrigues de Lima, Willem Hoogmoed, and Lijbert Brussaard. "Soil quality assessment in rice production systems: establishing a minimum data set". *Journal of Environmental Quality*, 37(2):623–630, 2008. doi:10.2134/jeq2006.0280.

[44] MK Shukla, R Lal, and Mç Ebinger. "Determining soil quality indicators by factor analysis". *Soil and tillage research*, 87(2):194–204, 2006. doi:10.1016/j.still.2005.03.011.

[45] Jingyu Zhang, Miles Dyck, Sylvie A Quideau, and Charlotte E Norris. "Assessment of soil health and identification of key soil health indicators for five long-term crop rotations with varying fertility management". *Geoderma*, 443:116836, 2024. doi:10.1016/j.geoderma.2024.116836.

[46] Sophie Q van Rijssel, GF Veen, Guusje J Koorneef, JMT Bakx-Schotman, Freddy C Ten Hooven, Stefan Geisen, and Wim H van Der Putten. "Soil microbial diversity and community composition during conversion from conventional to organic agriculture [dataset]", 2022. URL: https://doi.org/10.5061/dryad.00000005t.

[47] Sophie Q van Rijssel, GF Veen, Guusje J Koorneef, JMT Bakx-Schotman, Freddy C Ten Hooven, Stefan Geisen, and Wim H van Der Putten. "Soil microbial diversity and community composition during conversion from conventional to organic agriculture". *Molecular Ecology*, 31(15):4017–4030, 2022. URL: https://doi.org/10.1111/mec.16571.

[48] Renato Cordeiro Amorim. "Feature Relevance in Ward's Hierarchical Clustering Using the Lp Norm". *J. Classif.*, 32(1):46–62, apr 2015. doi:10.1007/s00357-015-9167-1.

[49] Ehsan Amid and Manfred K. Warmuth. "TriMap: Large-scale Dimensionality Reduction Using Triplets", 2022. arXiv:1910.00204.

[50] Leland McInnes, John Healy, and James Melville. "Umap: Uniform manifold approximation and projection for dimension reduction", 2020. arXiv:1802.03426.

[51] Laurens van der Maaten and Geoffrey E. Hinton. "Visualizing Data using t-SNE". *Journal of Machine Learning Research*, 9:2579–2605, 2008. URL: https://api.semanticscholar.org/CorpusID:5855042.

[52] William Nick, Joseph Shelton, Gina Bullock, Albert Esterline, and Kassahun Asamene. "Comparing dimensionality reduction techniques", 2015. doi:10.1109/SECON.2015.7132997.

[53] Laurens van der Maaten, Eric Postma, and H. Herik. "Dimensionality reduction: A comparative review". *Journal of Machine Learning Research - JMLR*, 10, 01 2007.

[54] J. Camacho, A.K. Smilde, E. Saccenti, and J.A. Westerhuis. "All sparse PCA models are wrong, but some are useful. Part I: Computation of scores, residuals and explained variance". *Chemometrics and Intelligent Laboratory Systems*, 196:103907, January 2020. doi:10.1016/j.chemolab.2019.103907.

[55] Cosma Shalizi. "Lecture 8: Mathematics and Interpretation of Principal Components". Lecture, Statistics 36-350, Carnegie Mellon University, October 2006. URL: https://www.stat.cmu.edu/~cshalizi/350-2006/lecture-08.pdf.

[56] Jarkko Venna and Samuel Kaski. "Local Multidimensional Scaling with Controlled Tradeoff Between Trustworthiness and Continuity". 2005. URL: https://api.semanticscholar.org/CorpusID:16318128.

[57] Spyridon Stasis, Ryan Stables, and Jason Hockman. "Semantically Controlled Adaptive Equalisation in Reduced Dimensionality Parameter Space". *Applied Sciences*, 6:116, 04 2016. doi:10.3390/app6040116.

[58] "Quality assessment of dimensionality reduction: Rank-based criteria". *Neurocomputing*, 72(7):1431–1443, 2009. Advances in Machine Learning and Computational Intelligence. doi:10.1016/j.neucom.2008.12.017.

[59] Haipeng Shen and Jianhua Z Huang. "Sparse principal component analysis via regularized low rank matrix approximation". *Journal of multivariate analysis*, 99(6):1015–1034, 2008. doi:10.1016/j.jmva.2007.06.007.

[60] Hui Zou, Trevor Hastie, and Robert Tibshirani. "Sparse principal component analysis". *Journal of computational and graphical statistics*, 15(2):265–286, 2006. doi:10.1198/106186006X113430.

[61] Queensland Government. "Soil pH", 2023. URL: https://www.qld.gov.au/environment/land/management/soil/soil-properties/ph-levels.

[62] European Commission. "Commission Decision (EU) 2022/1244". Commission Decision EU) 2022/1244, European Commission, 2022. Notified under document C(2022) 4758. URL: https://eur-lex.europa.eu/eli/dec/2022/1244/oj.

[63] G. Tóth, T. Hermann, M.R. Da Silva, and L. Montanarella. "Heavy metals in agricultural soils of the European Union with implications for food safety". *Environment International*, 88:299–309, 2016. doi:10.1016/j.envint.2015.12.017.

[64] Douglas L Karlen and Diane E Stott. "A framework for evaluating physical and chemical indicators of soil quality". *Defining soil quality for a sustainable environment*, 35:53–72, 1994. doi:10.2136/sssaspecpub35.c4.

[65] Andrea Ferrarini, Flavio Fornasier, and Claudio Bini. *Development of a Soil Health Index based on the ecological soil functions for organic carbon stabilization with application to alluvial soils of northeastern Italy*, page 216. 04 2014. doi:10.3920/978-90-8686-788-2_9.

[66] John W. Doran and Timothy B. Parkin. "Quantitative Indicators of Soil Quality: A Minimum Data Set". In John W. Doran and Alice J. Jones, editors, *Methods for Assessing Soil Quality*, volume 49, chapter 2. SSSA Special Publications, 1997. doi:10.2136/sssaspecpub49.c2.

[67] Márcio R Nunes, Kristen S Veum, Paul A Parker, Scott H Holan, Joseph P Amsili, Harold M van Es, Skye A Wills, Cathy A Seybold, and Douglas L Karlen. "SHAPEv1. 0 Scoring curves and peer group benchmarks for dynamic soil health indicators". *Soil Science Society of America Journal*, 88(3):858–875, 2024. doi:10.1002/saj2.20668.

[68] Ana Cláudia Rodrigues de Lima, Willem Hoogmoed, and Lijbert Brussaard. "Soil quality assessment in rice production systems: establishing a minimum data set". *Journal of Environmental Quality*, 37(2):623–630, 2008. doi:10.2134/jeq2006.0280.

[69] Guilin Li, Jie Chen, Zhiying Sun, and Manzhi Tan. "Establishing a minimum dataset for soil quality assessment based on soil properties and land-use changes". *Acta ecologica sinica*, 27(7):2715–2724, 2007. doi:10.1016/S1872-2032(07)60059-6.

[70] Mingli Zang, Xiaodong Wang, Yunling Chen, and Seyedeh Ensieh Faramarzi. "Estimation of soil health in the semi-arid regions of northwestern Iran using digital elevation model and remote sensing data". *Environmental Monitoring and Assessment*, 196(4):353, 2024. doi:10.1007/s10661-024-12527-z.

[71] Mingxin Guo. "Soil Health Assessment and Management: Recent Development in Science and Practices". *Soil Systems*, 5:61, 10 2021. doi:10.3390/soilsystems5040061.

[72] Bhoopander Giri and Ajit Varma. *Soil Health*. 01 2020. doi:10.1007/978-3-030-44364-1.

[73] Hanna Williams, Tino Colombi, and Thomas Keller. "The influence of soil management on soil health: An on-farm study in southern Sweden". *Geoderma*, 360:114010, 2020. doi:10.1016/j.geoderma.2019.114010.

[74] Muhamad A Arshad and Gerald M Coen. "Characterization of soil quality: physical and chemical criteria". *American Journal of Alternative Agriculture*, 7(1-2):25–31, 1992. doi:10.1017/S0889189300004410.

[75] Natural Resources Conservation Service (NRCS). "Soil health card". https://www.nrcs.usda.gov/sites/default/files/2022-10/Cropland_InField_Soil_Health_Assessment_Guide.pdf.

[76] Indian Council of Agricultural Research. "Leucerne". AICRP on Forage Crops, Indian Council of Agricultural Research. URL: https://aicrponforagecrops.icar.gov.in/pdfs/Leucerne.pdf.

[77] Gary Gillespie, Liam Brennan, Jim Burke, and Dermot Forristal. "Spelt wheat food potential". UCD, Dublin and Teagasc, Oak Park, Carlow. URL: https://www.teagasc.ie/media/website/crops/crops/Spelt-Wheat-Food-potential.pdf.

[78] Vern Grubinger. "Winter rye: A reliable cover crop", 2021. Vegetable and Berry Specialist, University of Vermont Extension. URL: https://www.uvm.edu/vtvegandberry/factsheets/winterrye.html.

Appendix A

Complete tables of indicators and sources

	Indicator	References
Chemical	Nitrogen (N)	[42, 66, 71, 32, 1, 30, 8, 72, 64, 40, 44, 65, 69, 70, 39] [38, 11, 4, 73, 31, 29]
	pH	[4, 42, 14, 66, 71, 1, 29, 30, 74, 13, 8, 64, 40, 44, 65] [69, 70, 38, 11]
	Electrical conductivity	[4, 42, 14, 66, 71, 29, 30, 74, 13, 8, 64, 44, 65, 69, 70] [38, 11]
	Phosphorus (P)	[4, 42, 66, 71, 32, 1, 30, 13, 72, 64, 40, 65, 69, 70, 39] [66, 38, 11]
	Potassium (K)	[1, 11, 13, 30, 32, 38, 40, 42, 64, 70, 69, 66, 71, 72]
	CEC	[66, 71, 74, 8, 64, 40, 65, 69, 38]
	Iron (Fe)	[14, 71, 32, 13, 1, 40, 70, 39, 38]
	Heavy metals	[4, 71, 1, 13, 8, 40, 70, 39, 38]
	Manganese (Mn)	[14, 71, 13, 1, 40, 70, 39, 38]
	Calcium (Ca)	[71, 1, 72, 32, 40, 38]
	Magnesium (Mg)	[71, 13, 29, 40, 38]
	Aluminum (Al)	[14, 71, 32, 40]

TABLE A.1: Indicators and full references for the chemical pillar

	Indicator	References
Biological	Respiration	[4, 42, 73, 71, 31, 32, 29, 30, 14, 13, 8, 72, 40, 65] [39, 66, 67, 38]
	Organic Carbon	[1, 4, 8, 11, 14, 29, 32, 38, 39, 44, 65, 66, 71, 72, 67]
	Organic Matter	[42, 28, 75, 71, 30, 74, 73, 14, 13, 29, 72, 64, 40, 69, 70]
	Microbial biomass	[4, 74, 8, 64, 40, 39, 66, 38, 11]
	Earthworms	[4, 28, 71, 29, 30, 1, 64, 40, 65, 38]
	Active Carbon	[13, 65, 64, 71, 73]

TABLE A.2: Indicators and full references for the biological pillar

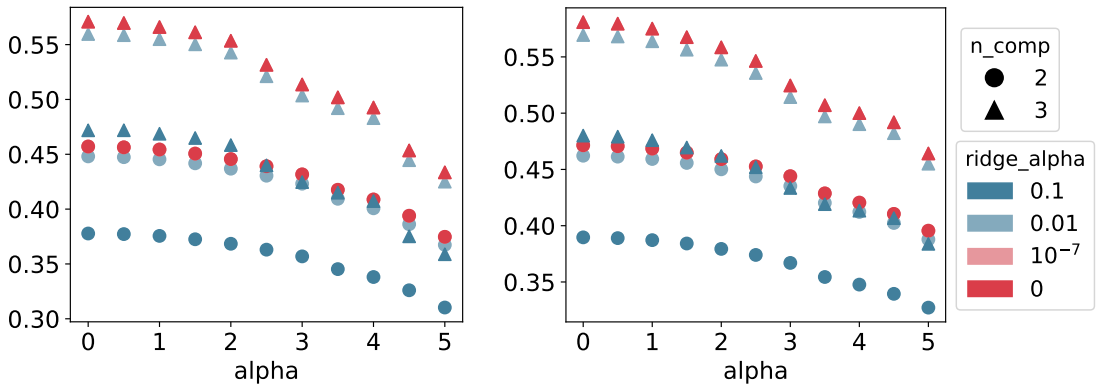
	Indicator	References
Physical	Bulk density	[4, 42, 28, 75, 1, 66, 71, 29, 30, 74, 8, 72, 64, 40, 44] [65, 69, 70, 39, 66, 38, 11, 14]
	Aggregate stability	[42, 28, 75, 71, 31, 29, 30, 74, 8, 72, 64, 44, 65, 39, 66] [67, 38, 11, 66, 73, 40]
	Water holding capacity	[4, 66, 71, 29, 74, 13, 8, 72, 64, 44, 70, 39, 66, 38, 11]
	Texture	[4, 42, 71, 29, 30, 66, 13, 64, 69, 70, 66]
	Water infiltration	[28, 75, 66, 71, 30, 74, 8, 29, 72, 66]
	Porosity	[64, 40, 44, 65, 69, 66, 38, 29]
	Mean weight diameter	[64, 40, 44, 65, 38, 11]
	Compaction	[74, 13, 8, 72, 38]
	Erosion	[4, 14, 1, 30]

TABLE A.3: Indicators and full references for the physical pillar

Appendix B

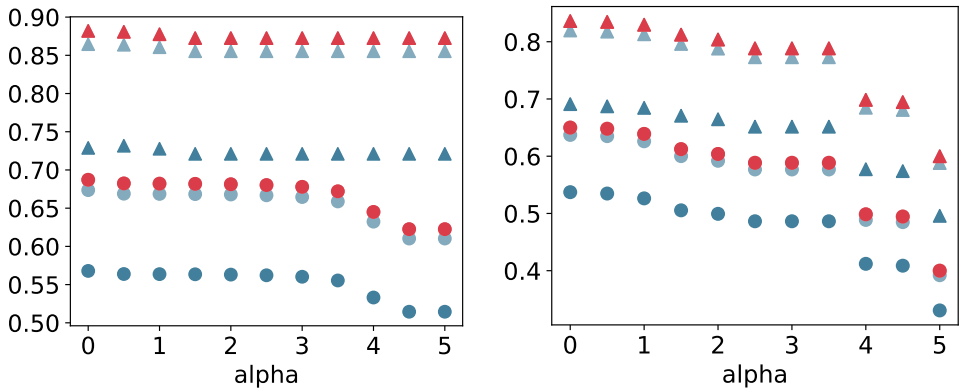
Additional plots for RQ1

Additional SPCA trade-off plots



(A) Marine clay (B) Sand

FIGURE B.1: SPCA trade-off plots for all pillars combined, separated by soil texture type



(A) Marine clay (B) Sand

FIGURE B.2: SPCA trade-off plots for the adapted chemical pillar

Correlation coefficients

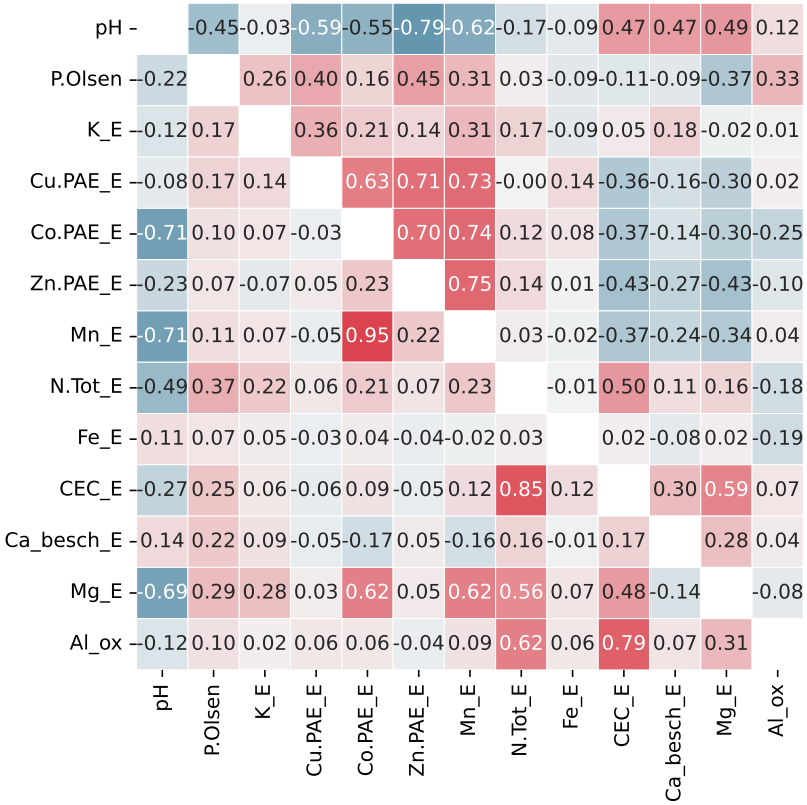
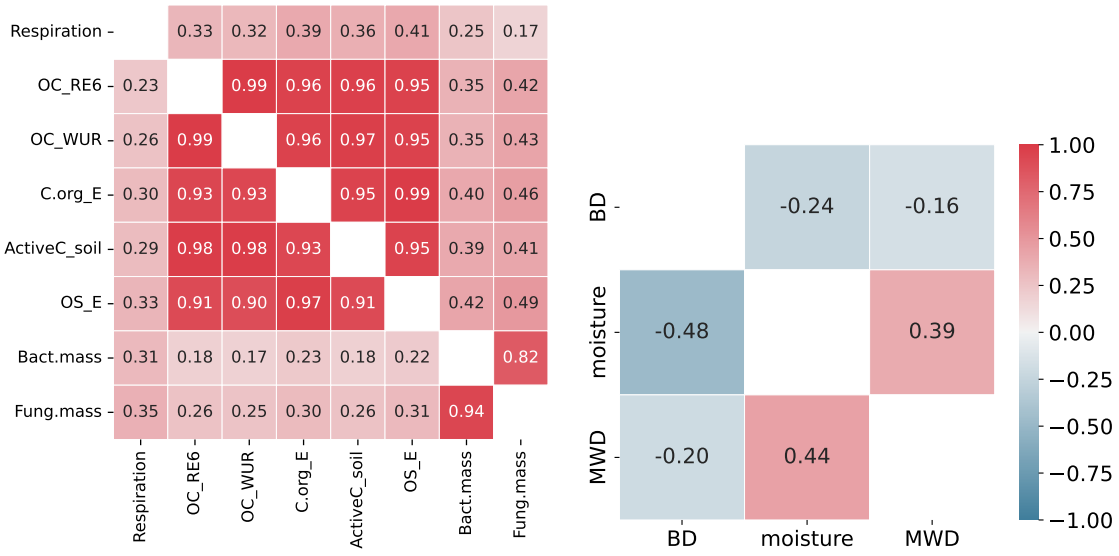


FIGURE B.3: Correlation coefficients for the chemical pillar. The lower triangular half is marine clay, upper triangular is sand.



(A) Biological (B) Physical

FIGURE B.4: Correlation coefficients for the biological and physical pillars. The lower triangular half is marine clay, upper triangular is sand.

Appendix C

Evaluation metrics for DRTs

Here are the full tables with the results for the evaluation metrics. The values are computed for each pillar. as well as for all pillars combined. as a comparison. The metrics are applied separately for each texture type (marine clay and sand) and the number of extracted principal components (n).

R^2 is the coefficient of determination (also the explained variance in this case). *Cont* is continuity (formula 4.3). *Trust* is trustworthiness (equation 4.2). r shows the Pearson correlation coefficient between the pairwise distances of the points from the original space and the distances in the lower dimensional space. This metric aims to evaluate how the original "shape" of the data changes in the new representation. A higher correlation coefficient would indicate the points retain proportional distances to the initial ones, whereas a weak r signifies that the points are now scrambled.

The dimensionality reduction techniques used are principal component analysis (PCA), space PCA (SPCA), incremental PCA (IPCA), and kernel PCA (KPCA). The latter is tested with five different kernels: polynomial of degree 2 (p2), polynomial of degree 3 (p3), RBF, sigmoid (sig), and cosine.

	MARINE CLAY								**SAND**							
	n = 2				n = 3				n = 2				n = 3			
	R^2	Cont	Trust	r	R^2	Cont	Trust	r	R^2	Cont	Trust	r	R^2	Cont	Trust	r
PCA	0.77	0.97	0.90	0.94	1	1	1	1	0.79	0.96	0.89	0.94	1	1	1	1
SPCA	0.77	0.96	0.90	0.93	1	1	1	1	0.78	0.97	0.91	0.95	1	1	1	1
IPCA	0.77	0.96	0.90	0.94	1	1	1	1	0.79	0.96	0.90	0.94	1	1	1	1
KPCA p2	0.67	0.92	0.84	0.82	0.90	0.97	0.95	0.87	0.76	0.94	0.87	0.86	0.95	0.98	0.97	0.90
KPCA p3	0.65	0.88	0.83	0.69	0.91	0.94	0.95	0.72	0.77	0.92	0.86	0.73	0.94	0.96	0.96	0.77
KPCA rbf	0.50	0.94	0.86	0.41	0.66	0.97	0.91	0.52	0.54	0.94	0.85	0.56	0.67	0.97	0.93	0.58
KPCA sig	0.63	0.96	0.91	0.87	0.82	0.99	0.99	0.96	0.65	0.96	0.88	0.90	0.80	0.99	0.99	0.97
KPCA cos	0.54	0.94	0.85	0.50	0.73	0.97	0.92	0.58	0.57	0.93	0.86	0.57	0.77	0.97	0.93	0.67

TABLE C.1: Biological pillar

	MARINE CLAY								SAND							
	n = 2				n = 3				n = 2				n = 3			
	R^2	Cont	Trust	r	R^2	Cont	Trust	r	R^2	Cont	Trust	r	R^2	Cont	Trust	r
PCA	0.85	0.98	0.94	0.96	1	1	1	1	0.80	0.97	0.92	0.92	1	1	1	1
SPCA	0.83	0.97	0.92	0.95	1	1	1	1	0.80	0.97	0.92	0.93	1	1	1	1
IPCA	0.85	0.98	0.94	0.96	1	1	1	1	0.80	0.97	0.92	0.92	1	1	1	1
KPCA p2	0.83	0.97	0.93	0.88	0.93	0.97	0.95	0.90	0.80	0.96	0.91	0.85	0.98	0.99	0.99	0.91
KPCA p3	0.82	0.95	0.94	0.77	0.92	0.94	0.93	0.80	0.77	0.92	0.86	0.69	0.95	0.94	0.94	0.72
KPCA rbf	0.64	0.94	0.86	0.63	0.80	0.98	0.95	0.68	0.60	0.93	0.84	0.55	0.66	0.96	0.92	0.62
KPCA sig	0.78	0.98	0.94	0.94	0.90	1	1	0.99	0.70	0.97	0.92	0.91	0.84	1	1	0.99
KPCA cos	0.67	0.94	0.87	0.71	0.83	0.97	0.94	0.74	0.64	0.93	0.84	0.63	0.83	0.97	0.91	0.69

TABLE C.2: Physical pillar

	MARINE CLAY								SAND							
	n = 2				n = 3				n = 2				n = 3			
	R^2	Cont	Trust	r	R^2	Cont	Trust	r	R^2	Cont	Trust	r	R^2	Cont	Trust	r
PCA	0.44	0.88	0.79	0.80	0.56	0.93	0.87	0.93	0.45	0.91	0.85	0.70	0.60	0.96	0.93	0.90
SPCA	0.42	0.89	0.79	0.76	0.52	0.89	0.79	0.88	0.42	0.91	0.83	0.66	0.53	0.96	0.91	0.74
IPCA	0.44	0.87	0.79	0.82	0.55	0.92	0.86	0.91	0.44	0.93	0.86	0.83	0.59	0.96	0.93	0.91
KPCA p2	0.36	0.82	0.76	0.79	0.53	0.85	0.76	0.85	0.44	0.87	0.83	0.77	0.57	0.92	0.87	0.83
KPCA p3	0.35	0.81	0.77	0.75	0.41	0.80	0.75	0.76	0.45	0.86	0.79	0.69	0.56	0.90	0.83	0.72
KPCA rbf	0.24	0.86	0.74	0.29	0.29	0.90	0.79	0.34	0.29	0.87	0.79	0.42	0.36	0.91	0.85	0.44
KPCA sig	0.21	0.89	0.81	0.65	0.25	0.93	0.86	0.68	0.19	0.91	0.84	0.68	0.22	0.95	0.91	0.78
KPCA cos	0.26	0.87	0.75	0.21	0.36	0.91	0.80	0.20	0.37	0.90	0.81	0.47	0.51	0.93	0.89	0.52

TABLE C.3: Chemical pillar

	MARINE CLAY								SAND							
	n = 2				n = 3				n = 2				n = 3			
	R^2	Cont	Trust	r	R^2	Cont	Trust	r	R^2	Cont	Trust	r	R^2	Cont	Trust	r
PCA	0.69	0.96	0.90	0.80	0.65	0.95	0.88	0.87	0.65	0.95	0.88	0.87	0.84	0.98	0.95	0.95
SPCA	0.67	0.95	0.89	0.76	0.59	0.94	0.88	0.80	0.59	0.94	0.88	0.80	0.79	0.98	0.93	0.94
IPCA	0.68	0.96	0.90	0.82	0.64	0.94	0.86	0.86	0.64	0.94	0.86	0.86	0.83	0.98	0.94	0.95
KPCA p2	0.65	0.91	0.84	0.79	0.69	0.94	0.87	0.86	0.69	0.94	0.87	0.86	0.84	0.97	0.95	0.91
KPCA p3	0.60	0.86	0.81	0.75	0.64	0.90	0.83	0.66	0.64	0.90	0.83	0.66	0.83	0.95	0.93	0.76
KPCA rbf	0.50	0.93	0.85	0.29	0.50	0.91	0.85	0.55	0.50	0.91	0.85	0.55	0.60	0.95	0.90	0.62
KPCA sig	0.58	0.95	0.90	0.65	0.47	0.94	0.88	0.84	0.47	0.94	0.88	0.84	0.57	0.98	0.94	0.93
KPCA cos	0.52	0.93	0.85	0.21	0.52	0.92	0.84	0.63	0.52	0.92	0.84	0.63	0.70	0.95	0.90	0.71

TABLE C.4: Adjusted chemical pillar

Appendix D

Feature coefficients for 2 and 3 principal components per pillar

Feature coefficients are obtained after applying regular PCA, sparse PCA (SPCA) with `alpha` and `ridge_alpha` tuned as explained in subsection 5.2 and incremental PCA (IPCA). Results are for 2 and 3 resulting principal components (n). PCA does not change coefficients when increasing the number of PCs, so it is not split like SPCA and IPCA.

		PCA			**SPCA**					**IPCA**				
					n = 2		n = 3			n = 2		n = 3		
		P1	P2	P3	S1	S2	S1	S2	S3	I1	I2	I1	I2	I3
M CLAY	Respiration	0.62	-0.19	0.76	0.71	0	0	0	1	0.63	-0.24	0.62	-0.19	0.76
	Bact.mass	0.59	-0.53	-0.61	0.71	0	1	0	0	0.58	-0.48	0.59	-0.53	-0.61
	OC_RE6	0.52	0.83	-0.22	0	0	0	1	0	0.51	0.84	0.52	0.83	-0.22
SAND	Respiration	0.55	0.75	-0.36	0	1	0	1	0	0.55	0.80	0.55	0.75	-0.36
	Bact.mass	0.57	-0.65	-0.50	0.70	0	-1	0	0	0.59	-0.58	0.57	-0.65	-0.50
	OC_RE6	0.61	-0.07	0.79	0.71	0	0	0	1	0.60	-0.17	0.61	-0.07	0.79

TABLE D.1: Biological pillar

		PCA			**SPCA**					**IPCA**				
					n = 2		n = 3			n = 2		n = 3		
		P1	P2	P3	S1	S2	S1	S2	S3	I1	I2	I1	I2	I3
M CLAY	BD	-0.55	0.67	-0.50	-0.71	0	-1	0	0	-0.56	0.68	-0.55	0.67	0.50
	moisture	0.65	-0.03	-0.76	0.71	0	0	0	1	0.64	-0.02	0.65	-0.03	0.76
	MWD	0.52	0.74	0.42	0	1	0	1	0	0.53	0.74	0.52	0.74	-0.42
SAND	BD	0.47	0.86	0.19	0	-1	0	-1	0	-0.49	0.84	-0.47	0.86	0.19
	moisture	-0.64	0.19	0.74	0.71	0	0	0	1	0.61	0.13	0.64	0.19	0.74
	MWD	-0.60	0.47	-0.64	0.70	0	1	0	0	0.63	0.53	0.60	0.47	-0.64

TABLE D.2: Physical pillar

		PCA			SPCA					IPCA				
					n = 2		n = 3			n = 2		n = 3		
		P1	P2	P3	S1	S2	S1	S2	S3	I1	I2	I1	I2	I3
M CLAY	pH	-0.51	-0.21	-0.10	-0.59	0	-0.59	0	0	0.49	-0.24	-0.49	0.23	-0.10
	P.Olsen	0.26	-0.48	0.13	0	0.78	0	0.78	0	-0.29	-0.37	0.29	0.44	0.12
	K_E	0.19	-0.37	-0.33	0	0	0	0	0	-0.21	-0.40	0.23	0.25	-0.42
	Cu.PAE_E	0.07	-0.21	0.05	0	0	0	0	0	-0.06	-0.21	0.09	0.07	-0.25
	Co.PAE_E	0.44	0.36	0.00	0.52	0	0.52	0	0	-0.43	0.50	0.43	-0.51	0.16
	Zn.PAE_E	0.14	0.15	0.65	0	0	0	0	1	-0.13	0.29	0.12	0.02	0.68
	N.Tot_E	0.40	-0.30	0.02	0.26	0.32	0.26	0.32	0	-0.42	-0.34	0.41	0.39	-0.04
	Fe_E	0.02	-0.11	-0.50	0	0	0	0	0	-0.02	-0.16	0.03	0.09	-0.26
	Ca_besch_E	-0.04	-0.55	0.38	0	0.55	0	0.55	0	-0.03	-0.35	0.01	0.52	0.38
	Mg_E	0.51	0.05	-0.20	0.56	0	0.56	0	0	-0.49	-0.02	0.49	-0.05	-0.20
SAND	pH	-0.47	-0.06	-0.13	-0.57	0	-0.57	0	0	0.46	0.15	0.47	-0.05	0.15
	P.Olsen	0.29	0.16	-0.45	0.18	0	0.15	0	0.31	-0.27	0.40	-0.28	0.15	0.40
	K_E	0.12	0.49	-0.24	0	0	0	0	0	-0.08	0.26	-0.11	0.49	0.28
	Cu.PAE_E	0.43	0.26	-0.05	0.52	0	0.52	0	0	-0.41	0.02	-0.42	0.26	0.03
	Co.PAE_E	0.39	0.23	0.23	0.47	0	0.48	0	0	-0.37	-0.21	-0.38	0.23	-0.19
	N.Tot_E	-0.05	0.56	0.21	0	0.85	0	0.86	0	0.08	-0.23	0.04	0.55	-0.20
	Fe_E	0.04	0.01	0.40	0	0	0	0	0	-0.03	-0.50	-0.04	0.06	-0.50
	CEC_E	-0.38	0.40	-0.01	-0.22	0.52	-0.23	0.51	0	0.40	-0.04	0.38	0.36	-0.03
	Ca_besch_E	-0.25	0.28	-0.19	0	0	0	0	0	0.27	0.22	0.24	0.36	0.23
	Mg_E	-0.38	0.23	0.12	-0.32	0	-0.31	0	0	0.40	-0.16	0.39	0.19	-0.15
	Al_ox	-0.03	-0.08	-0.64	0	0	0	0	0.95	0.03	0.58	0.03	-0.10	0.57

TABLE D.3: Chemical pillar

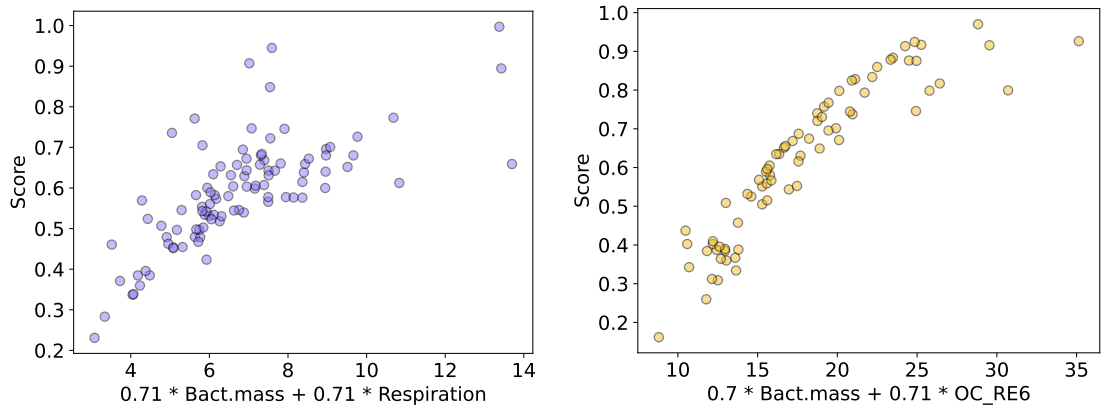
		PCA			SPCA					IPCA				
					n = 2		n = 3			n = 2		n = 3		
		P1	P2	P3	S1	S2	S1	S2	S3	I1	I2	I1	I2	I3
M CLAY	pH	-0.53	0.43	0.44	-0.63	0	-0.71	0	0	-0.51	0.12	-0.50	0.32	0.40
	P.Olsen	0.48	0.08	0.83	0.31	0	0	0	1	0.48	-0.43	0.48	-0.02	0.86
	K_E	0.34	0.88	-0.32	0	1	0	1	0	0.34	0.89	0.35	0.92	-0.18
	N.Tot_E	0.61	-0.18	-0.09	0.71	0	0.71	0	0	0.63	-0.06	0.63	-0.23	-0.25
SAND	pH	-0.62	-0.20	0.45	-0.71	0	-0.71	0	0	0.65	0.03	0.64	-0.14	0.48
	P.Olsen	0.49	0.37	0.15	0.71	0	0.71	0	0	-0.52	0.38	-0.52	0.28	0.27
	K_E	0.12	0.46	0.75	0	0	0	0	1	-0.16	0.67	-0.14	0.54	0.64
	N.Tot_E	-0.19	0.66	-0.46	0	0.71	0	0.71	0	0.12	0.46	0.15	0.66	-0.54
	CEC_E	-0.57	0.42	-0.04	0	0.71	0	0.71	0	0.52	0.44	0.53	0.41	0.01

TABLE D.4: Adjusted chemical pillar

Appendix E

Alternative approach to extracting summary scoring curves

Various combinations of indicator values can yield the same X-axis values while producing different associated scores, as shown in figure E.1, thus making the extraction of summary scoring curves more intricate. The **biological pillar** is used as a case study in attempting to extract a singular summary scoring curve. The summary values are computed using the linear combinations of the first principal components, based on the equations in table 5.12 (using raw indicator values), and the corresponding scores are derived by substituting the values with the scores in the same equations and dividing by the sum of the weights.



(A) PC1 of Marine Clay

(B) PC1 of Sand

FIGURE E.1: Scored summary indicators for the biological pillar

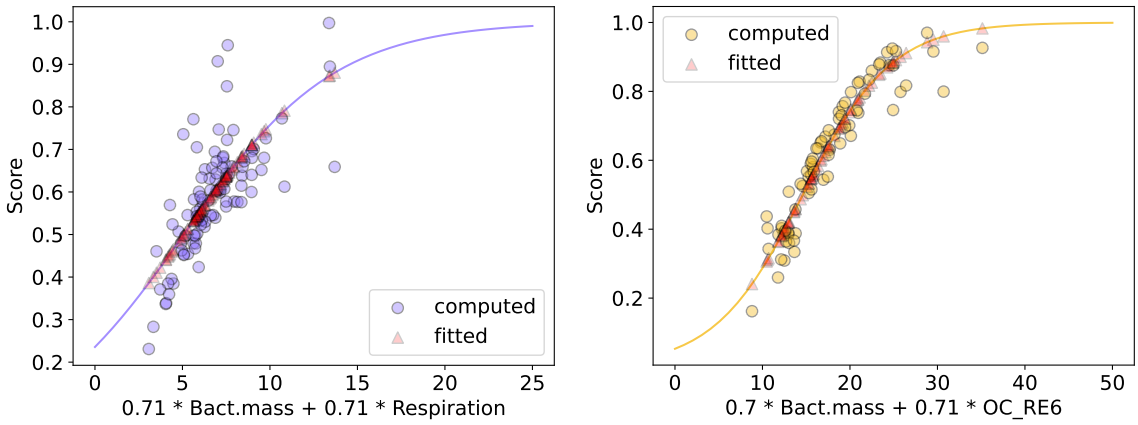
For both textures, but especially for the sand texture, a general trend is visible in the scattered points, with the score typically increasing as the summary indicator value rises. However, this trend is inconsistent across all combinations of individual indicator values, as seen in E.1a. For example, when the summary indicator equals 6, one sample is scored around 0.45, while similar values receive higher scores, ranging from 0.55 to 0.8. A similar pattern is observed for the sand texture in figure E.1b, where values approaching 15 are scored between 0.35 and 0.55.

The points appear to be primarily distributed according to a non-linear equation, which can be approximated. One method to derive this equation involves mathematically processing the formulas for each first principal component concerning their scores. This approach is challenging due to the complexity of the individual scoring functions, which are difficult

to manipulate in a manner which can produce a precise result. An alternative approach that more easily captures the trend involves applying a curve-fitting function from the `scipy` library in Python. For both texture types, the curve was approximated as a logistic function (same form as the ones for `OC_RE6` and `Bact.mass` in table 5.10). The program approximated both textures' b and c parameters, keeping a set to 1. The values for each parameter and the logistic function are shown in table E.1. The approximated summary scoring curves for each of the first principal components in the biological pillar are displayed in figures E.2a and E.2b, with the triangles being the new summary scores computed according to the fitted functions. The bullets for both texture types are computed identically as in figure E.1.

	a	b	c	Equation
M	1	3.244	0.231	$y = a/[1 + b \cdot \exp(-c \cdot x)]$
S	1	17.908	0.197	

TABLE E.1: Approximated parameters for fitting the summary scoring curves for the biological pillar in both texture types.



(A) PC1 of Marine Clay

(B) PC1 of Sand

FIGURE E.2: Fitted summary scoring curves for the first principal component of both soil texture types for the biological pillar

Fitting a curve to the points provides a singular equation for scoring summary indicators, which is the preferred outcome. However, this approach requires additional input and analysis within the framework, as the approximation functions need an initial equation as a starting point. This means users must review the point distribution and scoring formulas for each indicator contributing to the linear combination of the summary indicator. In cases where the summary indicator is composed of just two measurements (`Bact.mass` and `Respiration` for marine clay texture, or `Bact.mass` and `OC_RE6` for sand texture, as specific examples), or when all indicators follow the same base function, this task is achievable with a certain level of field knowledge. When the number of indicators per principal component increases or the equations vary significantly, it becomes difficult for a typical user to determine an appropriate equation to fit the points, deeming this method unfeasible.

Appendix F

Additional SMAF parameters and classes for scoring

Site-specific parameters translated to classes

texture	texture class
clay (> 60% clay)	1
sandy clay, clay loam, silty clay loam, silty clay, or clay (< 60% clay)	2
silt loam, silt	3
sandy loam (> 8% clay), sandy clay loam, or loam	4
sand, loamy sand, sandy loam (< 8% clay)	5

TABLE F.1: Soil texture classes

season	class
spring	1
summer	2
fall	3
winter	4

TABLE F.2: Season classes

degree days / avg. ppt	class
hi/hi: ≥ 170°days & ≥ 550 mm	1
hi/lo: ≥ 170°days & < 550 mm	2
lo/hi: < 170°days & ≥ 550 mm	3
lo/lo: < 170°days & < 550 mm	4

TABLE F.3: Climate classes. ppt = precipitation.

slope (%)	class
0 - 2	1
2 - 5	2
5 - 9	3
9 - 15	4
15+	5

TABLE F.4: Slope classes

The organic matter (OM) class is determined using soil suborders, ordered by OM content. This information is not available in the current dataset, therefore the assumption is that the samples have a high OM content (class 10). The weathering parameter is also

based on soil types, having a total of 3 classes. The first two classes consider only a couple of suborders, so class 3 (*all others* - only slightly weathering) is set. Mineralogy also has 3 classes, the first two being unfit for The Netherlands as they are of volcanic nature (smectitic and glassy); *other* — class 3 — is deemed appropriate for this context.

parameter	**set class**
season	1
climate	1
organic matter	1
mineralogy	3
slope	1
weathering	3

TABLE F.5: Classes for the parameters which are applied for all samples

Curve equations and parameters

Although most of the classes for the site-specific features remain constant in this analysis, the tables with the values of the equation parameters are kept mostly complete. The changes in values can offer valuable insight into how and the degree to which each parameter affects the score.

Organic Carbon

$$y = a/(1 + b \cdot e^{-c \cdot x}) \tag{F.1}$$

$$c = (c_1 \cdot c_2) + (c_1 \cdot c_2 \cdot c_3) \tag{F.2}$$

a	**b**	**OM**	$\mathbf{c_1}$	**texture**	$\mathbf{c_2}$	**climate**	$\mathbf{c_3}$
1	50.1	1	1.30	1	1.60	1	0.15
		2	1.55	2	1.25	2	0.05
		3	2.17	3	1.10	3	-0.05
		4	3.81	4	1.05	4	-0.10
				5	1		

TABLE F.6: Lookup table for fixed parameters and the site-specific factors for scoring organic carbon

Microbial biomass

$$y = a/(1 + b \cdot e^{-c \cdot x}) \tag{F.3}$$

$$c = (c_1 \cdot c_2 \cdot c_3) \tag{F.4}$$

a	b	OM	c_1	texture	c_2	season, climate	c_3
1	40.748	1	0.0062097	1	1.100	1	1
		2	0.0124192	2	1.025	2,1	0.920
		3	0.0186290	3	1	2,2	0.930
		4	0.0212898	4	0.980	2,3	0.940
				5	0.950	2,4	0.950
						3,1	0.980
						3,2	0.975
						3,3	0.970
						3,4	0.965
						4,1	0.880
						4,2	0.900
						4,3	0.920
						4,4	0.940

TABLE F.7: Lookup table for fixed parameters and the site-specific factors for scoring microbial biomass. A **season, climate** of 1 indicates spring.

pH

$$y = a \cdot e^{-(x-b^2)/2c^2} \tag{F.5}$$

crop	b pH opt	c pH range / 2	Source
barley	6.25	0.75	SMAF
winter wheat	6.25	0.75	SMAF
clover	6.5	1	SMAF
oats	6	2	SMAF
triticale	6.5	1	SMAF
lucerne	7	0.5	[76]
spelt	6	0.75	[77]
winter rye	6	1	[78]

TABLE F.8: Parameters for scoring the pH indicators based on crop type. **b** is the optimal pH value and **c** is half of the accepted range for fluctuating the pH.

Phosphorus

The scoring curve for phosphorus is defined by two separate functions, and choosing which to apply depends on the maximum level of P (P_{max}) and *EnvProtect*, as seen in equation F.6. The latter is relative to the slope class and is a limit meant to consider the environment, as phosphorus runoff can cause damage to water bodies. P_{max} is the maximum optimal for crop use and is computed as $P_{max} = P_{opt} + 6$. P_{opt} (optimal value for P) is specific to each crop type. Since not all crops in the dataset (as seen in the pH section

above) can be matched to a value, P_{opt} is set to 21 mg/kg (values range from 19 to 21 in the SMAF file, thus the difference in results would not be considerable).

$$y = \begin{cases} f_1 & \text{if } x \cdot m \leq P_{max} \\ f_2 & \text{if } x \cdot m > EnvProtect \\ 1 & \text{otherwise} \end{cases} \tag{F.6}$$

Equation F.7a describes f_1, and b is chosen based on table F.11. f_2 corresponds to equation F.8a and c is a function of slope, organic matter and texture (same table F.11).

$$f_2 = \frac{a \cdot b + c \cdot (x \cdot m)^d}{b + (x \cdot m)^d} \tag{F.7a}$$

$$b = b_1 + b_1 \cdot b_2 \cdot b_3 \tag{F.7b}$$

$$b_1 = 213.96744 + 39.579185 \cdot P_{opt} + 2.3020512 \cdot P_{opt}^2 \tag{F.7c}$$

$$f_2 = a - b \cdot e^{-c \cdot (x \cdot m)^d} \tag{F.8a}$$

$$c = (c_1 + c_1 \cdot c_2) \cdot c_3 \tag{F.8b}$$

The phosphorus scoring scheme depends on multiple site-specific factors, as well as on the method used for measuring, m. Besides the actual method, m is also relative to the weathering class of the soil, which can be seen in table F.10. As it is not specified in the dataset, all samples are classified as *slightly* (class 3). The method used for obtaining the dataset phosphorus is *Olsen*, equivalent to class 4 in the methods column. The remaining 5 methods are not mentioned in this appendix, as they are irrelevant to this case study.

	a	**b**	**c**	**d**
f_1	$9.25 \cdot 10^6$	Tab F.11	1	3.06
f_2	1	4.5	Tab F.11	-2

TABLE F.9: Lookup table for the constant factors in both phosphorus functions

method	**weathering**	**m**
4	1	2.4
4	2	1.8
4	3	1.7

TABLE F.10: Lookup table for the method factor parameter (m)

slope	**EnvProtect**	$\mathbf{c_1}$	**OM**	$\mathbf{b_2}$	$\mathbf{c_2}$	**texture**	$\mathbf{b_3}$	$\mathbf{c_3}$
1	160	110000	1	0.1250	0.250	1	0.98	0.9
2	140	90000	2	0.0250	0.050	2	0.99	1
3	115	70000	3	0.0175	0.035	3	1.00	1.1
4	85	35000	4	0.0100	0.020	4	1.01	1.4
5	60	20000				5	1.03	1.6

TABLE F.11: Lookup table for the site-specific factors for scoring phosphorus

Potassium

$$y = a \cdot (1 - e^{-b \cdot x}) \qquad \text{(F.9)}$$

texture class	a	b
1, 2	1.054133	-0.00981
3, 4, 5	1.074490	-0.01340

TABLE F.12: Lookup table for the site-specific factors for scoring potassium

Bulk density

This indicator is dependent on texture type. For a texture class ≥ 4, the method for computing the b, c, and d parameters changes and includes soil mineralogy. This method is not described here, since this research deals exclusively with classes 1 and 2.

$$y = a - b \cdot e^{-c \cdot x^d} \qquad \text{(F.10)}$$

a	texture	b	c	d
0.994	1	0.792	321.34	-12.990
	2	0.794	88.025	-12.061
	3	0.796	32.189	-11.297

TABLE F.13: Lookup table for scoring BD

Available water capacity

$$y = \begin{cases} (a \cdot b + c \cdot x^d)/(b + x^d) & \text{if } region = arid \\ a + b \cdot \cos(c \cdot x + d) & \text{if } region = humid \end{cases} \qquad \text{(F.11)}$$

As the samples are only collected from The Netherlands, the region is *humid*, thus only the second equation is applied. Table F.14 shows how the parameters are computed for this specific function.

a	b	c	texture	d
0.4772	0.52675	6.87765	1	-1.8928800
			2	-2.3483425
			3	-2.4717800
			4	-2.2356867
			5	-2.0427200

TABLE F.14

Appendix G

Additional crop plots

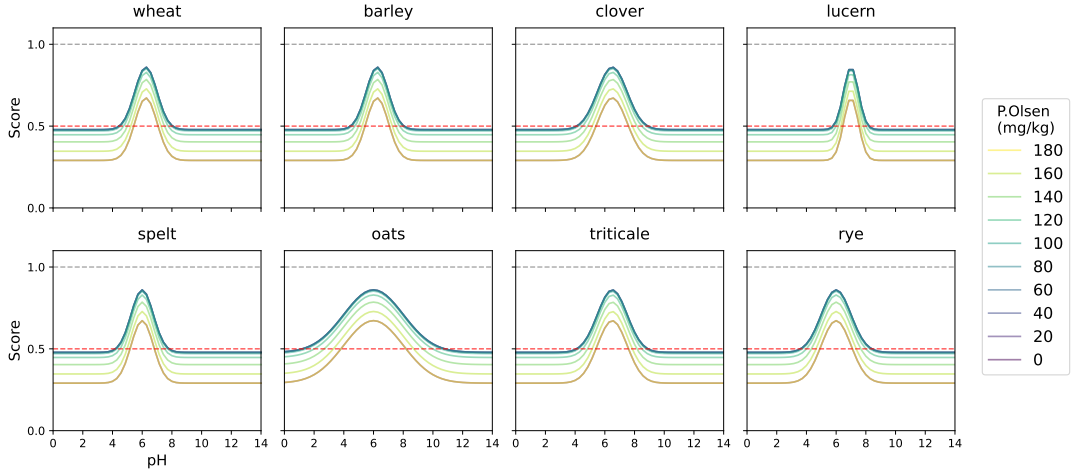


FIGURE G.1: Summary scoring curves for the first principal component of the chemical pillar, marine clay texture. Curves are plotted according to the same values, only changing the crop type.

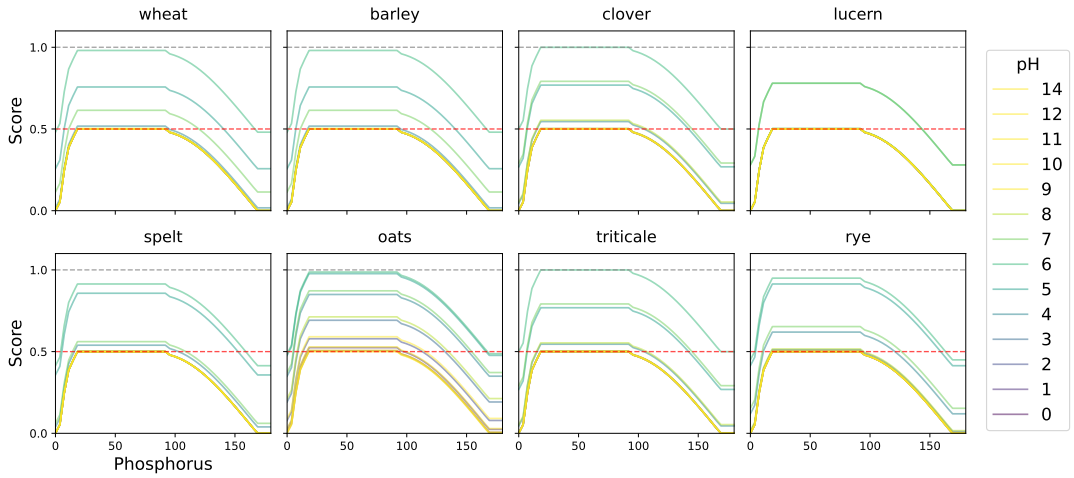
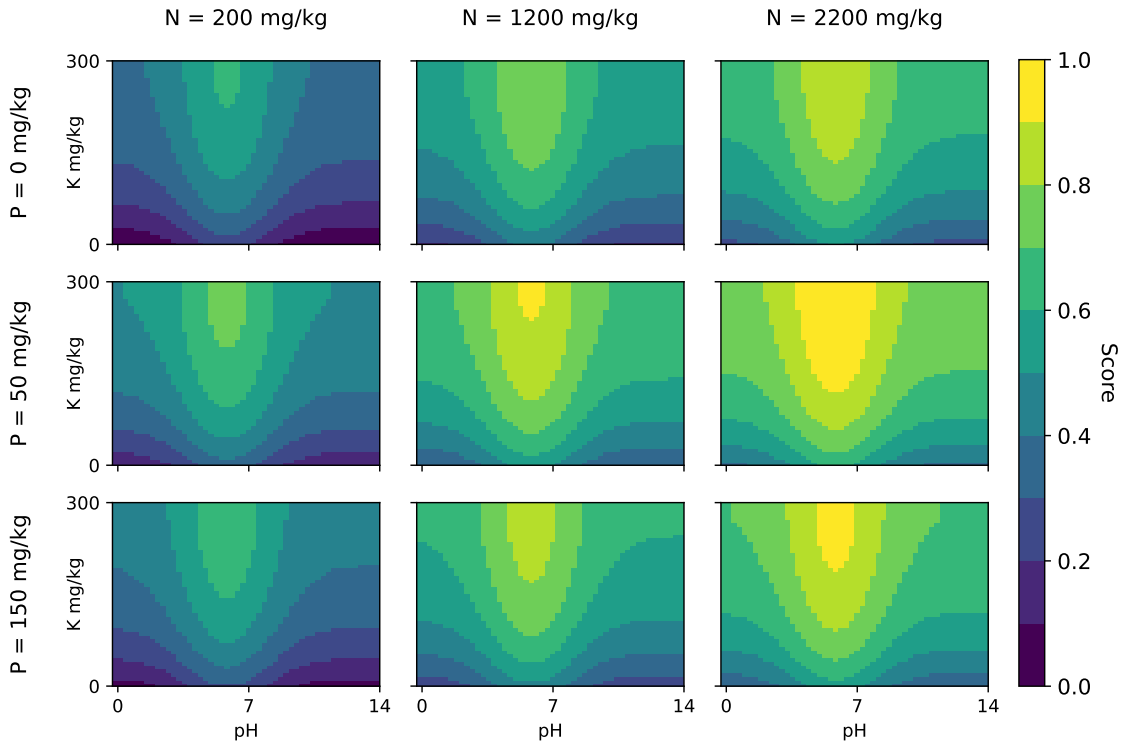
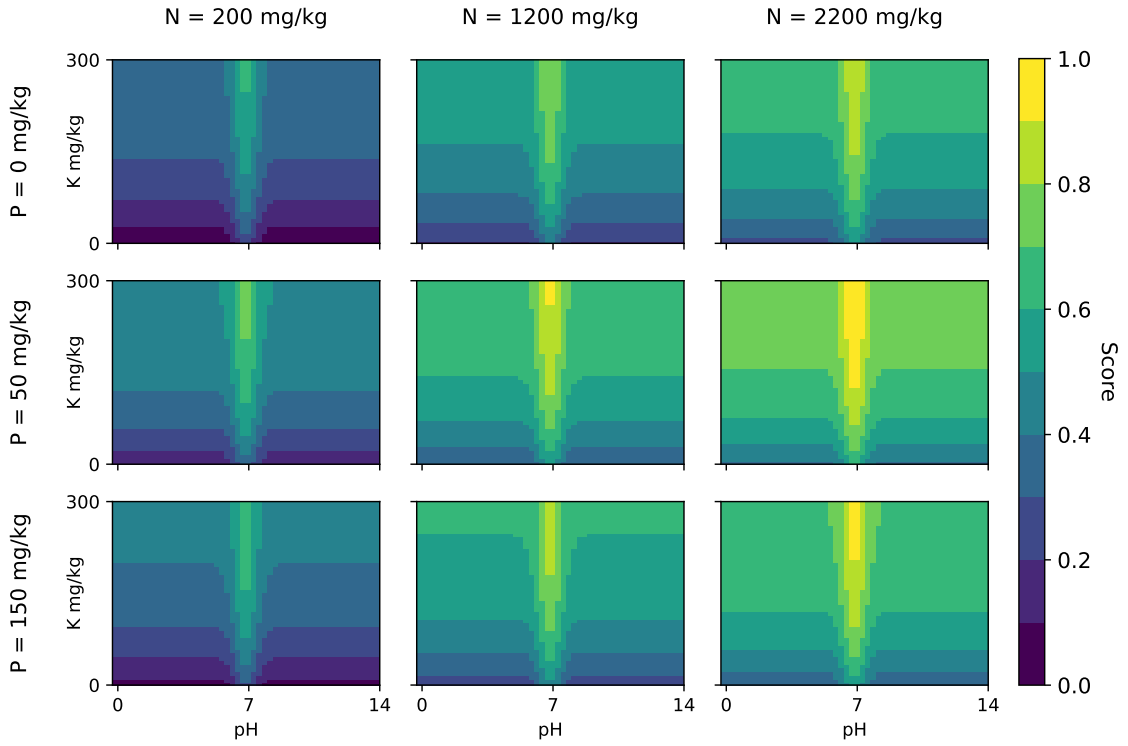


FIGURE G.2: Plotting Phosphorus as x-axis instead of pH. Same conditions used as in fig. G.1.



(A) **Oats**



(B) **Lucerne**

FIGURE G.3: Scoring heatmaps for the chemical pillar, marine clay texture. N = Nitrogen (`N.Tot_Eurofins`), P = Phosphorus (`P.Olsen`), K = Potassium (`K_Eurofins`). Different heatmaps for 2 types of crops.