

Selection of cropland sites for long-term monitoring soil microbiomes as indicators for soil health in context of the German Agricultural Soil Inventory (BZE-LW)

Sainur M. Samad, Christopher Poeplau, Axel Don, Christoph C. Tebbe

Dr. Sainur M. Samad
ORCID: 0000-0003-1409-7985
Thünen Institute of Biodiversity

PD Dr. Christopher Poeplau
ORCID: 0000-0003-3108-8810
Thünen Institute of Climate-Smart Agriculture

Prof. Dr. Axel Don
ORCID: 0000-0001-7046-3332
Thünen Institute of Climate-Smart Agriculture

Prof. Dr. Christoph C. Tebbe
ORCID: 0000-0003-4861-0214
Thünen Institute of Biodiversity

Braunschweig/Germany, March 23, 2024

Table of contents / Inhaltsverzeichnis

Summary	2
Zusammenfassung	2
1 Introduction	2
2 Methods	3
3 Results	4
4 Discussion and conclusion	4
5 References	14
Appendix A	15
Step-1: Import data and merging	15
Step-2: Filtering parameters	16
Step-3: Subset data for PCA analysis	16
Step-4: Data standardization:	17
Step-5: PCA analysis	17
Step-6: Data extraction from PC1 and PC2	23

Summary

In this report we present a selection of 192 cropland sites across Germany for prospective monitoring of soil microbiomes in order to assess their status and potential long-term changes in German agricultural soils. We explain the process by which the soils were selected from the database information generated by the first German Agricultural Soil Inventory (BZE-LW) during 2009 to 2018 which included a total of 2,234 cropland sites. It was the objective to reduce the impact of physicochemical soil properties which are well known to affect the microbial community structure, i.e. pH, total organic carbon (TOC) and CN ratio, in favour of finding indicators for implications of agricultural soil management practices or climate change. More specifically, sites with pH < 5.5 and > 7.5, TOC with < 10 g/kg and > 20 g/kg, CN ratio > 13 and a sand content > 90% were upfront excluded. Multidimensional scaling of the selected properties, including physicochemical soil parameters, textures, crop cover and other factors are presented in this report to demonstrate the range of soil properties which will be analysed for their indicator potential for soil health during the further course of the BZE. – This work is a contribution of the MonViA project (National Monitoring of Biodiversity in Agricultural Landscapes).

Keywords: monitoring, microbiomes, agricultural soils, soil management, climate change

Zusammenfassung

In diesem Bericht stellen wir eine Auswahl von 192 Ackerstandorten in ganz Deutschland für ein prospektives Monitoring von Bodenmikrobiomen vor, um deren Status und mögliche langfristige Veränderungen in deutschen landwirtschaftlichen Böden zu bewerten. Wir erläutern den Prozess, mit dem die Böden aus den Datenbankinformationen ausgewählt wurden, die im Rahmen der ersten deutschen landwirtschaftlichen Bodeninventur (BZE-LW) in den Jahren 2009 bis 2018 generiert wurden und insgesamt 2.234 Ackerstandorte umfassten. Ziel war es, den Einfluss physikochemischer Bodeneigenschaften, von denen bekannt ist, dass sie sich auf die Struktur der mikrobiellen Gemeinschaft auswirken, d. h. pH-Wert, gesamter organischer Kohlenstoff (TOC) und CN-Verhältnis, zugunsten der Suche nach Indikatoren für die Auswirkungen landwirtschaftlicher Bodenbewirtschaftungspraktiken oder des Klimawandels zu reduzieren. Konkret wurden Standorte mit einem pH-Wert < 5,5 und > 7,5, einem TOC-Wert < 10 g/kg und > 20 g/kg, einem CN-Verhältnis > 13 und einem Sandgehalt > 90 % ausgeschlossen. Eine mehrdimensionale Skalierung der ausgewählten Eigenschaften, einschließlich physikalisch-chemischer Bodenparameter, Texturen, Bodenbedeckung und anderer Faktoren, wird in diesem Bericht vorgestellt, um das Spektrum der Bodeneigenschaften zu veranschaulichen, die im weiteren Verlauf der BZE auf ihr Indikatorpotenzial für die Bodengesundheit untersucht werden sollen. - Diese Arbeit ist ein Beitrag des MonViA-Projekts (Nationales Monitoring der Biodiversität in Agrarlandschaften).

Schlüsselwörter: Monitoring, Mikrobiome, Landwirtschaftliche Böden, Bodenbearbeitung, Klimawandel

1 Introduction

Natural microbiomes provide essential functions to allow a sustainable agricultural use of soils. They carry out crucial steps in the biogeochemical cycling of carbon and nitrogen, they contribute to the build-up of soil organic matter, they induce the formation of soil aggregates and thereby improve soil structure, and they provide nutrients to promote plant growth and they degrade pesticides. Therefore, they are indispensable for soil health. Agricultural management practices as well as climate change factors have been shown to alter the abundance, composition and activity of soil microbiomes, typically with yet poorly predictable impacts on soil and ecosystem health. A better understanding how soil microbiomes change over time, should support the development of environmentally friendly agricultural management practices, e.g. by advising on crop rotation or soil tillage options.

Therefore, as initiated by research conducted with the MonViA project, an effort is now made to include assessments of the soil microbiome in context of the second German Soil Inventory (BZE- -LW II; <https://www.thuenen.de/de/fachinstitute/agrarklimaschutz/projekte/bodenzustandserhebung-landwirtschaft-bze-lw>). It was decided that an important approach to generate such data is based on soil DNA analyses, thus omitting the need to cultivate microorganisms in laboratory media but instead directly access and assess the soil-intrinsic genetic information. Soil microbial diversity is assessed from directly extracted soil DNA utilizing PCR to amplify structural and functional microbial marker genes and assess their diversity by DNA sequencing, followed by a range of bioinformatic and statistical analyses.

Basic information was gathered and published how to best handle field samples and store soils in the laboratory before DNA is extracted and they are characterized for the abundance and diversity of microbiomes (Finn et al., 2023). Currently it is also investigated, whether soil microbiomes can be recovered from air-dried soil samples ("archived soils") of the first BZE campaign (BZE -LW I) in a way, that their composition may give information on the soil status as it was at the time of sampling.

The objective of monitoring the soil microbiome is to evaluate the indicator potential of soil microbial diversity for assessing the status of agricultural soils and their vulnerability to environmental changes. Previous studies conducted in other countries, including the United Kingdom and France, had already demonstrated that pH and soil organic carbon were main factors affecting the microbial community structure (Griffiths et al., 2011; Karimi et al., 2018). However, for assessing the interplay between soil microbial diversity and the status of an agricultural soil, it deemed more important to evaluate the response of soil microbial communities to agricultural land use and management (Plassart et al., 2019; Thomson et al., 2015). Utilizing the database provided by BZE-LW, stored at the Thünen Institute, based on data collected from a grid of 8 km x 8 km, during the years 2009 to 2018, we selected in this study a total of 2,234 cropland sites which were then analysed in terms of their land use and soil properties. In order to reduce the influence of pH and total soil organic carbon, sites with a pH < 5.5 and > 7.5, TOC with < 10 g/kg and > 20 g/kg, CN ratio > 13 and a sand content > 90% and also peat sites were excluded. As the main outcome of this work presented here, we identified 198 sites which we now propose to be included for the second National Soil Inventory Agriculture (BZE II-LW).

2 Methods

The key dataset of the BZE-LW I is stored at Poeplau et al. 2020 (for direct link: https://www.openagrar.de/receive/openagrar_mods_00054877). The data are distributed in 10 files either "CSV" or "txt" format. These files show a wide range of variables related to soil classification, physico-chemical parameters of soils, vegetation, land management and history, and weather-related data of each site (as indicate as PointID). The data were processed and analysed in R (version 3.6.1) using "plyr (1.8.4)", "dplyr (1.0.3)", "readr (1.3.1)", "tidyverse (1.3.0)", "factoMineR (2.4)", "factoextra (1.0.7)" and "ggplot (3.2.1)" packages and followed the below steps for analyses.

Firstly, all data were imported and merged in R.

Secondly, some additional variables were calculated, for example, the CN ratio ($C/N = TC/TN$), mean annual temperature, mean annual precipitation, mean soil moisture, mean drought index.

Thirdly, the data were filtered by excluding samples based on pH (excluded sites if pH >7.5 and pH < 5.5), TOC (excluded sites if TOC with < 10 and > 20), CN-ratio (excluded sites if C:N >13), sand content (excluded sites if sand >90%), peat sites (excluded if "yes"), land.use.history.heathland.(excluded if "yes"). The above filtering process was supposed to reduce the data set to approximately 800 to 900 cropland sites. As there are more than 100 variables in the BZE dataset that describe the abiotic properties of a site or sample, we selected the following (what we considered most important) 16 to be included into the subsequent multivariate analysis: the proportion of (1) silt, (2) clay and (3) sand, (4) medium sized silt, (5) fine sized silt, (6) medium sized sand, (7) fine sized sand, (8) total soil carbon content (TC), (9) total soil organic carbon (TOC), (10) total nitrogen content (TN), (11) CN

ratio, (12) soil moisture*, (13) drought index*, (14) precipitation**, (15) pH value measured in CaCl₂ solution, and (16) air temperature* (*all annual mean values; **annual sum).

Fourthly, the variables related to those sites were standardized and then visualized with principle component analyses (PCA) in R.

Finally, we selected the sites that are close to the center (within a radius of < 2 unit in x, y coordinate distances) of principle component 1 (PC1) and 2 (PC2) in PCA plot. This filtering approach was targeted to retain approx. 200 cropland sites. A detailed analysis and R script are attached below (see Appendix A.).

3 Results

The BZE-LW I dataset showed a total of 2,234 cropland sites. After prefiltration, based on pH, TOC and CN ratio, sand content, peat land, and land use history, we narrowed this selection down to a total of 835 sites. The prefiltration step excluded all sites that showed low/high pH, low/high TOC, high sandy and peatland sites. The PCA graph explains a total of 57.3 % variation in two-dimensional scale (Fig. 1 A.), with PC1 (38.5 %) and PC2 (18.8 %), respectively. It should be noted that PC3 to PC10 also explained a considerable amount of variation (total 41.2 %), which however, was not considered further in the selection process applied here (Fig. 1 B.). The scattered plot of PCA (Fig. 1 C.) shows the actual position of all 835 sites in a two-dimensional scale. The sites which are close to the centre of PCA-plot represent a highly representative mean of the selected soil properties and thereby a considerable similarity to each other. On the other hand, the sites which are far away from the centre of PCA-plot represent a strong association with respective to soil and environmental properties. In order to exclude sites which are further away shifted towards different PC 1 and PC 2, a circular zone was drawn where the centre of the zone was zero and radius was 2 within the PCA-plot (Fig. 1 C.). Through this selection process, a total of 192 sites were then finally selected (Fig. 1 D.). Within these sites, the pH was between 5.5 and 7.4 (Table 1), while the actual vegetation at the time of sampling varied, e.g., some with clover, silage maize, oat, barley, rye, or wheat (Table 2). The selection included different soil types, i.e. luvisols, gleysols, stagnosols, cambisols, phaeozems, regosols, fluvisols, anthrosols and vertisols.

4 Discussion and conclusion

Soil microbiomes are strongly selected by their habitat conditions, which define water regimes, nutrient availabilities, colonizable soil particle surface areas, soil pH, or the availability of electron acceptors. Our intention was to exclude the impact that such factors, which are mainly defined by pedogenesis and long-term history of land use, of the abundance and diversity of the soil microbiome. Instead, we selected here similar field sites and then assess in a prospective long-term monitoring, how the microbiome would change in the light of different agricultural management practices (crop rotations, soil tillage, application of fertilizers and pesticides) or altered climatic conditions. Ultimately, this approach reduced the number of cropland sites by more than 90 % to 192 sites, which are considered sufficient to gather information about the status of microbiome on representative common cropland field sites in Germany. However, it should be noted that more extreme soils, e.g. peat land or sandy soils, or soils with a more extreme pH could be more vulnerable to environmental changes, and thus the selected soils here are not representative for the wide range of soils that occur across Germany. A limitation of the selected 192 sites also applies to the fact linked to temporal variation and the importance of cropping systems implemented at the time of sampling. For the temporal variation, we currently analyse a data set generated from the MonViA project, where at one farm in Northern Germany (near Hildesheim, Lower Saxony) the microbiome of three neighbouring fields differing in soil texture and soil tillage were characterized over a period of two years at two week-intervals (Wang et al., manuscript in preparation). The importance of cropping systems for shaping the soil microbiome during seasonal changes has recently been demonstrated in other studies (Liu et al., 2022; Liu et al. 2023). Thus, differences between soil microbiomes as emerging from comparing samples from different years taken at the same field site may not immediately demonstrate a tangible beneficial or adverse effect. To identify such, however, promising results come from analysing larger data sets by bioinformatics and complex

statistical approaches, i.e., network analyses, the latter allowing to distinguish between functionally important and less important members of the soil microbiome. For the evaluation of such approaches, the 192 sites selected here should provide valuable new results in the future.

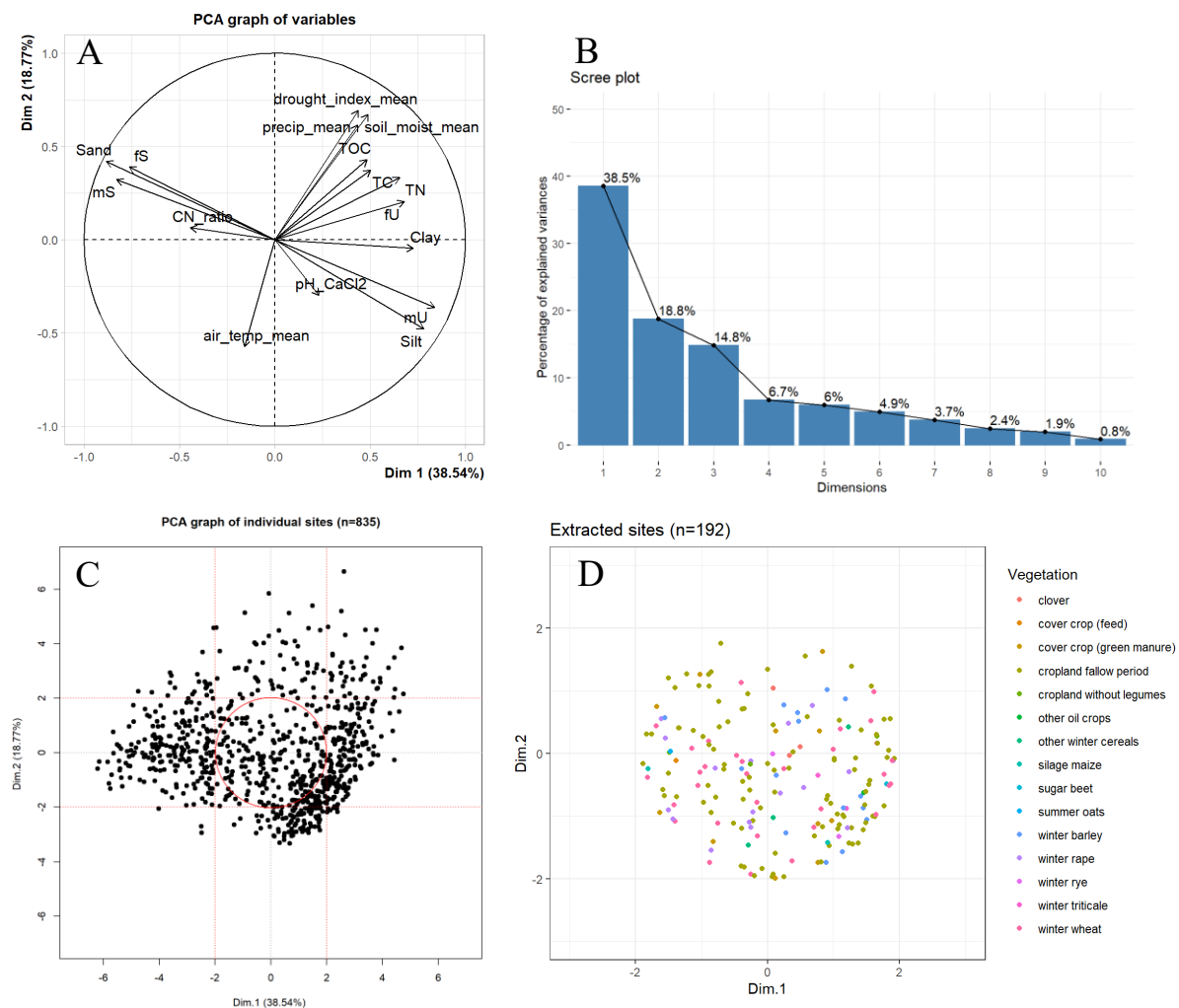


Figure 1. Steps and outcomes in the selection of soils for long-term monitoring soil microbiomes in context of future soil inventories for agriculture sites. **A.** Characterization of the impact of the 16 selected soil variables to distinguish soils based on Principal Component Analysis using the first two Principle Components. **B.** Contribution of the first ten Principle Components on the variation of the dataset. **C.** Variation of the 835 cropland sites selected for this study and indication of the area which was selected to further select the prospective monitoring sites. **D.** Distribution of the 192 sites and their vegetation at the time of sampling. Abbreviations: TOC: total soil organic C; TC: Total soil C (incl. carbonates C); TN: Total bound nitrogen; fU: Fine silt content; fS: Fine sand content; mU: Medium sized silt content; mS: Medium sized sand content.

Table 1. Range of soil properties of the 192 selected soils

	Silt (%)	Sand (%)	Clay (%)	pH (CaCl₂)	TOC (g/kg)	TC (g/kg)	TN (g/kg)	C:N
Min.	19.80	1.90	8.50	5.53	10.22	10.28	0.99	8.04
Max.	85.60	59.70	40.60	7.44	19.82	22.87	2.25	12.91
Average	52.85	28.48	18.67	6.42	14.04	14.50	1.40	10.34

Table 2. Properties of the 192 selected soils for a prospective monitoring of soil microbiomes on context of the 2nd agricultural soil inventory (BZE II-LW). Sites are shown here as “PointID” in accordance to the BZE database. The column “Eco” shows the cropland under organic (1) or convention farming (0). The column “WRB” shows the soil classification according to World Reference Base. Samples indicated yellow background (n = 20) were selected for a study on the recovery of soil microbiomes from archived soil samples (Samad *et al.*, manuscript in preparation). Main soil type (Hauptbodentyp) acronym refers to the German Classification System (Bodenkundliche Kartieranleitung, 2005)

Point ID	Eco	WRB	Farm type	Main soil type	Vegetation
1141	0	Cambisols	mixed	AB	winter rape
116	0	Luvisols	mixed	LL	cropland fallow period
1183	1	Cambisols	mixed	AB	other winter cereals
124	0	Gleysols	arable	MN	winter rape
1307	0	Stagnosols	arable	SS	winter wheat
136	0	Stagnosols	arable	SS	cover crop (green manure)
1382	0	Stagnosols	arable	SS	cropland fallow period
151	1	Anthrosols	mixed	ZZ	clover
1607	0	Cambisols	mixed	BB	cropland fallow period
1609	0	Luvisols	mixed	LL	other oil crops
1615	0	Cambisols	mixed	BB	cropland fallow period
1620	0	Regosols	arable	RQ	cropland fallow period

1635	0	Cambisols	arable	AB	winter rape
167	0	Stagnosols	arable	SS	winter rape
1672	0	Anthrosols	mixed	YE	cropland fallow period
1673	0	Gleysols	mixed	GG	cropland fallow period
1677	0	Stagnosols	mixed	SS	cover crop (green manure)
1678	0	Cambisols	mixed	BB	cover crop (green manure)
1741	0	Luvisols	mixed	LL	cropland fallow period
1744	0	Phaeozems	arable	YK	cropland fallow period
1812	0	Stagnosols	mixed	SH	cropland fallow period
1822	0	Luvisols	mixed	LL	cover crop (green manure)
1825	0	Regosols	arable	RQ	cropland fallow period
1834	0	Regosols	arable	RZ	winter barley
1881	0	Stagnosols	arable	SS	winter rape
1882	0	Luvisols	arable	LL	winter barley
1894	0	Phaeozems	arable	YK	sugar beet
194	0	Luvisols	mixed	LL	cropland fallow period
1943	0	Cambisols	mixed	BBc	cropland fallow period
199	0	Phaeozems	arable	YK	cropland fallow period
2031	0	Luvisols	mixed	LL	cropland fallow period
2032	0	Cambisols	mixed	BB	winter wheat
2034	0	Cambisols	arable	AB	other winter cereals
2037	0	Luvisols	mixed	LL	winter barley
2080	0	Luvisols	mixed	LL	cover crop (green manure)
2098	0	Stagnosols	mixed	SS	cropland fallow period
2107	0	Phaeozems	mixed	YK	winter barley
2122	0	Regosols	arable	RZ	cropland fallow period
2170	0	Stagnosols	mixed	SS	winter barley

2188	0	Anthrosols	arable	ZZ	winter barley
219	0	Gleysols	mixed	MN	cropland fallow period
2229	0	Luisols	mixed	LLh	cropland without legumes
2242	0	Cambisols	arable	BB	cropland fallow period
2254	0	Cambisols	mixed	BB	winter wheat
2259	0	Cambisols	mixed	BB	summer oats
2284	0	Stagnosols	mixed	SS	cropland fallow period
230	0	Luisols	mixed	LL	cropland fallow period
2304	0	Luisols	arable	LL	cropland fallow period
2318	0	Phaeozems	mixed	YK	cropland fallow period
2320	0	Phaeozems	mixed	YK	cropland fallow period
2330	0	Luisols	mixed	LL	cropland fallow period
2338	0	Phaeozems	mixed	YK	cropland fallow period
2362	0	Gleysols	arable	GGe	winter rape
2389	0	Luisols	mixed	LL	winter rape
2402	0	Phaeozems	mixed	YK	winter rape
2406	0	Vertisols	arable	DD	winter triticale
2407	0	Phaeozems	arable	YK	winter rye
2408	0	Phaeozems	arable	YK	cropland fallow period
2411	0	Cambisols	arable	BB	silage maize
2417	0	Phaeozems	mixed	YK	cropland fallow period
2418	0	Regosols	arable	RZ	winter wheat
2430	0	Phaeozems	arable	YK	cropland fallow period
2456	0	Luisols	mixed	LL	cropland fallow period
2460	0	Cambisols	mixed	BB	cropland fallow period
2465	0	Luisols	mixed	LL	cropland fallow period
2479	0	Fluvisols	mixed	AZ	cropland fallow period

2486	0	Cambisols	arable	BB	winter wheat
2487	0	Stagnosols	mixed	SS	winter wheat
2488	0	Luisols	mixed	LL	cropland without legumes
2509	0	Luisols	arable	LL	cropland fallow period
2512	0	Luisols	arable	LL	cropland fallow period
2519	0	Cambisols	mixed	BB	winter wheat
2572	0	Phaeozems	mixed	YK	winter wheat
2573	0	Regosols	mixed	RZ	cropland fallow period
2634	0	Luisols	mixed	LL	winter rape
2636	0	Stagnosols	mixed	SS	winter wheat
2652	0	Anthrosols	mixed	ZZ	winter barley
2660	0	Chernozems	mixed	TT	cropland fallow period
2663	0	Stagnosols	mixed	SS	winter wheat
2664	0	Stagnosols	arable	SS	winter wheat
2665	0	Luisols	mixed	LLd	cropland fallow period
2672	0	Cambisols	arable	BB	cover crop (green manure)
2677	0	Cambisols	mixed	BB	cropland fallow period
2681	0	Luisols	mixed	LL	cropland fallow period
2712	0	Luisols	mixed	LL	silage maize
2721	0	Cambisols	mixed	BB	winter rape
274	0	Luisols	arable	LL	winter rape
2755	0	Stagnosols	mixed	SH	cropland fallow period
276	0	Luisols	arable	LL	winter rape
2760	0	Stagnosols	mixed	SS	cover crop (green manure)
2762	0	Gleysols	mixed	GG	winter rape
2773	0	Luisols	mixed	LL	cropland fallow period
278	0	Stagnosols	mixed	SS	cropland fallow period

2794	0	Cambisols	mixed	BB	winter wheat
2840	0	Luvisols	mixed	LL	cropland fallow period
2845	0	Gleysols	arable	GG	cropland fallow period
2895	0	Cambisols	arable	AB	cropland fallow period
2946	0	Vertisols	mixed	DD	winter rape
2950	0	Stagnosols	arable	SS	winter barley
300	0	Stagnosols	arable	MD	cropland fallow period
3016	0	Cambisols	mixed	BB	cropland fallow period
3033	0	Chernozems	mixed	TT	cropland fallow period
3091	0	Luvisols	mixed	LL	winter rape
3093	0	Luvisols	mixed	LL	cropland fallow period
3120	0	Luvisols	mixed	LL	cropland fallow period
3163	1	Cambisols	mixed	BB	winter rye
3164	0	Cambisols	mixed	AB	cropland without legumes
318	0	Stagnosols	mixed	SS	cropland fallow period
3181	0	Cambisols	mixed	AB	cropland fallow period
3185	0	Stagnosols	mixed	SS	cropland fallow period
3203	0	Cambisols	arable	BB	cropland fallow period
3205	0	Luvisols	mixed	LL	cropland fallow period
3206	0	Luvisols	arable	LL	winter rape
3207	0	Luvisols	arable	LL	winter barley
3310	0	Phaeozems	mixed	YK	cropland fallow period
3347	0	Cambisols	mixed	AB	cropland fallow period
3404	0	Phaeozems	arable	YK	winter triticale
3406	0	Luvisols	mixed	LL	cropland fallow period
3407	0	Regosols	arable	RQ	cropland fallow period
3410	0	Phaeozems	mixed	YK	cropland fallow period

3455	0	Cambisols	mixed	BB	cropland fallow period
3470	0	Phaeozems	arable	YK	winter rape
3511	0	Cambisols	mixed	BB	cropland fallow period
3525	0	Stagnosols	mixed	SS	cover crop (green manure)
353	0	Stagnosols	arable	SS	winter wheat
3537	0	Cambisols	mixed	BB	winter barley
357	0	Stagnosols	mixed	SS	cropland fallow period
358	0	Luvisols	arable	LL	cropland fallow period
3591	0	Regosols	arable	RQ	winter wheat
3624	0	Luvisols	mixed	LL	cropland fallow period
3651	0	Phaeozems	mixed	YK	clover
3710	0	Gleysols	mixed	GG	cropland fallow period
3726	0	Phaeozems	mixed	YK	cropland fallow period
373	0	Stagnosols	mixed	SS	winter wheat
3816	0	Stagnosols	mixed	SS	winter wheat
385	0	Gleysols	mixed	MC	cropland fallow period
3855	0	Cambisols	mixed	AB	winter wheat
3865	0	Phaeozems	arable	YK	winter wheat
3919	0	Vertisols	mixed	DD	cropland fallow period
3921	0	Stagnosols	mixed	SS	other winter cereals
3975	0	Phaeozems	arable	YK	cropland fallow period
3979	0	Cambisols	mixed	BB	winter barley
401	0	Stagnosols	mixed	SS	cover crop (feed)
408	0	Stagnosols	mixed	SS	winter wheat
4084	0	Stagnosols	mixed	SS	winter wheat
4090	0	Cambisols	mixed	BB	winter wheat
4095	0	Stagnosols	mixed	SS	winter barley

4114	0	Regosols	arable	RQ	cropland fallow period
4144	0	Cambisols	arable	BB	cropland fallow period
4207	0	Stagnosols	mixed	SS	winter barley
4235	0	Cambisols	mixed	BB	winter wheat
4251	0	Stagnosols	arable	SS	cropland fallow period
4254	0	Stagnosols	mixed	SS	cropland fallow period
4263	0	Cambisols	mixed	BB	cover crop (green manure)
4308	0	Cambisols	mixed	BB	clover
4344	1	Cambisols	mixed	AB	cropland fallow period
4413	0	Stagnosols	mixed	SS	cover crop (green manure)
4520	0	Stagnosols	arable	SS	cropland fallow period
4579	0	Cambisols	mixed	BB	winter barley
4583	0	Cambisols	arable	AB	winter wheat
46	0	Gleysols	arable	MC	cropland fallow period
4633	0	Luvisols	mixed	LL	cropland fallow period
4674	0	Cambisols	mixed	BB	cropland fallow period
4675	0	Phaeozems	arable	YK	cover crop (green manure)
4676	0	Cambisols	mixed	BB	cropland fallow period
4776	0	Phaeozems	arable	YK	cropland fallow period
4779	0	Phaeozems	mixed	YK	cropland fallow period
4828	0	Cambisols	mixed	BB	winter barley
4832	0	Luvisols	mixed	LL	cropland fallow period
4850	0	Luvisols	mixed	LL	winter wheat
4886	0	Luvisols	mixed	LL	cropland fallow period
4954	0	Stagnosols	mixed	SS	cropland fallow period
6023	0	Luvisols	mixed	LLh	winter wheat
6079	0	Cambisols	mixed	AB	cropland fallow period

6097	0	Luvisols	mixed	LL	cropland fallow period
6103	0	Fluvisols	mixed	AQ	cropland fallow period
6112	0	Cambisols	mixed	AB	cropland fallow period
6127	0	Stagnosols	arable	SS	winter wheat
6135	0	Stagnosols	arable	SS	winter wheat
6140	0	Phaeozems	mixed	YK	winter wheat
6142	0	Luvisols	arable	LL	cropland fallow period
6145	0	Phaeozems	arable	YK	cropland fallow period
6152	0	Cambisols	arable	AB	winter wheat
6153	0	Phaeozems	mixed	YK	cropland fallow period
6166	0	Regosols	mixed	RQ	winter wheat
6167	0	Phaeozems	arable	YK	cropland fallow period
6172	0	Cambisols	mixed	BB	cropland fallow period
6175	0	Phaeozems	arable	YK	cropland fallow period
6177	0	Luvisols	mixed	LL	winter wheat
69	0	Stagnosols	arable	SS	cropland fallow period
721	0	Chernozems	arable	TT	winter wheat
89	0	Luvisols	mixed	LL	cropland fallow period

5 References

Bodenkundliche Kartieranleitung, 5. Auflage (2005) Bundesanstalt für Geowissenschaften und Rohstoffe in Zusammenarbeit mit den Staatlichen Geologischen Diensten (Hrsg.), 5. Auflage, Stuttgart: Schweizerbart, Hannover. ISBN 978-3-510-95920-4

Finn DR, Schroeder J, Samad MS, Poeplau C, Tebbe CC (2023) Importance of sample pre-treatments for the DNA-based characterization of microbiomes from cropland and forest soils. *Soil Biology & Biochemistry*. <https://doi.org/10.1016/j.soilbio.2023.109077>

Griffiths RI, Thomson BC, James P, Bell T, Bailey M, Whiteley, AS (2011) The bacterial biogeography of British soils. *Environmental Microbiology*, 13(6), 1642-1654. doi:10.1111/j.1462-2920.2011.02480.x

Karimi B, Terrat S, Dequiedt S, Saby NPA, Horrigue W, Lelievre M, . . . Ranjard L. (2018) Biogeography of soil bacteria and archaea across France. *Sci Adv*, 4(7), eaat1808. doi:10.1126/sciadv.aat1808

Liu B, Schroeder J, Ahnemann H, Poeplau C, Tebbe CC (2023) Crop diversification improves the diversity and network structure of the prokaryotic soil microbiome at conventional nitrogen fertilization. *Plant and Soil*. <https://doi.org/10.1007/s11104-023-06011-9>

Liu B, Arlotti D, Huyghebaert B, Tebbe CC (2022) Disentangling the impact of contrasting agricultural management practices on soil microbial communities – Importance of rare community members. *Soil Biology & Biochemistry*. 266. 108573 doi: 10.1016/j.soilbio.2022.108573

Plassart P, Prevost-Boure NC, Uroz S, Dequiedt S, Stone D, Creamer R, . . . Lemanceau P (2019) Soil parameters, land use, and geographical distance drive soil bacterial communities along a European transect. *Scientific Reports*, 9. doi:10.1038/s41598-018-36867-2

Poeplau C, Don A, Flessa H, Heidkamp A, Jacobs A, Prietz R (2020) Erste Bodenzustandserhebung Landwirtschaft – Kerndatensatz. <https://doi.org/10.3220/DATA20200203151139>

Thomson B.C, Tisserant E, Plassart P, Uroz S, Griffiths RI, Hannula SE, . . . Lemanceau P (2015) Soil conditions and land use intensification effects on soil microbial communities across a range of European field sites. *Soil Biology & Biochemistry*, 88, 403-413. doi:10.1016/j.soilbio.2015.06.012

Appendix A

Detailed analysis of WP1 (R markdown format)

Sainur Samad, Christoph C. Tebbe

12/03/2021

Step-1: Import data and merging

```
setwd("S:/Mibi/Produkte/Sainur/WP1/Data/SoilSelect/")
file1 = read.table("S:/Mibi/Produkte/Sainur/WP1/Data/SoilSelect/soiltypes.csv", sep = ";", header = T)
file2 = read.table("S:/Mibi/Produkte/Sainur/WP1/Data/SoilSelect/type of farm.csv", sep = ";", header = T)
file3 = read.table("S:/Mibi/Produkte/Sainur/WP1/Data/SoilSelect/mean_inputs.txt", sep = "\t", header = T)
file4 = read.table("S:/Mibi/Produkte/Sainur/WP1/Data/SoilSelect/lab_data.txt", sep = "\t", header = T)
file5 = read.table("S:/Mibi/Produkte/Sainur/WP1/Data/SoilSelect/general_site_information.txt", sep = "\t", header = T)
file6 = read.table("S:/Mibi/Produkte/Sainur/WP1/Data/SoilSelect/current veg.csv", sep = ";", header = T)
file7 = read.table("S:/Mibi/Produkte/Sainur/WP1/Data/SoilSelect/additional info.csv", sep = ";", header = T)
file8 = read.table("S:/Mibi/Produkte/Sainur/WP1/Data/SoilSelect/nmin_mean.txt", sep = "\t", header = T)
file9 = read.table("S:/Mibi/Produkte/Sainur/WP1/Data/SoilSelect/Cropland_PointIDs.csv", sep = ";", header = T)
file10 = read.table("S:/Mibi/Produkte/Sainur/WP1/Data/200313_BZE.und.Klima.csv", sep = ";", header = T)
```

Required packages

```
library("plyr")
library("dplyr")
library("readr")
library("tidyverse")
library("FactoMineR")
library("factoextra")
installed.packages()[names(sessionInfo())$otherPkgs, "Version"]

## factoextra FactoMineR forcats stringr purrr tidyr tibble
## "1.0.7" "2.4" "0.4.0" "1.4.0" "0.3.3" "1.0.0" "3.0.5"
## ggplot2 tidyverse readr dplyr plyr
## "3.2.1" "1.3.0" "1.3.1" "1.0.3" "1.8.4"
```

Merge all the files based on the column PointID

```
final = list(file9, file1, file2, file3, file4, file5, file6, file7, file8, file10) %>% reduce(full_join, by = "PointID")
#str(final)
dim(final) #3107 rows, 121 columns

## [1] 3107 121

#head(final)

### Data polishing
# remove NA from the columns using tidyverse package
no_NA_final = final %>% drop_na("PointID", "cropland", "Vegetation") #2234 rows, 121 columns
dim(no_NA_final)

## [1] 2234 121

## Create a new variables 'CN_ratio'
no_NA_final_CN <- no_NA_final %>% select(PointID:vegetation) %>%
  mutate(CN_ratio = TC/TN)

## Create a new variables 'air_temp_mean' from the multiple columns (air_temp_mean_13 to air_temp_mean_17)
```

```
no_NA_final_CN_temp <- no_NA_final_CN %>% select(PointID:CN_ratio) %>%
  mutate(air_temp_mean = (air_temp_mean_13+air_temp_mean_14+air_temp_mean_15+air_temp_mean_16+air_temp_mean_17)/5 )

## Create a new variables 'drought_index_mean' from the two columns (drought_index_14 and drought_index_15 )
no_NA_final_CN_temp_dro <- no_NA_final_CN_temp %>% select(PointID:air_temp_mean) %>%
  mutate(drought_index_mean = (drought_index_14+drought_index_15)/2 )

## Create a new variables 'precip_mean' from the multiple columns (precip_13 to precip_17)
no_NA_final_CN_temp_dro_pre <- no_NA_final_CN_temp_dro %>% select(PointID:drought_index_mean) %>%
  mutate(precip_mean = (precip_13+precip_14+precip_15+precip_16+precip_17)/5 )

## Create a new variables 'soil_moist_mean' from the two columns (soil_moist_14 to soil_moist_15)
no_NA_final_CN_temp_dro_pre_moi <- no_NA_final_CN_temp_dro_pre %>% select(PointID:precip_mean) %>%
  mutate(soil_moist_mean = (soil_moist_14+soil_moist_15)/2 )

#str(no_NA_final_CN_temp_dro_pre_moi)
```

Step-2: Filtering parameters

```
data2 = filter(no_NA_final_CN_temp_dro_pre_moi, pH_CaCl2 < 7.5, pH_CaCl2 > 5.5, TOC < 20,
  TOC > 10, CN_ratio < 13, Sand < 90,
  BZE_peat..yes.no. < 1,
  land.use.history.heathland..yes.no. < 1 )

#data2 = filter(no_NA_final_CN_temp_dro_pre_moi, pH_CaCl2 < 7.5, pH_CaCl2 > 5.5, TOC < 20,
#  TOC > 10, CN_ratio < 13, Sand < 90,
#  BZE_peat..yes.no. < 1,
#  land.use.history.heathland..yes.no. < 1 )

data3 = data2 %>% drop_na("mean.C.input", "eco.yes.no")
dim (data3)

## [1] 835 126

#head(data3)
```

Step-3: Subset data for PCA analysis

```
# Below variables were consired for the PCA analysis
data4 <- data3 %>% select(PointID, Silt, Sand, Clay, pH_CaCl2, TOC, TC, TN, mU, fU, fS,mS, CN_ratio,
  air_temp_mean,drought_index_mean,precip_mean, soil_moist_mean)

dim (data4)

## [1] 835 17

data4$PointID <- as.factor(data4$PointID) # converting to categorical for one column

# Create 'PointID' as row names
library(tidyverse)
data5 <- data4 %>% remove_rownames %>% column_to_rownames(var="PointID")
```



```
print(res.pca)
```

```
## **Results for the Principal Component Analysis (PCA)**
## The analysis was performed on 835 individuals, described by 16 variables
## *The results are available in the following objects:
##
## name      description
## 1 "$eig"   "eigenvalues"
## 2 "$var"   "results for the variables"
## 3 "$var$coord" "coord. for the variables"
## 4 "$var$cor" "correlations variables - dimensions"
## 5 "$var$cos2" "cos2 for the variables"
## 6 "$var$contrib" "contributions of the variables"
## 7 "$ind"   "results for the individuals"
## 8 "$ind$coord" "coord. for the individuals"
## 9 "$ind$cos2" "cos2 for the individuals"
## 10 "$ind$contrib" "contributions of the individuals"
## 11 "$call" "summary statistics"
## 12 "$call$centre" "mean of the variables"
## 13 "$call$cart.type" "standard error of the variables"
## 14 "$call$row.w" "weights for the individuals"
## 15 "$call$col.w" "weights for the variables"
```

Explanation of PCA formula:

```
#PCA(X, scale.unit = TRUE, ncp = 5, graph = TRUE)
```

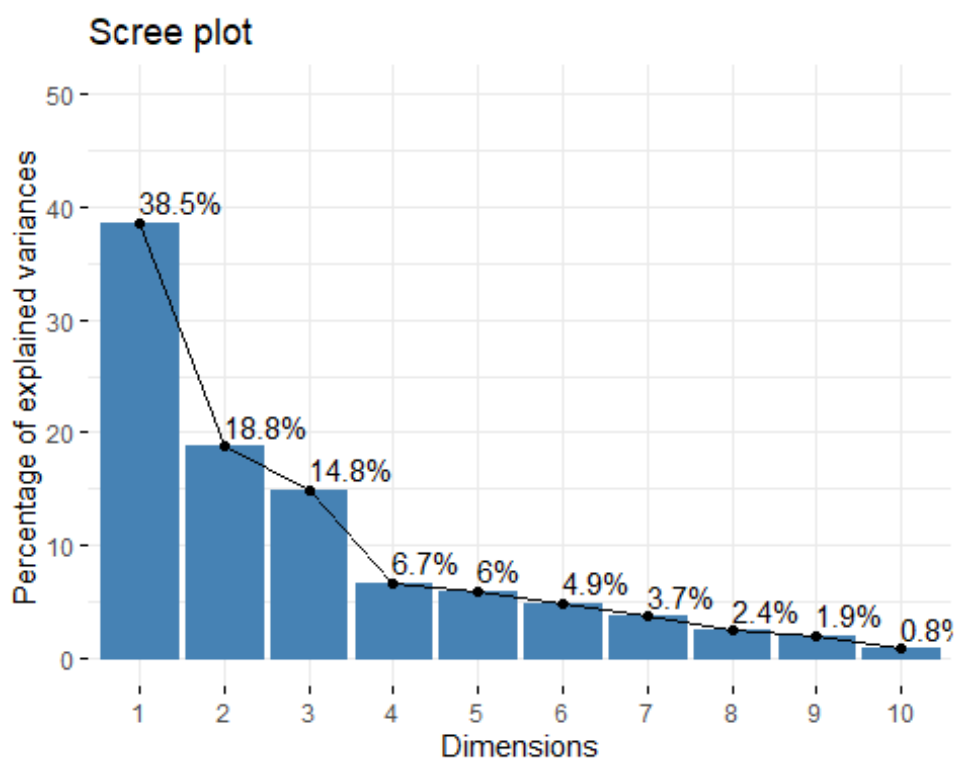
#X: a data frame. Rows are individuals and columns are numeric variables

#scale.unit: a logical value. If TRUE, the data are scaled to unit variance before the analysis. This standardization to the same scale avoids some variables to become dominant just because of their large measurement units. It makes variable comparable.

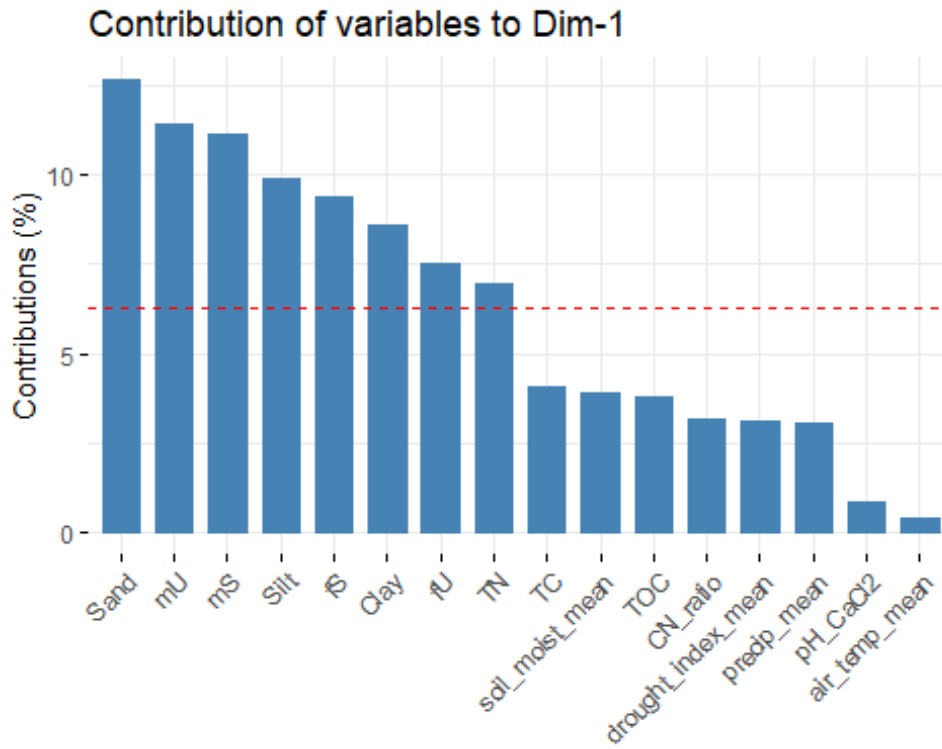
#ncp: number of dimensions kept in the final results.

#graph: a logical value. If TRUE a graph is displayed.

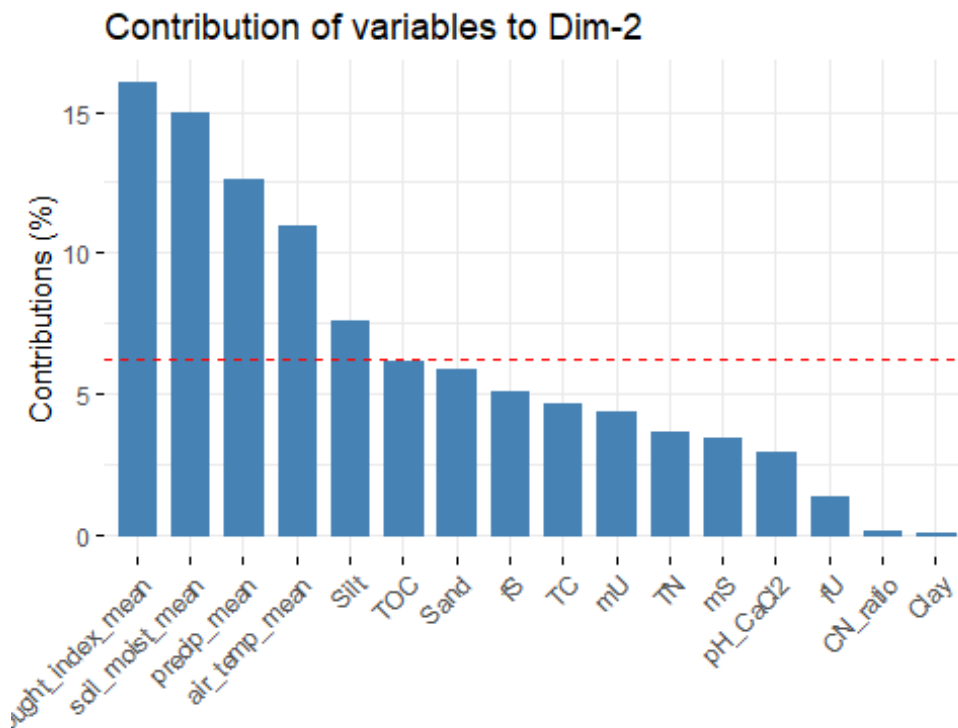
```
fviz_eig(res.pca, addlabels = TRUE, ylim = c(0, 50))
```



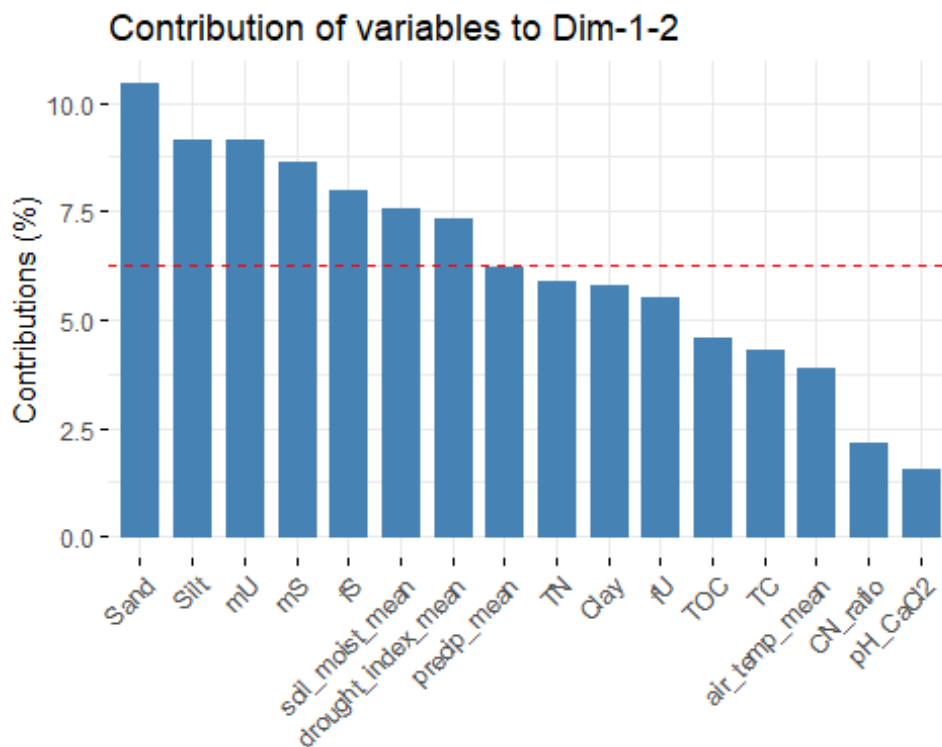
```
# Contributions of variables to PC1
fviz_contrib(res.pca, choice = "var", axes = 1, top = 20)
```



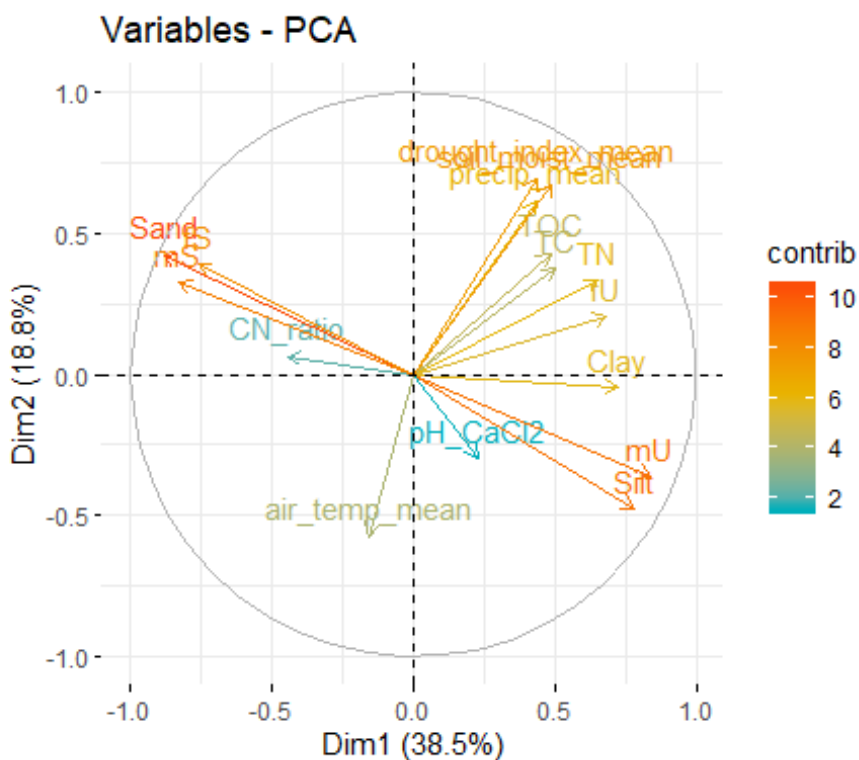
```
# Contributions of variables to PC2
fviz_contrib(res.pca, choice = "var", axes = 2, top = 20)
```



```
# Total contribution
fviz_contrib(res.pca, choice = "var", axes = 1:2, top = 20)
```



```
fviz_pca_var(res.pca, col.var = "contrib",
             gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07")
             )
```



```
res.desc <- dimdesc(res.pca, axes = c(1,2), proba = 0.05)
# Description of dimension 1
res.desc$Dim.1
```

```
## $quanti
##          correlation  p.value
## mU          0.8389172 2.777310e-222
## Silt         0.7817649 4.890990e-173
## Clay         0.7268773 4.385508e-138
## fU           0.6808217 9.278805e-115
## TN           0.6555709 1.042341e-103
## TC           0.5020643 1.594532e-54
## soil_moist_mean 0.4906634 8.496000e-52
## TOC          0.4842702 2.592485e-50
## drought_index_mean 0.4376860 2.145111e-40
## precip_mean   0.4358676 4.878633e-40
## pH_CaCl2     0.2324199 1.050909e-11
## air_temp_mean -0.1587782 4.016683e-06
## CN_ratio     -0.4419086 3.121622e-41
## fS           -0.7616636 3.855210e-159
## mS           -0.8289155 2.283565e-212
## Sand        -0.8827914 1.486612e-275
##
## attr("class")
## [1] "condes" "list"
```

```
# Description of dimension 2
res.desc$Dim.2
```

```
## $quanti
##          correlation  p.value
## drought_index_mean 0.6934567 1.018954e-120
## soil_moist_mean   0.6710877 2.275800e-110
## precip_mean       0.6144338 7.994775e-88
## TOC               0.4297076 7.606043e-39
## Sand             0.4206048 3.975427e-37
## fS               0.3898379 1.072137e-31
## TC              0.3742958 3.658187e-29
## TN              0.3329707 4.561807e-23
## mS              0.3232299 9.278431e-22
## fU              0.2039855 2.710056e-09
## pH_CaCl2       -0.2978115 1.455334e-18
## mU             -0.3634051 1.811438e-27
## Silt           -0.4770380 1.136646e-48
## air_temp_mean  -0.5732739 4.057202e-74
##
## attr("class")
## [1] "condes" "list"
```

Include Dim.1 and Dim.2 to the data fram

```
Dim1_5 <- res.pca$ind$coord
head(Dim1_5)
```

```
##   Dim.1 Dim.2 Dim.3 Dim.4 Dim.5
## 15 -2.1752496 2.5279150 0.3065205 0.4236078 -1.23362958
## 22 -0.8784968 3.0335975 1.5097615 1.4253389 -0.61861651
## 23 -1.0146731 1.9962075 -0.2571774 -0.2588570 -0.03235849
## 24 -0.4593895 2.2831376 1.7867695 -0.7026180 -0.97563428
## 33 -0.5543691 2.7568247 1.4157617 -0.1701138 -1.78707442
## 35 3.3477599 0.6497954 2.1927777 -0.1656028 0.96463077
```

```
Dim12 <- subset(Dim1_5, select = c("Dim.1", "Dim.2"))
head(Dim12)
```

```
##   Dim.1 Dim.2
## 15 -2.1752496 2.5279150
```

```

## 22 -0.8784968 3.0335975
## 23 -1.0146731 1.9962075
## 24 -0.4593895 2.2831376
## 33 -0.5543691 2.7568247
## 35 3.3477599 0.6497954

dim(Dim12)

## [1] 835 2

dim(data5)

## [1] 835 16

data7 <- merge(data5, Dim12, by = 0, all = T, sort = FALSE)
head(data7)

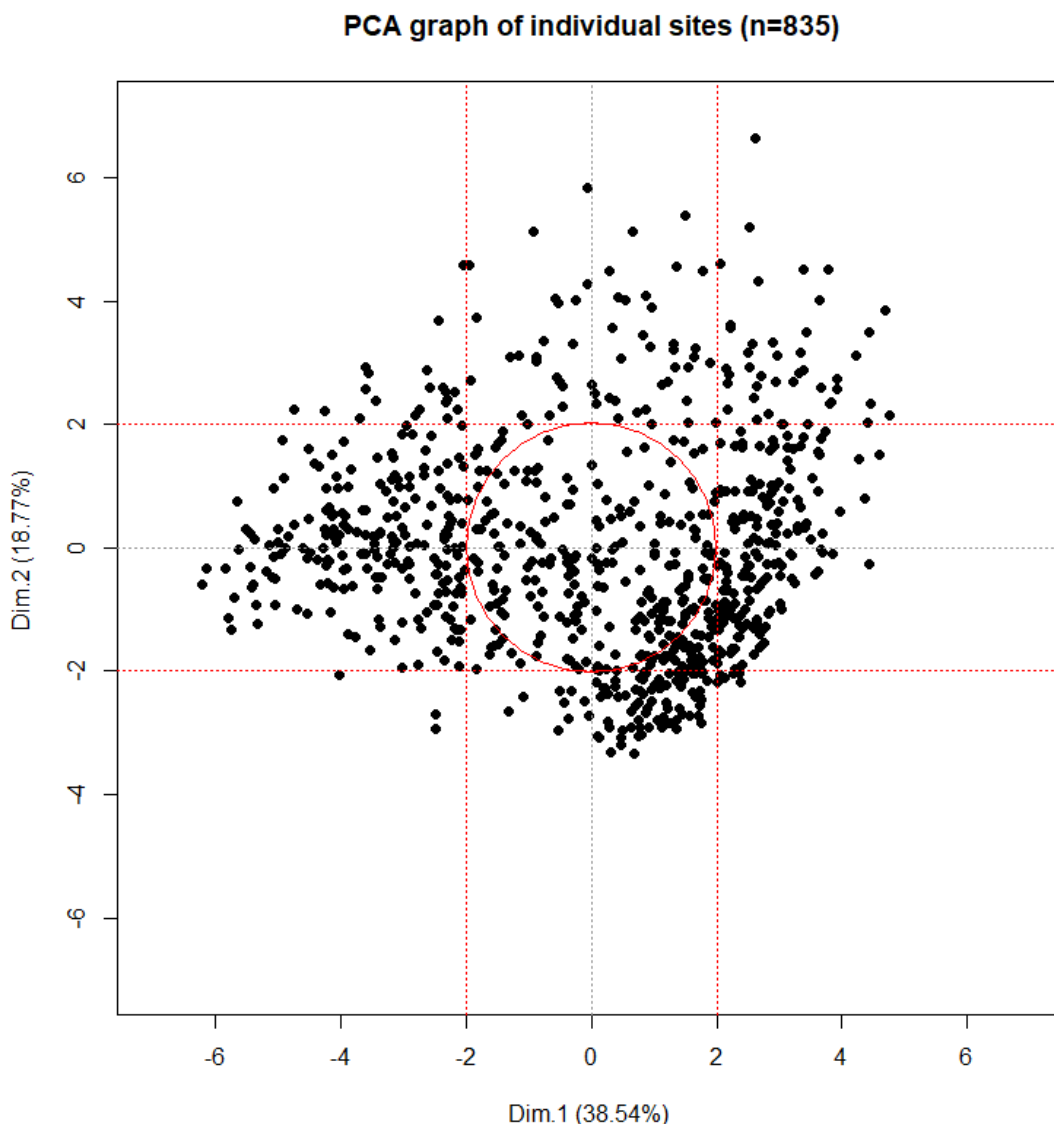
## Row.names Silt Sand Clay pH CaCl2 TOC TC TN mU fU fS mS
## 1 15 23.3 67.5 9.2 5.63 16.08 16.08 1.52 5.8 3.5 26.9 32.2
## 2 22 28.8 60.2 11.0 5.96 19.22 19.22 1.66 9.3 4.7 27.1 23.8
## 3 23 31.2 54.0 14.8 6.22 14.95 14.95 1.44 9.4 5.5 29.6 19.4
## 4 24 28.5 56.0 15.5 6.14 18.23 18.23 1.86 9.1 5.6 31.1 20.0
## 5 33 28.2 56.1 15.7 5.53 18.48 18.48 1.85 8.9 4.6 29.6 20.9
## 6 35 59.8 6.3 33.9 7.17 19.00 19.94 1.98 17.0 5.4 6.1 0.1
## CN_ratio air_temp_mean drought_index_mean precip_mean soil_moist_mean
## 1 10.578947 87.2 2.333333 331.2 75.66667
## 2 11.578313 86.8 2.500000 347.2 75.66667
## 3 10.381944 85.2 2.500000 345.2 75.16667
## 4 9.801075 85.6 2.000000 314.4 75.33333
## 5 9.989189 85.8 2.333333 317.6 75.16667
## 6 10.070707 88.2 2.333333 324.0 76.33333
## Dim.1 Dim.2
## 1 -2.1752496 2.5279150
## 2 -0.8784968 3.0335975
## 3 -1.0146731 1.9962075
## 4 -0.4593895 2.2831376
## 5 -0.5543691 2.7568247
## 6 3.3477599 0.6497954

dim(data7)

## [1] 835 19

# PCA graph of individual sites
plot(data7$Dim.1, data7$Dim.2, main="PCA graph of individual sites (n=835)",
      xlab="Dim.1 (38.54%)", ylab="Dim.2 (18.77%)", pch=19, xlim = c(-7,7), ylim = c(-7,7))
symbols(x=0, y=0, circles=2, fg="red", add=T, inches=F, xlim = c(-7,7), ylim = c(-7,7))
abline(h = 0, v = 0, col = "gray60", lty=3)
abline(h = -2, v = -2, col = "red", lty=3)
abline(h = 2, v = 2, col = "red", lty=3)

```

Step-6: Data extraction from PC1 and PC2

#Finally, filtering the dataset on the basis of scatter plot where the sample data points are located close to the center. (filtering those sites that their radius are less then 2). So, the formula we used is: $(x-0)^2 + (y-0)^2 = (radius)^2$

```
data9 = filter(data7, ((Dim.1)**2 + (Dim.2)**2) < 2**2)
```

```
head(data9)
```

```
## Row.names Silt Sand Clay pH_CaCl2 TOC TC TN mU fU fS mS
## 1 46 71.9 13.0 15.1 7.08 11.91 13.85 1.15 7.2 1.5 12.5 0.3
## 2 69 31.0 48.6 20.4 5.89 15.40 15.40 1.47 10.1 4.6 26.5 17.1
## 3 89 34.5 50.3 15.2 6.09 17.01 17.01 1.63 10.4 5.0 29.7 16.6
## 4 116 30.2 48.5 21.3 7.01 14.75 14.75 1.45 10.3 5.4 30.2 15.3
## 5 124 44.8 40.9 14.3 6.32 10.78 10.78 1.09 5.5 3.3 40.8 0.1
## 6 136 33.9 50.1 16.0 5.74 13.31 13.31 1.26 9.2 5.7 29.7 15.9
## CN_ratio air_temp_mean drought_index_mean precip_mean soil_moist_mean
## 1 12.043478 88.4 2.000000 306.4 76.16667
## 2 10.476190 90.0 1.833333 244.4 73.50000
## 3 10.435583 89.2 2.000000 293.2 74.00000
```

```
## 4 10.172414      89.6      2.000000      266.4      73.33333
## 5 9.889908      90.6      2.666667      340.8      76.00000
## 6 10.563492     88.4      2.000000      305.2      74.00000
##   Dim.1   Dim.2
## 1 -0.5164914 -1.1179016
## 2 -1.3376617 0.4024167
## 3 -0.8493524 1.2967375
## 4 -0.9765547 0.2636558
## 5 -1.6063284 0.5466100
## 6 -1.6799695 0.7455723
```

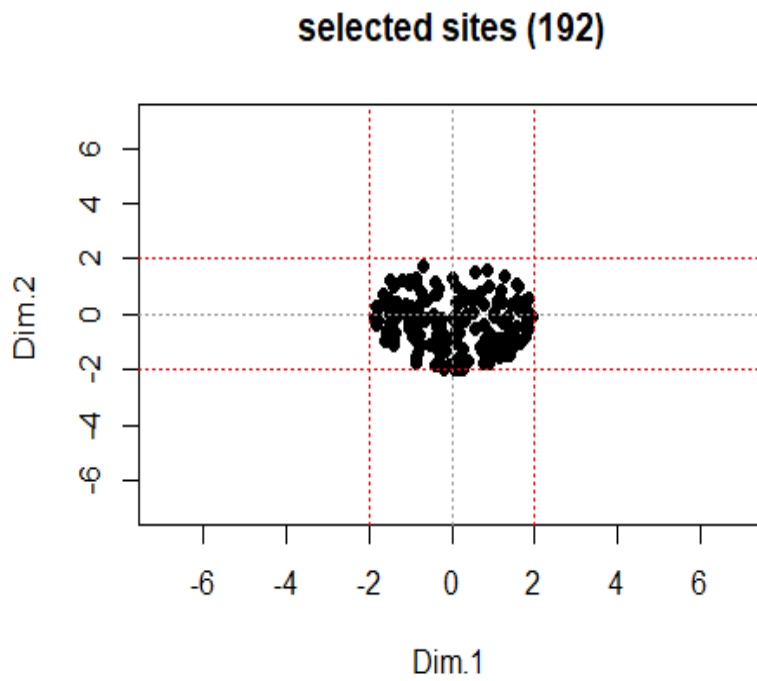
```
str(data9)
```

```
## 'data.frame': 192 obs. of 19 variables:
## $ Row.names      : 'AsIs' chr "46" "69" "89" "116" ...
## $ Silt           : num 71.9 31 34.5 30.2 44.8 33.9 56.9 30.3 33.8 37.5 ...
## $ Sand          : num 13 48.6 50.3 48.5 40.9 50.1 20.2 51 51.6 51 ...
## $ Clay          : num 15.1 20.4 15.2 21.3 14.3 16 22.9 18.7 14.6 11.5 ...
## $ pH_CaCl2      : num 7.08 5.89 6.09 7.01 6.32 5.74 7.27 6.88 6.23 6.26 ...
## $ TOC           : num 11.9 15.4 17 14.8 10.8 ...
## $ TC            : num 13.8 15.4 17 14.8 10.8 ...
## $ TN            : num 1.15 1.47 1.63 1.45 1.09 1.26 1.36 1.26 1.24 1.28 ...
## $ mU            : num 7.2 10.1 10.4 10.3 5.5 9.2 9.3 9.9 10.6 10.6 ...
## $ fU            : num 1.5 4.6 5 5.4 3.3 5.7 3.7 5.3 7.4 5.2 ...
## $ fS            : num 12.5 26.5 29.7 30.2 40.8 29.7 20 30.4 29.5 31.4 ...
## $ mS            : num 0.3 17.1 16.6 15.3 0.1 15.9 0.2 16.4 17.9 15.5 ...
## $ CN_ratio      : num 12.04 10.48 10.44 10.17 9.89 ...
## $ air_temp_mean : num 88.4 90 89.2 89.6 90.6 88.4 90.6 89.4 88.2 90 ...
## $ drought_index_mean: num 2 1.83 2 2 2.67 ...
## $ precip_mean   : num 306 244 293 266 341 ...
## $ soil_moist_mean : num 76.2 73.5 74 73.3 76 ...
## $ Dim.1         : num -0.516 -1.338 -0.849 -0.977 -1.606 ...
## $ Dim.2         : num -1.118 0.402 1.297 0.264 0.547 ...
```

```
dim (data9)
```

```
## [1] 192 19
```

```
plot(data9$Dim.1, data9$Dim.2, main="selected sites (192)",
      xlab="Dim.1 ", ylab="Dim.2 ", pch=19, xlim = c(-7,7), ylim = c(-7,7))
#symbols(x=0, y=0, circles= 2, fg = "red", add=T, inches=F, xlim = c(-7,7), ylim = c(-7,7))
abline(h = 0, v = 0, col = "gray60", lty=3)
abline(h = -2, v = -2, col = "red", lty=3)
abline(h = 2, v = 2, col = "red", lty=3)
```



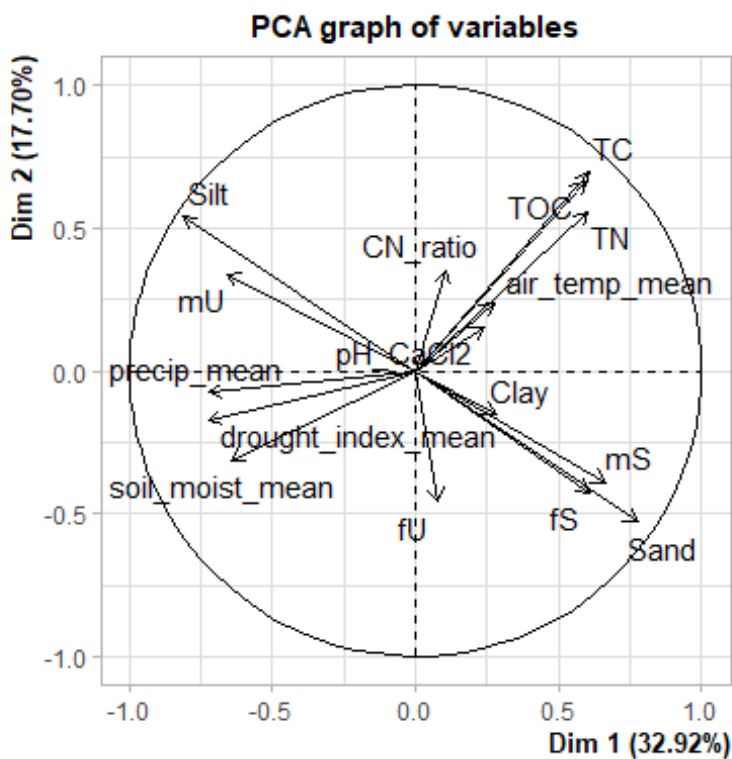
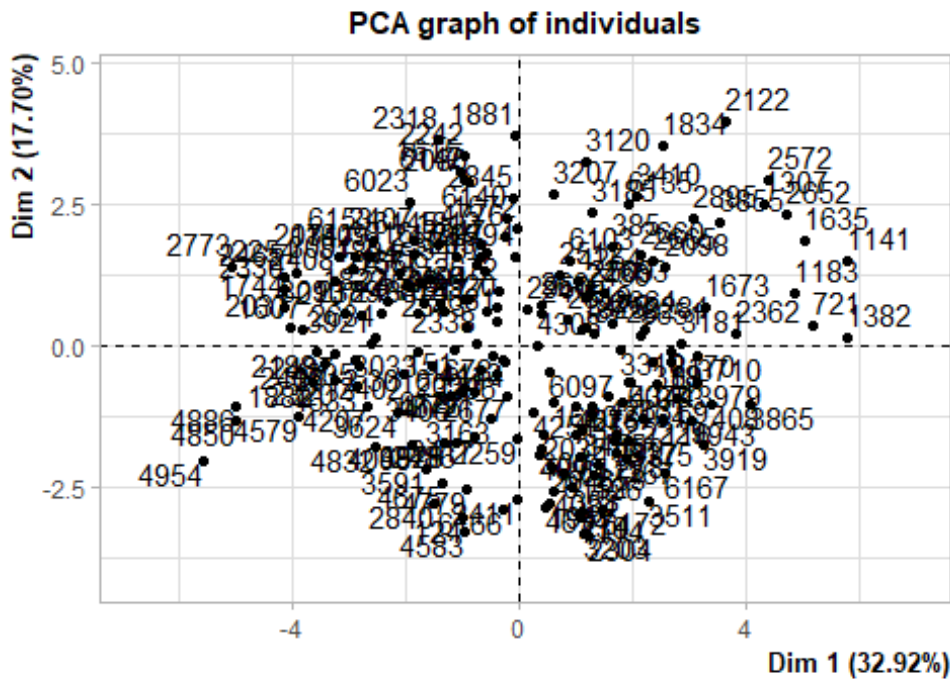
PCA plots after filtration

```
data9f <- data9 %>% remove_rownames %>% column_to_rownames(var="Row.names")

data9ff <- data9f %>% select(Silt, Sand, Clay, pH_CaCl2, TOC, TC, TN, mU, fU, fS, mS, CN_ratio,
  air_temp_mean, drought_index_mean, precip_mean, soil_moist_mean)

data9fff <- scale(data9ff)

res.pca2 <- PCA(data9fff, graph = TRUE)
```



Export 192 sites as csv format

```
data_e <- data3 %>% select(PointID, cropland, eco.yes.no, BodentypKlasse, WRB, farm.type, Hauptbodentyp, Vegetation)
data_e2 <- merge(data9, data_e, by.x = "Row.names", by.y = "PointID", all.x = TRUE)
print(data_e2$Row.names)

## [1] "1141" "116" "1183" "124" "1307" "136" "1382" "151" "1607" "1609"
## [11] "1615" "1620" "1635" "167" "1672" "1673" "1677" "1678" "1741" "1744"
```

```
## [21] "1812" "1822" "1825" "1834" "1881" "1882" "1894" "194" "1943" "199"
## [31] "2031" "2032" "2034" "2037" "2080" "2098" "2107" "2122" "2170" "2188"
## [41] "219" "2229" "2242" "2254" "2259" "2284" "230" "2304" "2318" "2320"
## [51] "2330" "2338" "2362" "2389" "2402" "2406" "2407" "2408" "2411" "2417"
## [61] "2418" "2430" "2456" "2460" "2465" "2479" "2486" "2487" "2488" "2509"
## [71] "2512" "2519" "2572" "2573" "2634" "2636" "2652" "2660" "2663" "2664"
## [81] "2665" "2672" "2677" "2681" "2712" "2721" "274" "2755" "276" "2760"
## [91] "2762" "2773" "278" "2794" "2840" "2845" "2895" "2946" "2950" "300"
## [101] "3016" "3033" "3091" "3093" "3120" "3163" "3164" "318" "3181" "3185"
## [111] "3203" "3205" "3206" "3207" "3310" "3347" "3404" "3406" "3407" "3410"
## [121] "3455" "3470" "3511" "3525" "353" "3537" "357" "358" "3591" "3624"
## [131] "3651" "3710" "3726" "373" "3816" "385" "3855" "3865" "3919" "3921"
## [141] "3975" "3979" "401" "408" "4084" "4090" "4095" "4114" "4144" "4207"
## [151] "4235" "4251" "4254" "4263" "4308" "4344" "4413" "4520" "4579" "4583"
## [161] "46" "4633" "4674" "4675" "4676" "4776" "4779" "4828" "4832" "4850"
## [171] "4886" "4954" "6023" "6079" "6097" "6103" "6112" "6127" "6135" "6140"
## [181] "6142" "6145" "6152" "6153" "6166" "6167" "6172" "6175" "6177" "69"
## [191] "721" "89"
```

```
range(data_e2$pH_CaCl2) #5.53 7.44
```

```
## [1] 5.53 7.44
```

```
count(data_e2, eco.yes.no==1)
```

```
## eco.yes.no == 1 n
## 1 FALSE 188
## 2 TRUE 4
```

```
count(data_e2, eco.yes.no==0)
```

```
## eco.yes.no == 0 n
## 1 FALSE 4
## 2 TRUE 188
```

```
unique V <- unique(data_e2$Vegetation)
```

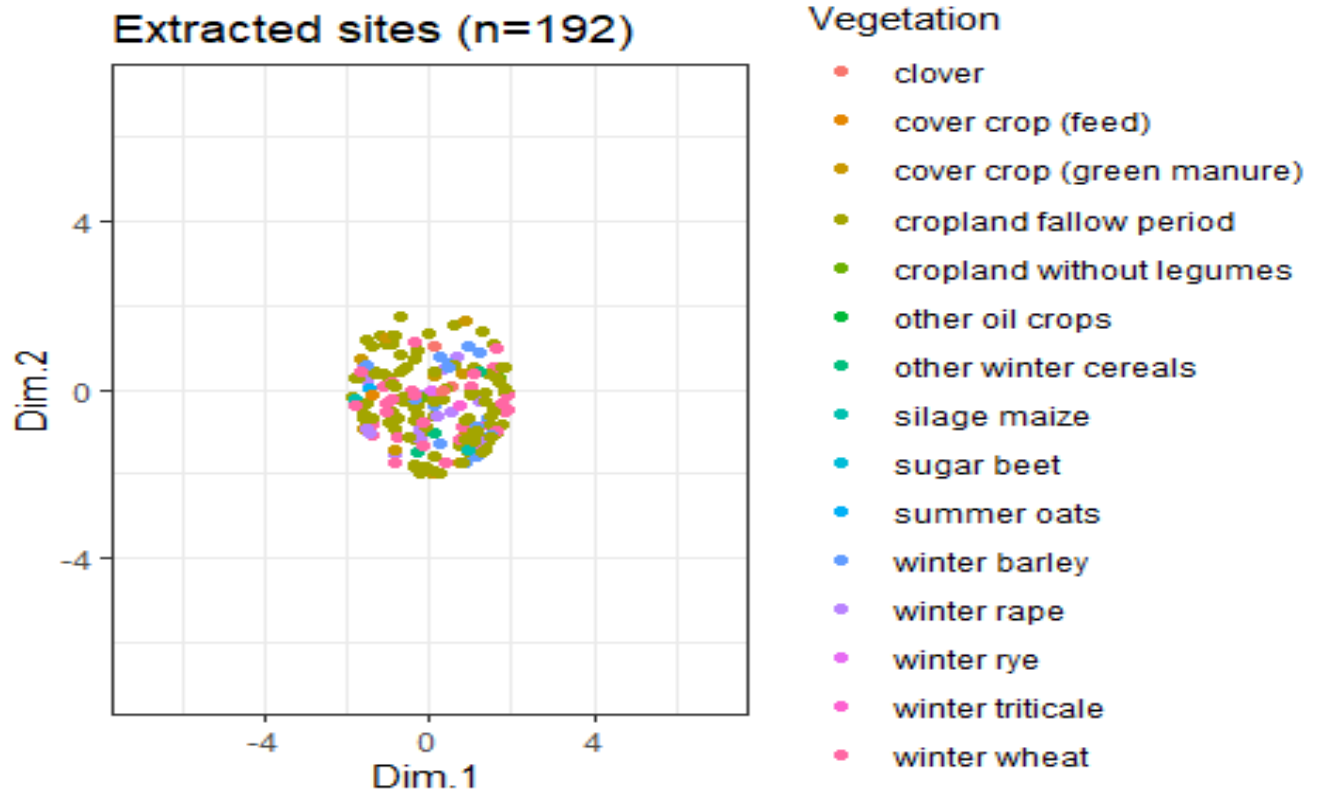
```
print(unique_V)
```

```
## [1] winter rape          cropland fallow period
## [3] other winter cereals   winter wheat
## [5] cover crop (green manure) clover
## [7] other oil crops        winter barley
## [9] sugar beet            cropland without legumes
## [11] summer oats           winter triticale
## [13] winter rye             silage maize
## [15] cover crop (feed)
## 37 Levels: 0 asparagus bush fruits clover ... winter wheat
```

```
library(ggplot2)
```

```
qq <- qplot(Dim.1, Dim.2, data = data_e2, colour = Vegetation, xlim = c(-7,7), ylim = c(-7,7), main="Extracted sites (n=192)")
```

```
qq+theme_bw()
```



```
#Rename 'Row.names' as PointID  
data_e3 <- rename(data_e2, PointID =Row.names)  
  
#dim(data_e2)  
write.csv("data_e3", "S:/Mibi/Produkte/Sainur/WP1/Data/192_sites.csv", row.names = F)
```

Zitationsvorschlag – *Suggested source citation:*

Samad, M.S., Poeplau, C., Don, A., Tebbe, C.C., 2024. Selection of cropland sites in Germany for long-term monitoring soil microbiomes as potential indicators for soil health in context of the National Soil Inventory (BZE). OpenAgrar, Göttingen. <https://doi.org/10.3220/DATA20240322093234-0>

Die Verantwortung für die Inhalte liegt bei den jeweiligen Verfassern bzw. Verfasserinnen.

The respective authors are responsible for the content of their publications.



Herausgeber/Redaktionsanschrift – *Editor/address*

Johann Heinrich von Thünen-Institut
Bundesallee 50
38116 Braunschweig
Germany

data@thuenen.de
www.thuenen.de

DOI: 10.3220/DATA20240322093234-0