

Exploring Crop Rotational Diversity and Its Role in Enhancing Soil Organic Carbon in European Agroecosystems

Nick van Dijk

Supervisor: dr. J.A. Helfenstein

Examiner: dr.ir. V.L. Mulder

March 7th, 2025

MSc Thesis

Soil Geography and Landscape Group
Wageningen University & Research

Title page background image: GRID-Arendal (2014). Flat agriculture landscape in the Netherlands.
<https://www.grida.no/resources/1701>

Abstract

The current shift towards increased intensification and specialization of European agricultural systems may lead to short-term yield benefits, but poses a threat to maintaining yield, soil health and economic stability in the long-term. Diverse cropping systems are expected to increase soil health and maintain yield, even under adverse climate conditions. Crop rotation is an example of such a diverse cropping system, which has been widely used all over Europe for centuries. Yet, there is a lack of studies on the spatial and temporal variability of crop rotational diversity (CRD) across Europe. Past research has predominantly been carried out in the United States and Asia, and has focused on comparing monoculture to a crop rotation system, with different levels of species diversity within crop rotation systems being largely overlooked. In this research, 6226 observations from the LUCAS database are combined with farm management data from the Farm Accountancy Data Network (FADN) collected over a time span of 11 years, as well as with climatic and socioeconomic variables. Based on this combined dataset, a large-scale analysis on the spatial and temporal variability of CRD in Europe was carried out, and the effect of increasing CRD on SOC stocks was investigated using Random Forest modelling. Current levels of CRD in Europe were found to be highest in Central Europe and lowest in the Alps and along the Mediterranean Coast. Overall, levels of CRD have been increasing between 2009-2020. The Random Forest model for explaining variability in CRD had a high model performance, with the share of arable area under organic farming, farm size and climatic factors being the most important predictors. The Random Forest models for explaining SOC stock changes had a low model performance, likely caused by high variability and noise within the LUCAS data. Soil OC and N content were the most important significant predictors of changes in SOC stocks. The low model performances and lack of established direct correlations between Δ CRD and Δ SOC means that the effects of varying levels of species diversity in rotational systems still remains unclear. OC data from LUCAS should be further refined in future research, in order to reduce noise and uncertainty and for clearing up what still remains unknown about the effects of CRD on SOC stocks, so that concrete policy and management recommendations can be made to help secure healthy soils and sufficient food production, even in a changing climate.

Acknowledgements

I would like to express my gratitude to my supervisor dr. Julian Helfenstein, and to dr. Sophie van Rijssel, for their valuable feedback and guidance throughout my research. I have been strongly inspired and motivated by their knowledge and dedication. Moreover, I am deeply grateful for my friends and fellow students, especially Ruben Brouwer and Kim van Etten, for thinking along and supporting me throughout my work, which has helped me to improve my research and reach my objectives.

Table of Contents

1. Introduction.....	1
2. Methodology.....	3
2.1 Data preprocessing	3
2.2 Variability and trend in crop rotational diversity.....	5
2.3 Climatic and socioeconomic variables	6
2.4. Combining FADN and LUCAS.....	7
2.5 Determining changes in SOC	7
2.6 Finding correlations between CRD, SOC and other factors	8
2.7 Predicting CRD across Europe.....	8
2.8 Explaining variability in SOC stock changes	9
3. Results.....	10
3.1 Current CRD and trends across Europe	10
3.2 Changes in SOC across Europe	11
3.3 Correlations between CRD, SOC and external variables	12
3.4 Relationships with crop types	14
3.5 Model for predicting CRD	15
3.6 Model for predicting changes in SOC	17
4. Discussion.....	20
4.1 Variability of CRD across Europe.....	20
4.2 Main factors driving SOC changes across Europe.....	22
4.3 Effect of CRD on changes in SOC stocks	23
4.4 Overcoming limitations in using LUCAS data for SOC analyses.....	23
5. Conclusion.....	25
References.....	26
Appendix I: Intermediate Model Results.....	40
Appendix II: Use of Artificial Intelligence	42

1. Introduction

CD: Crop Diversity

CR: Crop Rotation

CRD: Crop Rotational Diversity

Over the last decades, agricultural production systems have become increasingly specialized and intensified (Baldwin-Kordick et al., 2022), for example through increased fertilizer use (Tilman et al., 2002) and the reduction of crop diversity in rotational systems (Baldwin-Kordick et al., 2022; Barbieri et al., 2017). Although this intensification can lead to short-term increased productivity (Hufnagel et al., 2020), it can impose negative long-term effects on soil health (Baldwin-Kordick et al., 2022; Tilman et al., 2002) and the sustainability of the cropping systems themselves (Hufnagel et al., 2020). In order to maintain soil health and ultimately crop yields, especially in light of a growing demand for food production (Tilman et al., 2011) and the threat posed by climate change (Jägermeyr et al., 2021; Kumar, 2016; Rezaei et al., 2023), a shift towards more sustainable and future-proof agricultural systems is necessary (Bowles et al., 2020; Helfenstein et al., 2024; Yang et al., 2024).

Diversification of cropping systems is a renowned method for increasing soil health (Mortensen & Smith, 2020; Nunes et al., 2018; Yang et al., 2024). The use of a diverse cropping system enhances soil structure due to diverse root structures (Bodner et al., 2021), supports rich microbial communities (A. Williams et al., 2023), and mobilizes a wider range of nutrients (Smith et al., 2023). Moreover, crop diversification reduces the risk of climate change-induced yield reduction (Bowles et al., 2020; Bybee-Finley et al., 2024; Isbell et al., 2017; Marini et al., 2020) and increases resilience against pests and diseases (Isbell et al., 2017). Crop diversity (CD) can be increased through a number of methods, including the implementation of cover crops, intercropping or crop rotation, of which the latter is a commonly used cropping system that involves cultivating a varying sequence of different crops, in contrast to monoculture (Boincean & Dent, 2019; Y. Liu et al., 2023). Crop rotation (CR) is an effective, economic (Boincean & Dent, 2019) and sustainable system for maintaining and improving yield and soil physical and hydraulic properties under adverse growing conditions (Boincean & Dent, 2019; Costa et al., 2018; Degani et al., 2019; Kluger et al., 2022; Ouda et al., 2018).

Another important benefit of CR is the potential to increase carbon sequestration in soils (Lal, 2004a). Carbon sequestration has great potential in mitigating climate change as increasing soil organic carbon (SOC) stocks can significantly reduce atmospheric CO₂ (Lal, 2004b; Lugato et al., 2014). Several meta-analyses have emphasized this by finding positive effects on SOC stocks when shifting from a monoculture to a CR system (X. Liu et al., 2022; McDaniel et al., 2014; Schmer et al., 2020; Zheng et al., 2023). However, it remains unclear whether these benefits keep increasing with increasing levels of crop rotational diversity (CRD) (Isbell et al., 2017), i.e. the species diversity within a CR system. It is suggested that the effect may depend on the species involved, with only high residue-producing crops causing a significant effect (Omay et al., 1997). Chatterjee et al., 2016 found that only increasing CRD does not significantly increase SOC stocks. On the other hand, considering the strong negative correlation between bulk density and SOC (Fowler et al., 2023; Heuscher et al., 2005; Tsui et al., 2013), Iheshiulo et al., 2023 found that bulk density decreases with increasing CRD, which could be an indication that SOC stocks do in fact increase with increasing CRD.

In many regions of Europe, CR is embedded in traditional farming practice (Kollas et al., 2015; Kutsch et al., 2010). However, most research on the effects of CR is focused on North America, Africa and Asia, with only 4% of European case studies involved in several meta-analyses (Iheshiulo et al., 2023; Li et al., 2024; McDaniel et al., 2014; Zheng et al., 2023). This shows that there is still a shortage of European studies. Recently, research has been conducted on the spatial distribution of absolute CD (Machefer et al., 2024; Merlos & Hijmans, 2020), i.e. the number of species within a unit of area, and dominant crop sequences in rotation in Europe (Ballot et al., 2023). However, there is still a knowledge gap about the current trend and spatial variability in the diversity of crops in rotational systems. In addition, as indicated by Smith et al., 2023, most research on CRD is based on small-scale experiments or focuses on only comparing monoculture to a CR system, i.e. not on different levels of rotational diversity.

CD in Europe is threatened by the ongoing trend of agricultural specialization and intensification, as crops are increasingly grown in shorter rotations or even in monocultures (Costa et al., 2018; Marini et al., 2020; Stein & Steinmann, 2018). This trend has been described by Helfenstein et al., 2024 as a shift towards a productivist pathway characterized by high-intensity, low-diversity farming. Another threat to CD in Europe is caused by the dropping number of farmers in Europe, resulting in a relative increase in farm sizes (Piet, 2016). As CD in general is previously found to be negatively correlated with farm size (Machefer et al., 2024; Ricciardi et al., 2021), this suggests decreasing levels of CD. Accordingly, CRD may also be decreasing, under the assumption that in areas with a high CD, high CRD is found as well. On the contrary, increasing awareness among farmers about the benefits of diverse agricultural systems, such as increased soil health and lower fertilizer and pesticide requirements, incentivizes farmers to invest into the implementation of such systems (Olesen et al., 2011). This may compensate for CD loss from the shift towards more productivism, or possibly even outweigh this trend.

In this study, the knowledge gap about the spatial and temporal variability of CRD in Europe and the effects on SOC stocks will be addressed. The spatial and temporal trends of CRD in Europe over the last 15 years will be determined by using data from the Farm Accountancy Data Network (FADN) collected by the European Commission (European Commission, 2024). This will be done on a NUTS2-region scale, which closely resembles provincial area units. Furthermore, these results will be combined with data on a number of soil properties from the LUCAS Soil database (Orgiazzi et al., 2018), climate data (Beck et al., 2023; Fick & Hijmans, 2017; Zomer et al., 2022) and socioeconomic variables (Debonne et al., 2022; European Commission, 2019; Kruse et al., 2022; T. G. Williams et al., 2024). Random Forest models will be made that allow the exploration of possible relationships between CRD and soil, topography, farm management climate and socioeconomics, and how different levels of CRD affect SOC stocks on a European level. It is hypothesized that in Europe, CRD will be highest in regions dominated by small and low-intensity farms, with an overall decreasing trend due to the general shift of agricultural systems towards higher intensity farming and increased specialization. Moreover, it is expected that increasing CRD will have positive effects on enhancing SOC stocks, which the effect stagnating a certain level of diversity. Mapping the current levels and the trend in CRD across Europe, uncovering what variables relate most strongly to the spatial distribution of CRD levels, and identifying the effect on SOC stocks, offers insight in how the sustainability of agricultural systems in Europe can be improved.

2. Methodology

Before the data analysis of this research could advance, farm and soil data were preprocessed so that they could easily be implemented in the analysis. A large part of the preprocessing involved the collection of the farm management data from the FADN database. Important to note is that due to the privacy-sensitive nature of the FADN database, all data collection from the FADN database was carried out by dr. Julian Helfenstein. After the farm management data was collected, some last preprocessing steps were carried out to harmonize NUTS2 region codes to all be in the 2021 format (Eurostat, 2022b). The main analysis of this study can be divided into three sections. The first part of the methodology was devoted to the mapping of the current CRD and trend across Europe. Next, CRD and Δ SOC were further analyzed by creating Random Forest models based on soil, farm management, climate and socio-economic variables. Figure 1 shows a flowchart which highlights the main steps of the research.

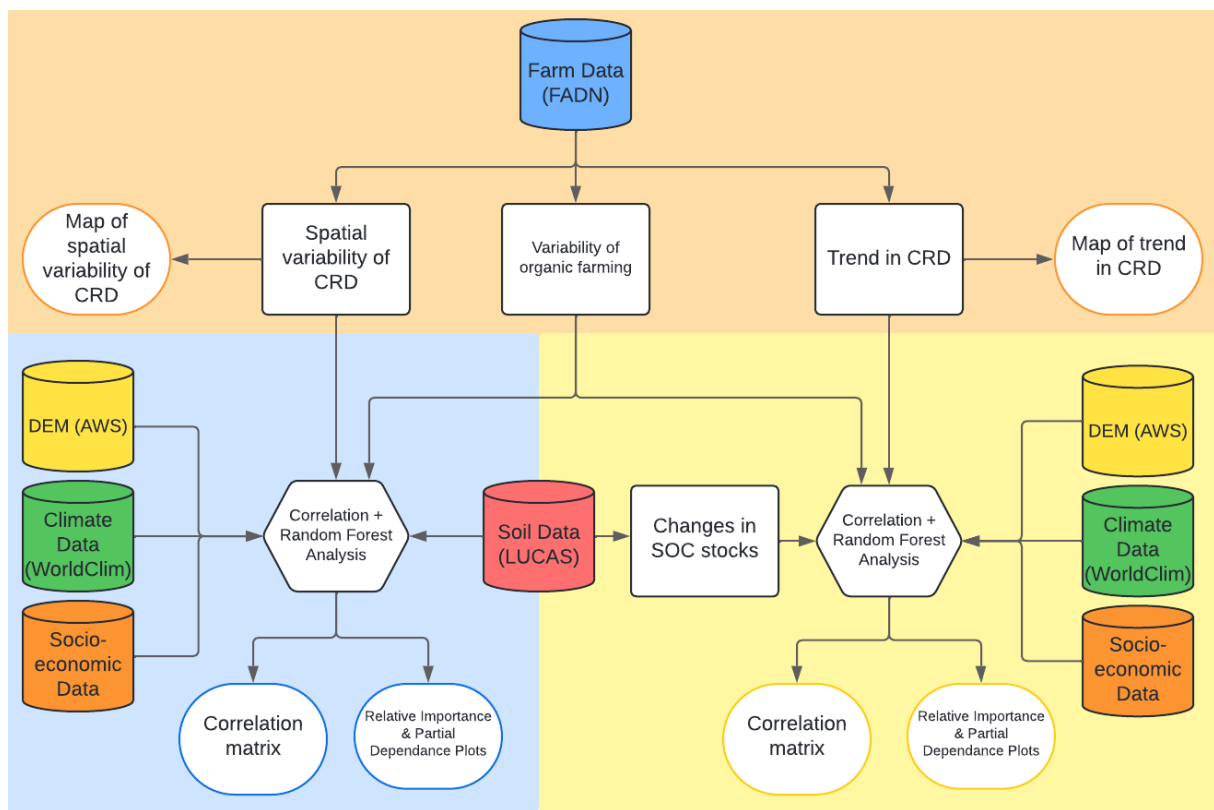


Figure 1: Schematic overview of the methodology

2.1 Data preprocessing

To protect farmers' privacy, farm-level data from the FADN was aggregated by NUTS2 region, altitude class, and crop type. A single prediction of farm management indicators was generated for each unique combination of these factors. This prediction is based on all farms in that same NUTS2 region and altitude class, which cultivated the same crop. Management indicators were predicted using weighted medians and standard deviations, in order to increase the aggregations' resemblance to reality as much as possible. The weights were determined for each farm by taking into account both the share of the cultivated crop within the total area of the farm, and the share of the cultivated area of the crop, relative to the total area in the NUTS2 region on which this crop is cultivated (Eq. 1):

$$w_i = \text{sqrt}\left(\frac{\text{area}_i}{UAA_i} * \frac{\text{area}_i}{\text{total area}}\right) \quad (\text{Eq. 1})$$

where w_i is the weight of each farm, area_i is the area on which the crop is cultivated on the farm and UAA_i is the total utilized agricultural area of the farm. Total area is the sum of all areas on which the crop is cultivated in the region ($= \sum \text{area}_i$). For analyzing the effects of the shares of different crops in rotation, the same crop categories used in Ballot et al., 2023 are used in this study (cereals, fodder legumes, maize, pulses, rapeseed, root crops, sunflower & ley (temporary grassland)), with the addition of industrial crops and vegetables. Different farm management indicators have been collected and derived from the FADN database: mineral fertilizer use, livestock density, area fraction under organic farming, crop intensity and two crop (rotational) diversity indicators: Richness and the Gini-Simpson Index.

Richness (R) was calculated as the weighted median number of arable crops per specified area, while CRD was determined using the Gini-Simpson Index (GSI). Unlike richness, which only counts the number of crops in rotation, GSI also considers crop abundance and distribution, providing a more accurate measure of true diversity. In short, the Gini-Simpson index is a widely used measure for approximating species diversity (Guiasu & Guiasu, 2012) by calculating the probability that two crops chosen at random are not part of the same species. GSI is also used in e.g. the FAO agroecology performance assessment tool (FAO, 2019). It was specifically chosen to use the GSI instead of the standard Simpson index, because the GSI represents the complimentary proportion of the standard Simpson index, thus putting the emphasis on diversity rather than uniformity. In this study, GSI is used to measure crop diversity in rotational systems, where a higher GSI signifies greater CRD. In essence, GSI is calculated as one minus the weighted mean of the proportional abundances of each crop, where the proportional abundances themselves simultaneously act as the weights (Eq. 2):

$$GSI = 1 - \sum_{i=1}^R p_i^2 \quad (\text{Eq. 2})$$

where R is richness, i.e. the total number of crops cultivated in the specified region, and p_i is the proportional abundance of each crop.

Mineral N, P, K fertilizers are calculated in kg ha⁻¹. Livestock density is calculated as the number of standard livestock units per ha. Standard livestock units are determined based on the feed requirement as compared to that of one adult cow (Eurostat, 2024). Organic farming is calculated as the crop-specific share of area under arable organic cultivation in the given region, thus equal to a probability.

The calculations of the crop intensity index (CI) were based on the C input of each crop category involved in the rotations. High crop intensities were assigned to regions with an abundance of crops having a low soil C input and or a high soil disturbance during planting, cultivation and or harvest (Engel et al., 2017; Novelli et al., 2017). More specifically, vegetables, maize and root crops were given a high CI; cereals, pulses sunflower and industrial crops a moderate CI; and fodder legumes and ley a low CI.

Although the FADN database does not contain any direct data on tillage, tillage was still deemed relevant for this study as it is often found that tillage impacts SOC stocks (Haddaway et al., 2017; Nunes et al., 2020). Tillage data (from 2016) was acquired from Eurostat (Eurostat, 2020). The data contains the areas of conventional, conservation, and zero tillage for each NUTS2 region, farm size category, and farm type combination. Tillage intensity (TI) was calculated as the area fraction under conventional tillage plus the area under conservation tillage times a factor of 0.2 (Eq. 3). No-tillage is not included in the equation, as the TI of no-tillage is 0 by definition.

$$TI = f_{conventional} + 0.2 * f_{conservation} \quad (\text{Eq. 3})$$

Therefore, a value of 100 indicates only conventional tillage being practiced in the specified area, whereas 0 indicates only no-tillage.

After gathering the farm management data, the LUCAS datasets were prepared for further analysis. The column structure of the LUCAS datasets was harmonized, and NUTS2 region codes were updated to all be in the 2021 format, in three steps. First, NUTS2 regions of which only the code had changed, were simply renamed. Points located in NUTS2 regions which were split up or merged, got assigned an updated NUTS2 region based on their NUTS3 codes. Lastly, points located in NUTS2 regions of which the geometry was heavily changed, were assigned a new NUTS2 region code by matching the 2021 region code based on overlaying a NUTS2 shapefile over the specified points. Points with an invalid land cover (LC) class, were removed, as matching these points with the FADN data would be impossible in a later stage. NUTS2 codes present in the provided FADN dataset were also harmonized with the updated LUCAS datasets.

2.2 Variability and trend in crop rotational diversity

For assessing current CRD across Europe, the weighted mean GSI for the most recent year (2020) was determined for each country and NUTS2 region and mapped using the *ggplot2* R package (Wickham, 2016). For the NUTS2 level data, the years 2018, 2019 and 2020 were aggregated under the assumption that the GSI did not change significantly between these years, in order to have sufficient data available for as many NUTS2 regions as possible.

The trend in CRD (absolute change in GSI in % per year) was calculated for each country and NUTS2 region based on linear regression from the yearly aggregations of weighted mean GSI. The standard deviations were calculated and the p-values were derived from t-tests within the context of a linear regression model with $\alpha = 0.05$. For NUTS2 regions that were split at some point in time, the GSI of the combined NUTS2-region was used for each of the breakaway regions, in order to still be able to determine a trend for these regions. For three NUTS2 regions (HR02, HR05 & HR06), only data from 2015 and 2020 was available, therefore t-tests were not applicable to assess significance for the trends in CRD for these regions, as these require at least three data points. To solve this issue, a z-test was carried out following Eq. 4, after which the p-values were determined using the *pnorm()* command in R.

$$Z = \frac{\bar{\mu}_1 - \bar{\mu}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad (\text{Eq. 4})$$

Where Z is the z-statistic, $\bar{\mu}$ the weighted mean GSI, σ the standard deviation and n the sample size for each of both years. An assumption that must be made for using z-tests is that the data is normally distributed, with a minimum suggested sample size of $n=30-50$ (Pocock, 2006). Given the large mean sample size of $n = 438$ and following the central limit theorem, the required assumption is made that the data for these three NUTS2 regions is normally distributed.

To determine the effect size, i.e. how much of the variation of CRD is attributable to differences in countries and NUTS2 regions, the η^2 -statistic was calculated. The η^2 -statistic can be calculated using ANOVA, under the assumption that the data is normally distributed and that variances are equal. Normality of the data is assumed given the median sample sizes of countries and NUTS2 regions to be 243 and 1397, respectively. However, it was found through a Levene's test that the variances between countries and NUTS2 regions are unequal, therefore an ANOVA was not considered suitable (Kim & Cribbie, 2018). For this reason, a Kruskal-Wallis test was used instead for calculating the effect sizes. As only one mean GSI and a standard deviation was available for each area, performing a Kruskal-Wallis test using only these means would strongly underestimate the spatial variation. Therefore, based on a normal distribution and using the standard deviations, 100 random simulations of the mean GSI were carried out for each NUTS2-region and country. The Kruskal-Wallis test was then done based on these simulated values. The η^2 -statistic was subsequently calculated according to Eq. 5 (Tomczak & Tomczak, 2014).

$$\eta^2 = \frac{H-k+1}{n-k} \quad (\text{Eq. 5})$$

Where H is the Kruskal-Wallis test statistic, k the number of groups (areas) and n the total number of (simulated) values. Suggested interpretations for η^2 are: 0.01-0.06 (small effect), 0.06-0.14 (moderate effect) and ≥ 0.14 (large effect) (Kassambara, 2023).

2.3 Climatic and socioeconomic variables

For allowing further in-depth analysis of the variation of CRD and variation of changes in SOC stocks across Europe, different climate and socioeconomic data that was considered to potentially be relevant, was gathered. Climate may impact levels of CRD due to differences in water availability and varying lengths of growing seasons, where longer growing seasons allow for longer and more diverse rotation schemes. Policy incentives and favorable economic conditions may cause farmers to implement higher levels of CRD (Bowles et al., 2020; Mortensen & Smith, 2020). Climate data was extracted from GeoTIFFs based on the point locations in LUCAS 2018. Climatic variables include annual mean precipitation and temperature (Fick & Hijmans, 2017), annual mean potential evapotranspiration and aridity index (AI) (Zomer et al., 2022), and Köppen climate classifications (Beck et al., 2023). The coordinate system used throughout the entire research is WGS84 (EPSG:4326). Socioeconomic variables were added to the master dataset based on either NUTS2 regions or countries, depending on the level of detail of the socioeconomic data. Socioeconomic variables used in this study are subsidy intensity, population density and GDP per capita (T. G. Williams et al., 2024), Regional Competitiveness Index (RCI) (European Commission, 2019), economic farm size typology (EFS) and age ratio (ratio old farmers (>55) to young farmers (<35)) (Debonne et al., 2022) and Environmental Policy Stringency (EPS) (Kruse et al., 2022).

2.4. Combining FADN and LUCAS

Due to issues with the standard elevation data in LUCAS, new elevation data was extracted using *get_elev_point()* from the *elevatr* R package (Hollister et al., 2023). From the newly extracted elevation data, each datapoint was assigned an altitude class matching the altitude classes in the farm management data (1 = <300 m, 2 = 300-600 m, 3 = 600< m). To make sure only agricultural land is included in the analysis, all data with a non-agricultural land use class ($\neq$ land use class U111 or U112) was removed. Also, all datapoints with a non-arable (permanent) crop land cover (LC) class were omitted. This included all points with permanent grassland not used in rotation (Costa et al., 2018; Eurostat, 2022a).

The linking of the FADN data with the LUCAS data was conducted in 3 steps. First, based on exact matches of NUTS2 region, LC class and altitude class between LUCAS and FADN datapoints, FADN data was added to the LUCAS points to generate one large combined dataset. The FADN data included fertilizer use, livestock units, share of organic farms and share of crop categories involved in rotations. 92.7% of LUCAS points had a matching NUTS2 region, LC class and altitude class with a FADN datapoint, meaning that only 7.3% of LUCAS points could not be allocated any farm management data from the FADN. In the second step, the remaining points that did not have an exact match were addressed, and the process was iterated. However, this time altitude class was ignored, thus only a matching NUTS2-region and LC class was required to successfully match FADN data to the LUCAS points. Only 3.9% of LUCAS points still could not be allocated any FADN data after this step. The last step to address the remaining datapoints without any match with the FADN data, was to look for the nearest point with the same LC class within the same country, and then copy that FADN data. In other words, the remaining datapoints got the FADN data allocated from the nearest point within the same country with the same LC class. Only 1.6% of LUCAS datapoints could not be matched with any FADN data after all three steps, so these were removed. A total of 6226 LUCAS datapoints were successfully matched with FADN data.

2.5 Determining changes in SOC

In order to calculate changes in SOC stocks, revisited points in 2009/2012 and 2018 were first identified based on reoccurring point IDs. Next, the trend in SOC in was calculated by taking the difference between 2009/2012 and 2018 and converting this to an absolute change per year, similar to De Rosa et al., 2024 (Eq. 6):

$$\Delta SOC \text{ (g C kg}^{-1}\text{y}^{-1}) = \frac{SOC_2 - SOC_1}{t_2 - t_1} \quad (\text{Eq. 6})$$

where SOC_1 and SOC_2 are the SOC contents in g C kg⁻¹ in the first year t_1 and last year t_2 . The fact that OC measurements in Bulgaria and Romania were done in 2012 instead of 2009, was taken into account. The relative change in SOC (ΔSOC_{rel}) was also calculated, according to Eq. 7.

$$\Delta SOC_{rel} \text{ (\% y}^{-1}) = \left[\left(\frac{SOC_f}{SOC_{in}} \right)^{\frac{1}{t}} - 1 \right] * 100 \quad (\text{Eq. 7})$$

Where SOC_{in} and SOC_f are the initial and final SOC contents in $g\ C\ kg^{-1}$, and t is the time period in years (6 years for Bulgaria and Romania, 9 years in all other cases). In order to increase the reliability of the used data, points were removed from the dataset when their SOC content, ΔSOC or ΔSOC_{rel} did not comply to a set of requirements:

- (1) To limit the potential effects induced by heavy peat soils, points with an SOC content of over $160\ g\ C\ kg^{-1}$ were removed, as done in a similar study by De Rosa et al., 2024. (27 points removed)
- (2) Points with a ΔSOC_{rel} beyond -50 and 100 % across the entire time period were also removed as these were considered to be nonsensical. These limits correspond with relative changes of less than -7.4 or more than 8.0 % $year^{-1}$ over a 9-year period. (324 points removed)
- (3) Points with a ΔSOC beyond -3 to $3\ g\ C\ kg^{-1}\ y^{-1}$ were removed after De Rosa et al., 2024 found approximately this range after they carried out a quality control on the LUCAS data to identify locations with extreme SOC changes unrelated to agricultural practices. (11 points removed)

5334 of 8996 total revisited points had a non-permanent agricultural land use. Of these points, 362 did not comply with all 3 requirements, leaving 4972 points with ΔSOC and ΔSOC_{rel} that comply to the specified requirements. The information on changes in SOC formed the basis of a new dataset to which FADN data was added following the previously elaborated three-step matching technique. This dataset therefore does contain datapoints without matched FADN data, but does not contain datapoints that are either not revisited or do not have a ΔSOC and ΔSOC_{rel} complying to the specified requirements. Climate and socioeconomic data was also added. Whereas the master dataset with 6226 points, all having successful matches with FADN data as explained in section 2.4, is used for the random forest model for CRD (section 2.7) , this second dataset with 4972 points is used for the analyses in sections 2.6 And 2.8. The calculated ΔSOC and ΔSOC_{rel} was mapped, and the effect size (η^2) of countries on the variance in SOC was determined with a method similar to section 2.2, however in this case simulating datapoints was not necessary.

2.6 Finding correlations between CRD, SOC and other factors

A correlogram with all relevant variables, including current CRD, ΔCRD , ΔSOC and ΔSOC_{rel} was made using the *corrplot* R package (Wei & Simko, 2024), and by using Spearman's rank correlation coefficients (ρ). Spearman's rank correlation coefficients were deemed more suitable than Pearson's correlation coefficients, as these are more robust and take into account non-linear relationships. The direct relationship between ΔSOC and ΔCRD was analyzed through the calculation of Spearman's ρ . For analyzing potential correlations between the share of crop type involved, current CRD, ΔCRD , ΔSOC and ΔSOC_{rel} , Spearman's ρ , as well as the statistical significances ($\alpha = 0.05$), were determined.

2.7 Predicting CRD across Europe

A random forest model was made to investigate which variables are most important in predicting CRD, i.e. are likely to be the biggest contributing factors to differences in CRD across Europe. As random forest analyses cannot be conducted with NA values in the dataset, this was resolved by imputing missing values using the *missForest* R package (Stekhoven & Bühlmann, 2012), with 100 decision trees. To limit the uncertainty of the imputed values, 3 variables (Ox_Al, Ox_Fe & CaCO₃) were not included in the random forest analysis due to a very high number of NAs. The random forest model was built using the *rfPermute* R package. This extends the random forest by

incorporating statistical significance testing for variable importance through permutation testing. Therefore, it allows the determination of which variables in the model are not only strongly contributing to the model's predictive performance, but also assesses which variables are statistically significant, i.e. an indicator of whether the importance is reliably above random noise. The model was made using the master dataset with imputed NA values and 500 decision trees. Individual effects of variables on the prediction of CRD were evaluated through partial dependance plots (PDPs), which were made for the 8 variables with the highest importance (%IncMSE) using the *pdp* R package (Greenwell, 2017).

2.8 Explaining variability in SOC stock changes

With the aim of assessing which variables have the largest impact on changes in SOC, and assessing whether CRD has a significant impact, two random forest models were made: one for ΔSOC and one for $\Delta\text{SOC}_{\text{rel}}$. In order to build the random forest models for ΔSOC and $\Delta\text{SOC}_{\text{rel}}$, missing values for variables other than ΔSOC and $\Delta\text{SOC}_{\text{rel}}$ were again first imputed using *missForest()* with 100 decision trees, after removing Ox_Al , Ox_Fe & CaCO_3 due to the very high number of NAs. The random forest models were then made using *rfPermute()* and 500 decision trees. Lastly, individual effects of variables in the models were again analyzed through partial dependance plots. It should be noted that the models have been altered and refined numerous times during the research, with not all steps and requirements being implemented initially. Therefore, it was only after many refinements that the final models were made.

3. Results

3.1 Current CRD and trends across Europe

The calculated current weighted mean GSI for each NUTS2 region (averages from 2018-2020) is depicted in Figure 2a. The overall lowest GSI was calculated for Crete (EL43, 0.017 ± 0.14) and the highest for Bratislava (SK01, 0.781 ± 0.18). Generally, we find low CRDs along the Mediterranean coast, the Alps, Ireland and northern Scandinavia, whereas we find high CRDs in Central Europe (predominantly east Germany, the Czech Republic and the (north) eastern regions of Slovakia and Hungary). The calculated effect sizes (η^2) from the Kruskal-Wallis tests for the NUTS2 region-level data and country-level data indicate that 32.8% of the variance in GSI is explained by the difference in NUTS2 regions, while 22.6% of the variance in GSI is explained by the difference in countries. The trend in GSI in percent change per year (2009-2020) is depicted in Figure 2b. NUTS2 regions with an insignificant trend are colored in as having no trend (0%). The most significantly negative trend is found Crete (EL43, -3.14 ± 1.37 % year⁻¹) and the most positive trend is seen in Castile and León (ES41, 2.94 ± 0.79 % year⁻¹). Generally, positive trends are found mainly in Italy, the Netherlands and northern inland Spain. Only few NUTS2 regions have a significant negative trend, mainly Crete and northern Croatia. Considering Europe as a whole, we see a generally increasing trend of CRD over 2009-2020.

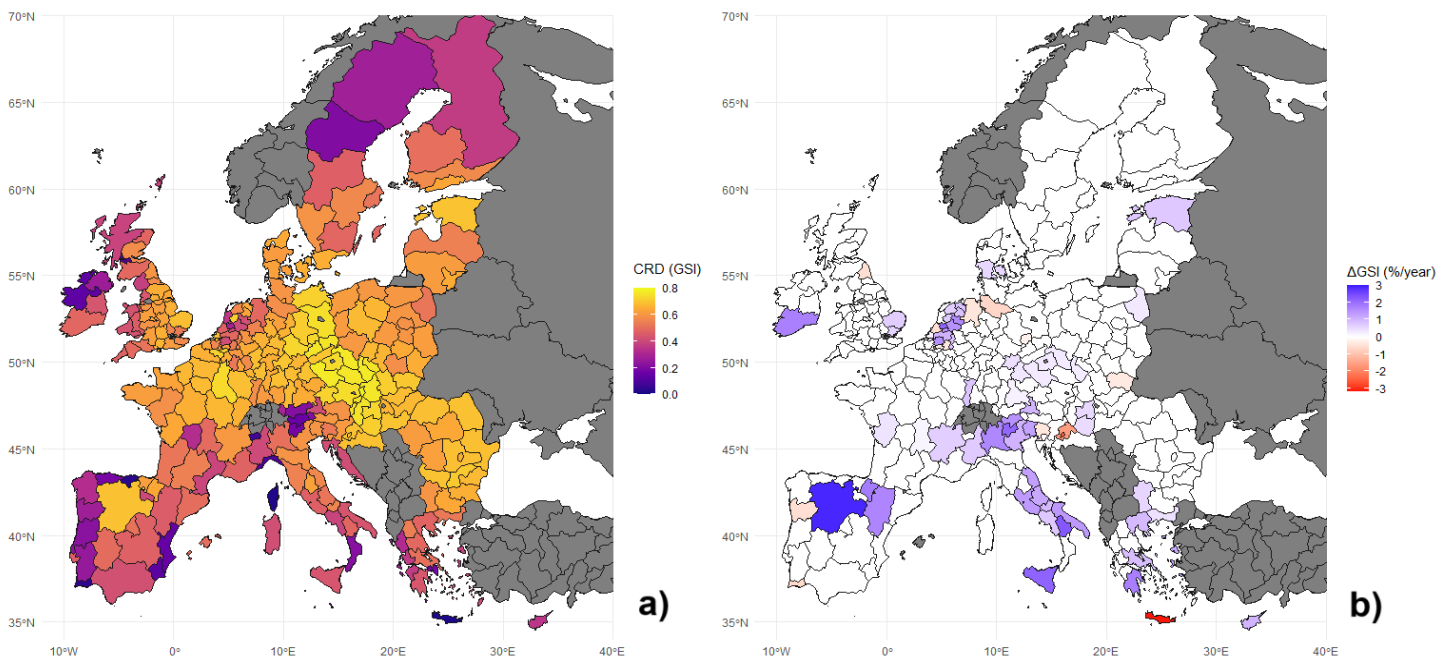


Figure 2: a) Current CRD across Europe (average 2018-2020) in GSI, b) Trend in CRD across Europe (2009-2020) in % change in GSI per year

Table 1 shows crop richness, current weighted mean GSI, the trend in GSI and the associated p-value for each country. As crop richness is closely related to CRD, lower crop richness is generally found in countries with also a low GSI. The country with the lowest GSI is Malta (0.17 ± 0.18) and the country with the highest GSI is the Czech Republic (0.759 ± 0.20). Only 10 out of 28 countries have had a significant trend in GSI over the years 2009-2018. Portugal has had the most negative trend (-0.41 ± 0.37 % year⁻¹), yet it was found to be statistically insignificant ($p > 0.05$). Considering only significant trends, no country has a negative trend in CRD. France has the smallest significant trend (0.14 ± 0.08 % year⁻¹), whereas Spain has the highest trend (1.48 ± 0.51 % year⁻¹).

Table 1: Crop richness, current weighted mean GSI, the trend (% year⁻¹) in GSI and the associated p-value for each country. Asterisks indicate statistical significance: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Country	Richness ($\pm\sigma$)	GSI ($\pm\sigma$)	Trend (%/y) ($\pm\sigma$)	P-value	Country	Richness ($\pm\sigma$)	GSI ($\pm\sigma$)	Trend (%/y) ($\pm\sigma$)	P-value
AUT	3.962 $\pm$ 1.305	0.698 $\pm$ 0.219	0.35 $\pm$ 0.36	0.118	HUN	4.604 $\pm$ 1.457	0.701 $\pm$ 0.261	0.19 $\pm$ 0.2	0.119
BEL	3.529 $\pm$ 1.284	0.61 $\pm$ 0.261	0.08 $\pm$ 0.19	0.44	IRL	2.05 $\pm$ 0.656	0.463 $\pm$ 0.259	1.16 $\pm$ 0.36	0.006 **
BGR	4.309 $\pm$ 1.463	0.663 $\pm$ 0.254	0.04 $\pm$ 0.49	0.872	ITA	2.935 $\pm$ 1.166	0.502 $\pm$ 0.278	0.96 $\pm$ 0.56	0.031 *
CYP	1.657 $\pm$ 0.704	0.352 $\pm$ 0.241	1.03 $\pm$ 0.46	0.015 *	LTU	3.745 $\pm$ 1.262	0.604 $\pm$ 0.22	0.24 $\pm$ 0.44	0.302
CZE	5.348 $\pm$ 1.656	0.759 $\pm$ 0.202	0.21 $\pm$ 0.1	0.016 *	LUX	3.412 $\pm$ 0.987	0.637 $\pm$ 0.182	0.09 $\pm$ 0.29	0.543
DEU	3.949 $\pm$ 1.168	0.698 $\pm$ 0.219	-0.01 $\pm$ 0.15	0.871	LVA	2.982 $\pm$ 1.083	0.542 $\pm$ 0.25	0.06 $\pm$ 0.29	0.659
DNK	2.688 $\pm$ 0.952	0.633 $\pm$ 0.204	0.57 $\pm$ 0.41	0.053	MLT	1.662 $\pm$ 0.498	0.17 $\pm$ 0.182	-0.24 $\pm$ 0.21	0.085
ESP	2.469 $\pm$ 1.08	0.518 $\pm$ 0.292	1.48 $\pm$ 0.58	0.011 *	NLD	2.817 $\pm$ 1.039	0.567 $\pm$ 0.284	0.78 $\pm$ 0.2	0.003 **
EST	4.098 $\pm$ 1.336	0.701 $\pm$ 0.265	0.75 $\pm$ 0.25	0.007 **	POL	3.462 $\pm$ 1.192	0.633 $\pm$ 0.222	0.04 $\pm$ 0.07	0.312
FIN	2.293 $\pm$ 0.912	0.467 $\pm$ 0.249	0.01 $\pm$ 0.31	0.953	PRT	1.886 $\pm$ 0.89	0.272 $\pm$ 0.239	-0.41 $\pm$ 0.37	0.091
FRA	3.595 $\pm$ 1.354	0.625 $\pm$ 0.277	0.14 $\pm$ 0.08	0.032 *	ROU	3.853 $\pm$ 1.189	0.65 $\pm$ 0.251	0.27 $\pm$ 0.69	0.441
GBR	2.984 $\pm$ 1.108	0.573 $\pm$ 0.291	0.37 $\pm$ 0.22	0.033 *	SVK	4.895 $\pm$ 1.505	0.745 $\pm$ 0.205	-0.09 $\pm$ 0.16	0.303
GRC	2.975 $\pm$ 1.172	0.479 $\pm$ 0.257	0.75 $\pm$ 0.34	0.016 *	SVN	3.515 $\pm$ 1.304	0.615 $\pm$ 0.277	-0.19 $\pm$ 0.22	0.143
HRV	4.05 $\pm$ 1.265	0.663 $\pm$ 0.242	-0.07 $\pm$ 0.52	0.814	SWE	2.86 $\pm$ 1.08	0.516 $\pm$ 0.267	0.08 $\pm$ 0.48	0.72

3.2 Changes in SOC across Europe

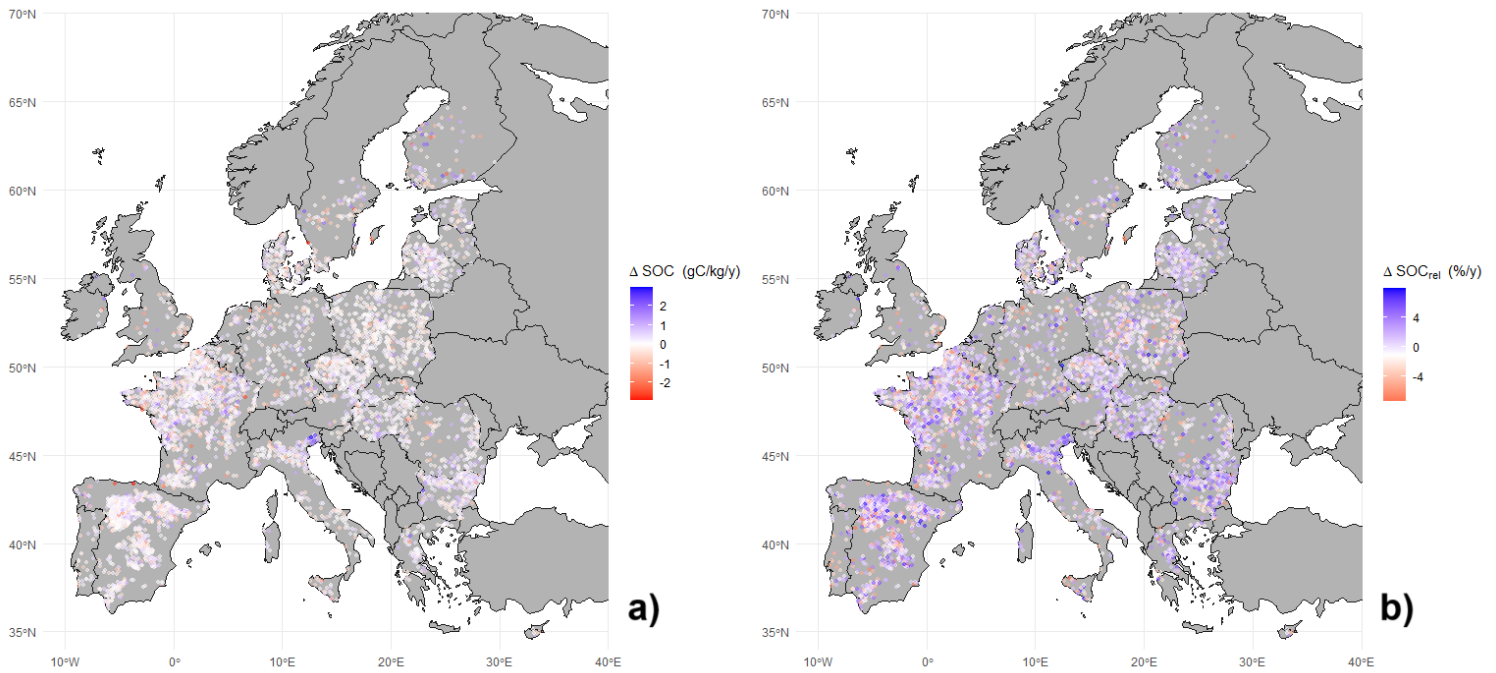


Figure 3: a) map of ΔSOC across Europe (2009/2012-2018), b) map of ΔSOC_{rel} across Europe (2009/2012-2018)

Figures 3a and 3b show the map of Europe with the locations of all used datapoints with the observed ΔSOC and $\Delta\text{SOC}_{\text{rel}}$ from 2009/2012 to 2018 in $\text{g C kg}^{-1} \text{y}^{-1}$ and $\% \text{ y}^{-1}$, respectively. Mean ΔSOC is $0.03 \pm 0.50 \text{ g C kg}^{-1} \text{y}^{-1}$ and mean $\Delta\text{SOC}_{\text{rel}}$ is $0.20 \pm 2.72 \% \text{ y}^{-1}$ compared to the initial value. ΔSOC and $\Delta\text{SOC}_{\text{rel}}$ do not show any clear spatial patterns, indicating that changes in SOC stocks are very much dependent on local factors. The only area with large scale increasing ΔSOC and $\Delta\text{SOC}_{\text{rel}}$ is northeastern Italy. The calculated effect sizes (η^2) are 0.014 and 0.012 for ΔSOC and $\Delta\text{SOC}_{\text{rel}}$, respectively, which points to there being no large spatial patterns that differ between countries.

3.3 Correlations between CRD, SOC and external variables

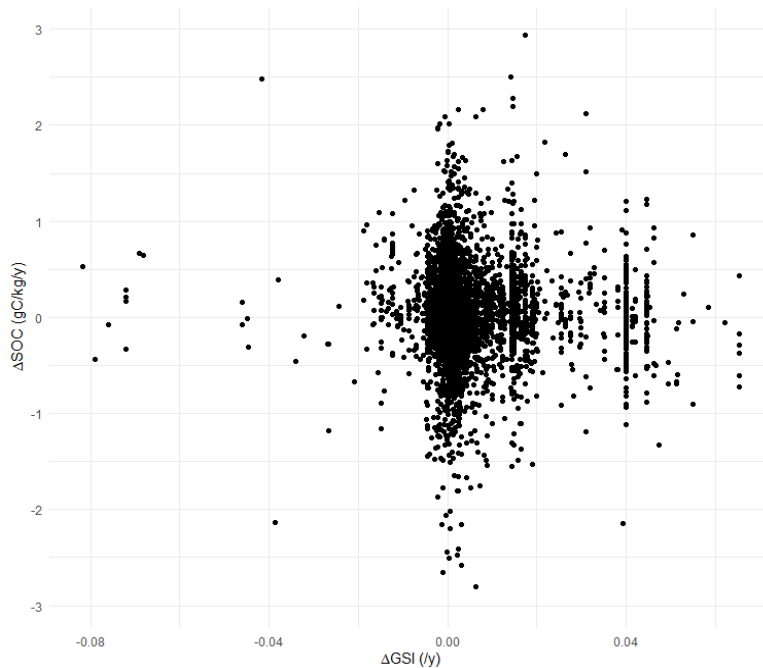


Figure 4: Absolute change in SOC versus absolute change in GSI

Figure 4 shows a scatterplot of the absolute change in SOC against the absolute change in GSI per year for each of the datapoints. The spread in ΔSOC appears to be larger than the spread in ΔGSI . With a Spearman's ρ of -0.02 ($p = 0.1037$), no direct correlation between trends in CRD and SOC stocks is found. Some vertical lines appear due to a high number of points in the same area (e.g. points in NUTS2-region ES41 where barley is cultivated) being assigned the same ΔGSI in the matching process, but having different ΔSOC .

Figure 5 shows the correlogram with the Spearman's ρ of all variables taken into account in the analyses. From the correlogram can be derived that current CRD has a high correlation with median arable area, and a negative correlation with the fraction under organic farming and tillage intensity. This indicates that generally, higher CRD is found in areas with large farm sizes, a low fraction under organic farming and a low tillage intensity. The correlations of ΔCRD with climatic factors are showing higher positive trends in areas with arid and warm climates. Moreover, ΔCRD is positively correlated to age ratio and negatively correlated to EPS, indicating that the ΔCRD is higher in areas with aging farmer populations and a low environmental policy stringency. ΔSOC and $\Delta\text{SOC}_{\text{rel}}$ show very few strong correlations with other variables. ΔSOC and $\Delta\text{SOC}_{\text{rel}}$ is however positively correlated to OC and N, which is a sign that soils which are high in OC and N also facilitate higher ΔSOC and $\Delta\text{SOC}_{\text{rel}}$.

From the correlation matrix can also be deduced that socioeconomic indicators such as EPS and RCI are generally correlated with higher precipitation, lower potential ET_0 , lower temperatures and lower elevation. Similar correlations are found for livestock intensity. Correlations of CRD, ΔCRD , ΔSOC and $\Delta\text{SOC}_{\text{rel}}$ with shares of crop types involved in rotation will be laid out in more detail in section 3.4.

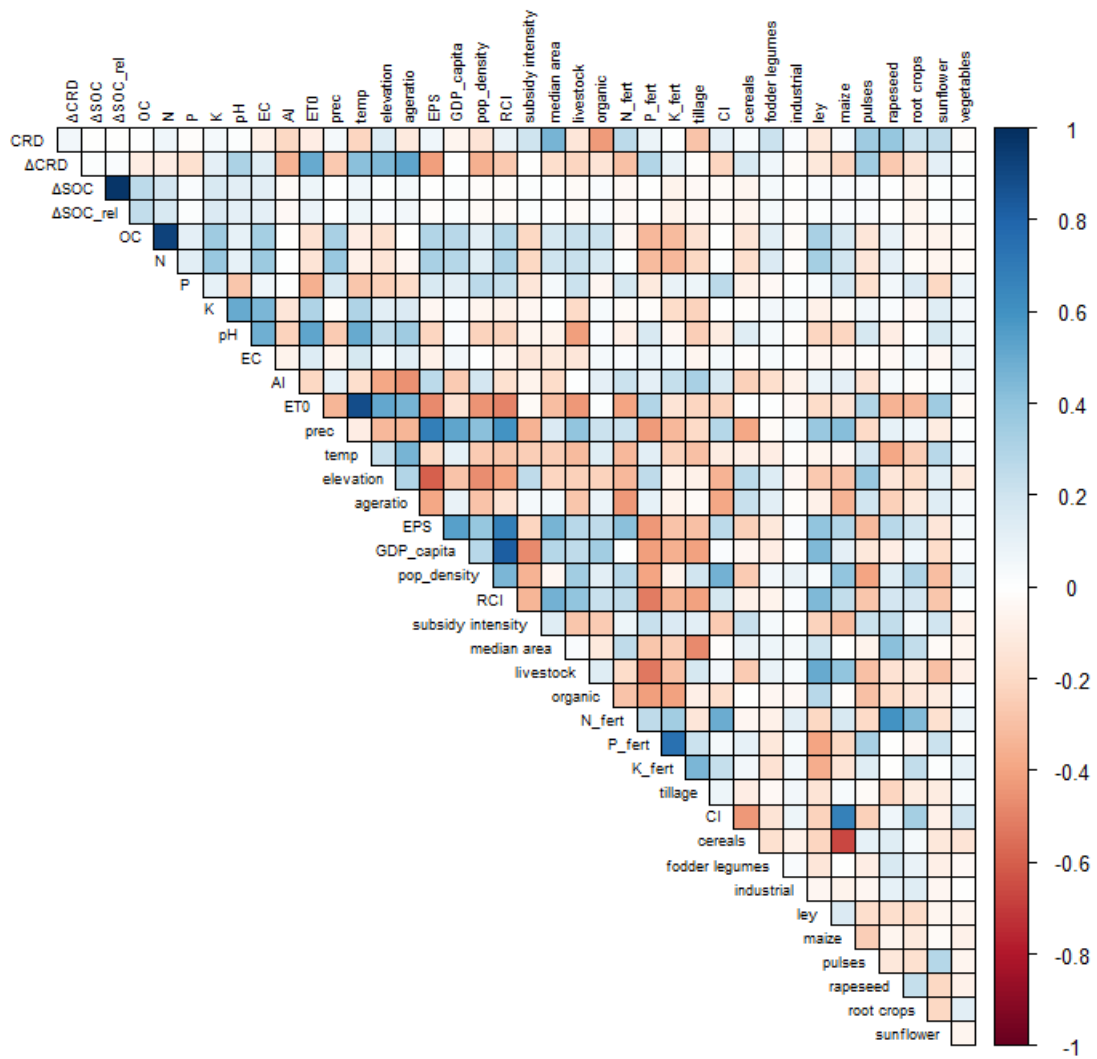


Figure 5: Correlogram including all variables used in the analyses.

3.4 Relationships with crop types

In order to check whether certain crop types are most often found in diverse rotations, and which crop types have the most positive effect on SOC stocks when included in a rotation the Spearman's ρ of correlations between current CRD (GSI), Δ CRD (Δ GSI), Δ SOC and Δ SOC_{rel}, and shares of crop types involved in rotation are shown in Table 2. The asterisks indicate significance levels, with more stars indicating a lower p-value and thus a higher significance.

Current CRD is most positively significantly related to root crops, pulses and rapeseed, indicating that these crops are most likely to be found in more diverse rotations. Δ CRD is most positively significantly related to

pulses, and negatively to rapeseed and maize, showing that the Δ CRD is most strongly increasing in areas with many pulses being cultivated, and decreasing in areas with a large share of maize and/or rapeseed being cultivated. Although they can be considered confounding factors, Δ SOC and Δ SOC_{rel} generally appear to be less strongly correlated to crop type, but there are still some statistically significant relationships. Both Δ SOC and Δ SOC_{rel} are most negative in areas with a high share of cereals and root crops, and most positive in areas with a high share of fodder legumes, maize and sunflower. Although it was hypothesized, no statistically significant relation with ley (temporary grassland) is found.

Table 2: Spearman's ρ of correlations between current CRD, Δ CRD, Δ SOC and Δ SOC_{rel}, and shares of crop types involved in rotation. Asterisks indicate significance: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Crop Type	ρ_{GSI}	$\rho_{\Delta\text{GSI}}$	$\rho_{\Delta\text{SOC}}$	$\rho_{\Delta\text{SOC}_{\text{rel}}}$
cereals	0.02	0.177***	-0.054***	-0.052***
fodder legumes	0.094***	-0.011	0.03*	0.035*
industrial	-0.025	-0.02	0.002	0.004
ley	-0.057***	-0.07***	0.026	0.02
maize	0.172***	-0.208***	0.039**	0.036*
pulses	0.298***	0.334***	0.011	0.009
rapeseed	0.365***	-0.228***	0.005	0.003
root crops	0.247***	-0.123***	-0.047***	-0.047**
sunflower	0.191***	-0.044**	0.059***	0.055***
vegetables	-0.008	0.025	0.009	0.011

3.5 Model for predicting CRD

The 6226 datapoints used for the Random Forest model for predicting and explaining variability in CRD are shown in Figure 6. The spatial spread in used datapoints shows an overall large spread across Europe, although there are a number of regions that either have a large concentration of datapoints, or completely lack any datapoints. The highest concentrations of points is found in central Spain, France, Poland and the Czech Republic. The Alps, the Spanish, French and Greek Mediterranean coasts, western Ireland and northern Scandinavia have very few to no datapoints.

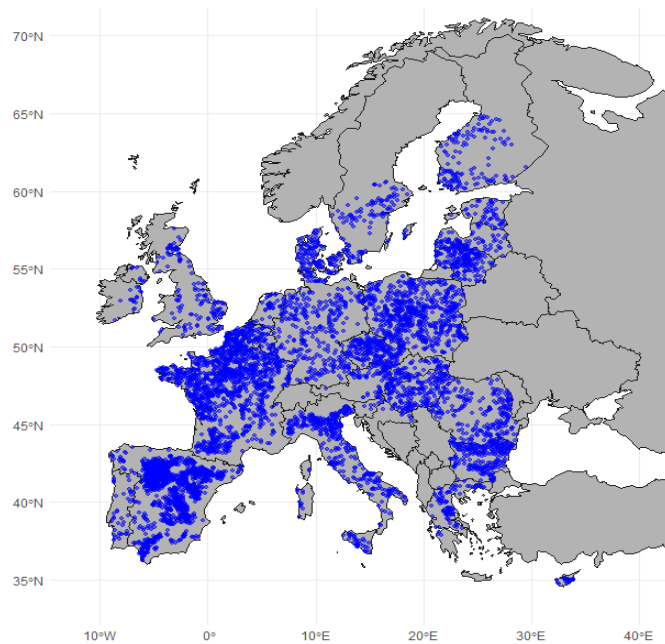


Figure 6: Locations of datapoints used for the RF model for predicting/explaining variability in CRD

Figure 7 shows the relative importance of each variable in the Random Forest model for CRD. Bars of significant variables ($p < 0.05$) are colored in. The overall R^2 of the model is 0.731 (RMSE = 0.028, ME = 0.00024, mean nodes = 4028). The fraction of arable land under organic cultivation is the most important variable, leading to a 57.5 % increase in mean squared error (%IncMSE) of the model if this variable is removed. Median area is the second most important predictor, followed by mean annual temperature, mean annual precipitation and elevation. Overall, the most important categories of predictors of CRD are farm management, climate and socioeconomics, respectively. Soil does not appear to play any significant role in predicting variation in CRD.

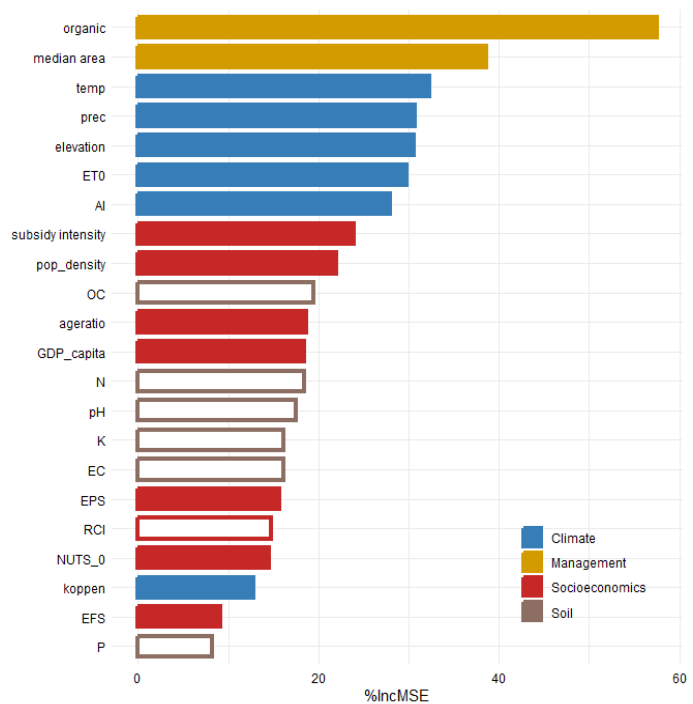


Figure 7: Relative importance of variables in the RF model for CRD. Bars without a fill indicate that this variable did not contribute significantly

Figure 8 shows the partial dependence plots of the 8 most important predictors of CRD in the model. It is important to note that Random Forest models do not necessarily express causality, even though the partial dependence plots indicate the effect of the explanatory variables in the predicted variable. The fact that an explanatory variable may show a certain effect on the predicted variable could also be caused due to the opposite: the predicted variable affecting the explanatory variable, thus showing a certain correlation in the data.

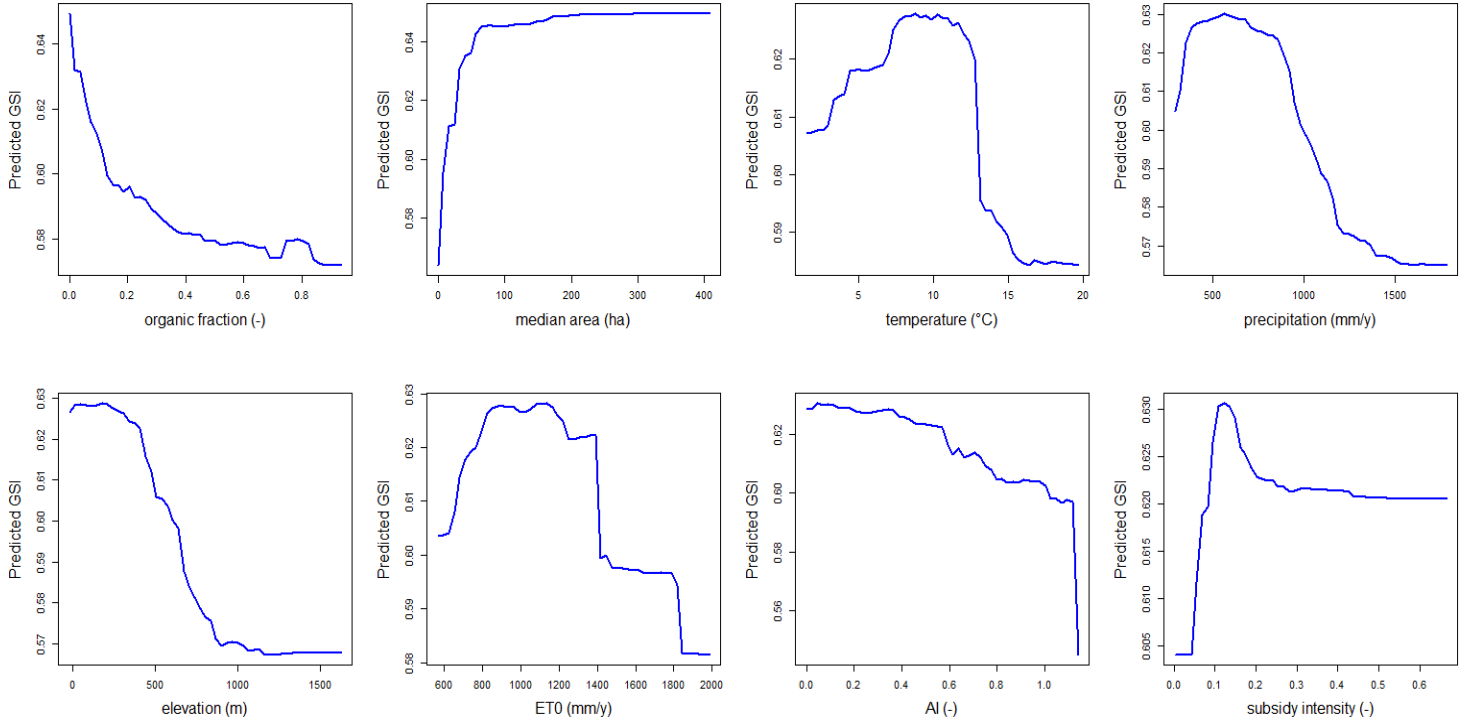


Figure 8: Partial dependence plots of the 8 most important variables in the RF model for CRD

The partial dependence plots highlight the negative correlation of the organic fraction as an isolated variable with CRD. In contrast, median arable area corresponds positively to the magnitude of CRD. Climatic/topographic variables except AI and elevation have a strongly non-linear impact, showing optimum values. The optimum values for mean annual temperature, mean annual precipitation and annual mean potential evapotranspiration are approximately 10°C, 600 mm y⁻¹ and 1000 mm y⁻¹, respectively. CRD decreases with increasing elevation and AI. Subsidy intensity also has an optimum value leading to the highest predicted CRD, namely ~0.15.

3.6 Model for predicting changes in SOC

Both Random Forest models for predicting changes in SOC have a low R^2 , indicating poor model performance. The model for ΔSOC has an R^2 of 0.099 (RMSE = 0.198, ME = -0.004, mean nodes = 3373) and the model for ΔSOC_{rel} an R^2 of 0.111 (RMSE = 1.071, ME = 0.002, mean nodes = 3369). Figures 9a and 9b show the relative importance of variables in the Random Forest model for ΔSOC and ΔSOC_{rel} , respectively. Overall, initial OC and N are the most important variables in the model, and contribute significantly in both models. The model for ΔSOC_{rel} has considerably more significantly contributing variables than the model for ΔSOC , regardless of having similar R^2 . Soil factors generally show the largest importance in predicting ΔSOC and ΔSOC_{rel} , followed by climate, management and socioeconomics equally, and lastly crop types. Although not all having a high relative importance all socioeconomic variables showed a significant contribution to the model. Countries (NUTS_0) appear to have a remarkably large importance in predicting SOC changes, meaning that the predictions are somewhat country-dependent, even though the effect sizes of countries on the variation in ΔSOC and ΔSOC_{rel} was found to be very small in section 3.2. CRD is the 10th most important predictor in the model for ΔSOC_{rel} , and is a significant contributor. It is also the most important management predictor in this model.

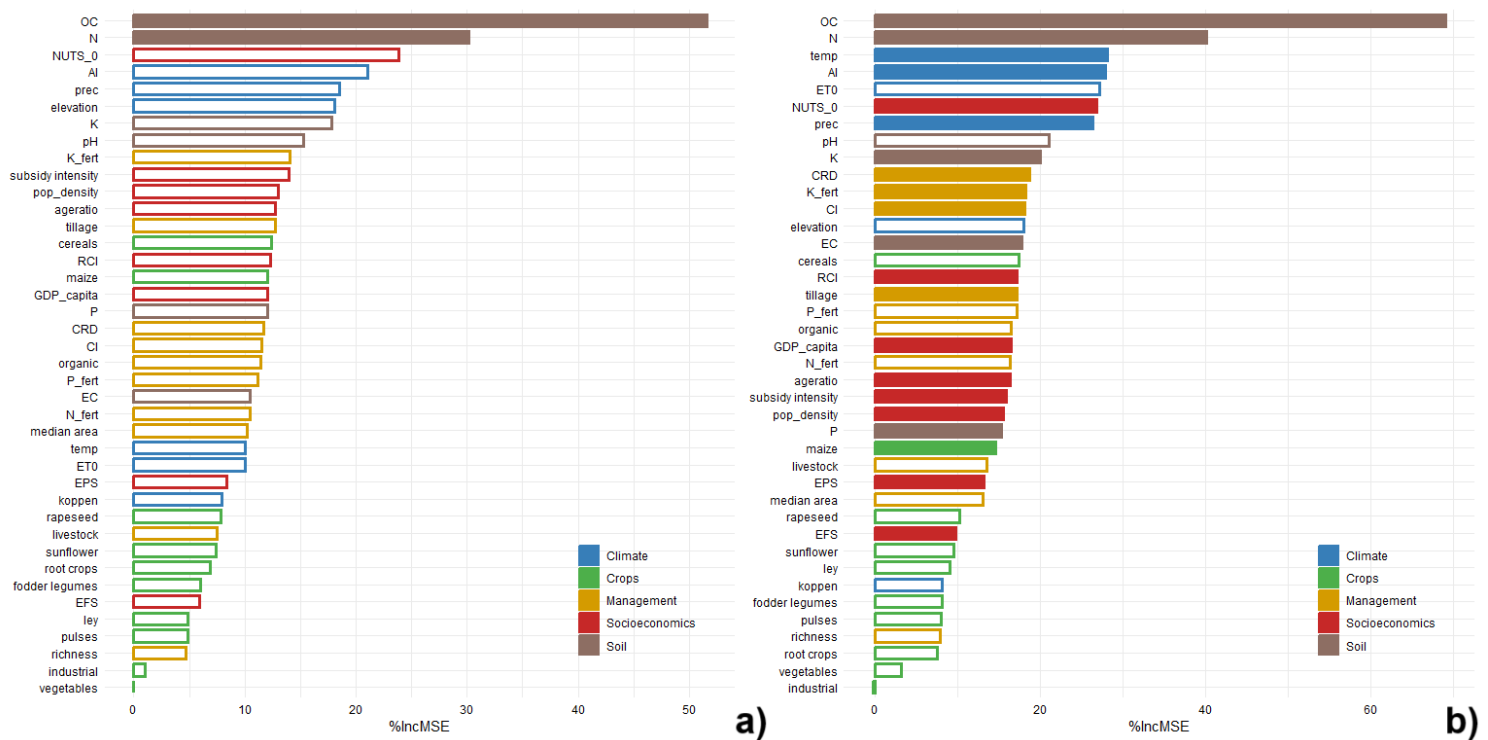


Figure 9: a) Relative importance of variables in the RF model for ΔSOC , b) Relative importance of variables in the RF model for ΔSOC_{rel} . Bars without a fill indicate that this variable did not contribute significantly

After the permutations which generated the relative importances with significances, two new Random Forest models were made for both ΔSOC and ΔSOC_{rel} with (1) the five least important variables removed and (2) all insignificant variables removed, in an attempt to increase model performance. However, model performance could not be notably improved, thus the final models remained unchanged. It is very important to note that due to the low model performance of predicting ΔSOC and ΔSOC_{rel} , and the observed high sensitivity of partial dependences to small changes to the model (underscored by intermediate model results in Appendix I), interpretations from the partial dependence plots from both models can only be made with proper caution and high uncertainty. The partial dependence plots of the 8 most important variables, except NUTS_0,

ranked by %IncMSE of the Random Forest models for ΔSOC and $\Delta\text{SOC}_{\text{rel}}$ are shown in Figures 10 and 11, respectively. As stated in section 3.5, the partial dependence plots do not necessarily show causality, but rather the relationship between the explanatory and the predicted variable within the context of the entire model.

For the model for ΔSOC , only OC and N contribute significantly. There is a strong positive relationship between OC and ΔSOC up to an OC content of approximately 50 g kg^{-1} , after which the line flattens. This indicates that according to the Random Forest model, soils which are high in OC may also have the largest positive SOC sequestration rate. There appears to be a strong negative relationship with N, which is contradictory to the positive relationship that has previously been found in the correlation matrix. Although insignificant, AI and elevation appear to cause lower or negative ΔSOC at their extremes. Either very high or low mean annual precipitation relates to the most positive ΔSOC . pH and K-fertilization show positive and negative relationships with ΔSOC , respectively. K-content has a clear optimum of approximately 500 mg kg^{-1} .

The 8 most important predictors in the model for $\Delta\text{SOC}_{\text{rel}}$ were all significant. The relationships of OC, N, AI, pH and K-content are similar to the model for ΔSOC . The relationship of precipitation with $\Delta\text{SOC}_{\text{rel}}$ differs from that of ΔSOC in the sense that it does not have a high peak at large annual mean precipitation, indicating a largely negative relationship between precipitation and ΔSOC . Mean annual temperature does show an effect similar to that of precipitation in the model for ΔSOC , as it appears that either low or high temperatures relate to the highest $\Delta\text{SOC}_{\text{rel}}$. The relationship between mean annual potential evapotranspiration and $\Delta\text{SOC}_{\text{rel}}$ is largely positive, with the line flattening at an ET_0 of approximately 1200 mm y^{-1} . The partial dependence plot of CRD for the model for $\Delta\text{SOC}_{\text{rel}}$ is shown in Figure 12. The relationship of CRD to $\Delta\text{SOC}_{\text{rel}}$ appears to be such that either a very low or a very high CRD relates to the highest $\Delta\text{SOC}_{\text{rel}}$.

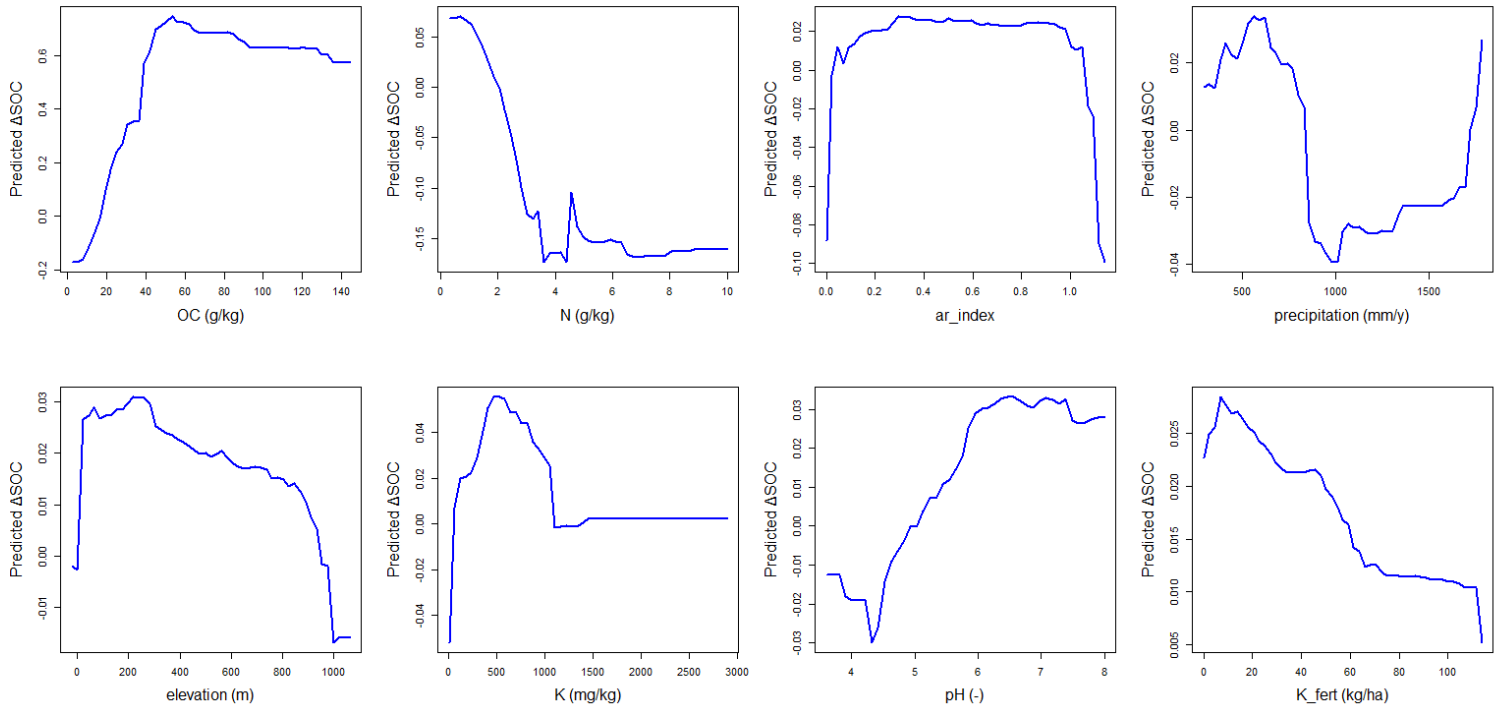


Figure 10: Partial dependence plots of the 8 most important variables in the RF model for ΔSOC

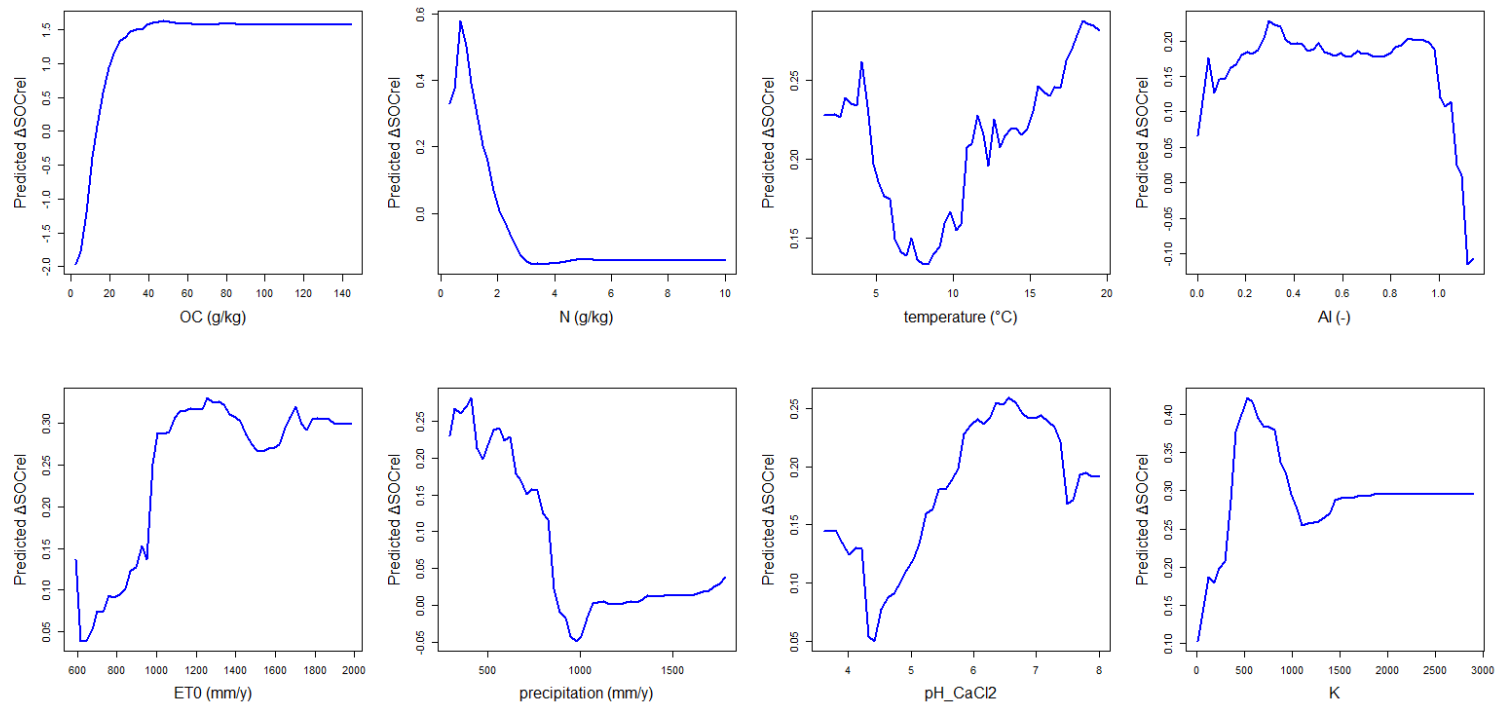


Figure 12: Partial dependence plots of the 8 most important variables in the RF model for ΔSOC_{rel}

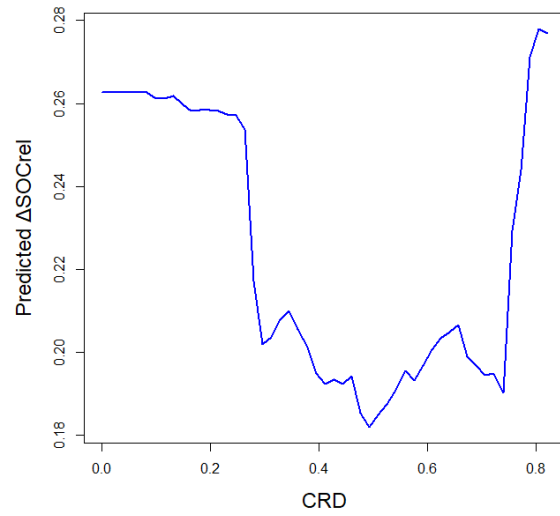


Figure 11: Partial dependence plot of CRD in the RF model for ΔSOC_{rel}

4. Discussion

4.1 Variability of CRD across Europe

According to the data derived from the FADN, CRD is currently highest in central Europe, particularly in eastern Germany, the Czech Republic and Slovakia. CRD is the lowest in the Alps, along the Mediterranean coast, Ireland and northern Scandinavia. This does not align with the hypothesis that the highest diversities would be in regions dominated by low-intensity farms. Comparing to a study by Debonne et al., 2022, in which farm intensities on a NUTS2-scale were determined, the regions with the highest CRD found in this study are dominated by medium-intensity farms, and the regions with the lowest CRD are dominated by low-intensity farms. The results also contradict Cutforth et al., 2001, who found that diversity is generally lowest on highly productive and relatively flat land. Similarly, Merlos & Hijmans, 2020 found that CRD is lowest for common crops, which are usually grown on extensive and highly productive farms. This would suggest that larger farms generally engage in less diverse rotation systems. It is however important to consider that levels of CRD depend on the applied scale (Machefer et al., 2024; Merlos & Hijmans, 2020). As CRD is calculated in this study at the farm level and subsequently averaged for NUTS2 regions, with favoring bigger farms by using weights depending on farm size, it does not necessarily reflect the crop diversity at the landscape scale. At the same time, if the aggregations would not have been based on farm-level data, the levels of diversity would likely have been higher than farm-level diversity, as on a large-scale, the diversity between farms is often larger than the diversity within them (Merlos & Hijmans, 2020). As the scales used by Cutforth et al., 2001 and Merlos & Hijmans, 2020 closely resemble farm-level diversities, comparison to these studies is appropriate. The same holds true for Ricciardi et al., 2021 found based on a meta-analysis that overall, smaller farms have higher crop diversities, as subsistence farmers cultivate a greater diversity of crops to meet nutritional needs. However, Ricciardi et al., 2021 also acknowledges that larger farms have more capacity to diversify their cropping systems. Apart from maintaining soil and plant health, this can lead to improved income stability, as economic risks are mitigated (Parré et al., 2024; Ricciardi et al., 2021). This may explain the strong positive relationship between median arable area (farm size) and CRD in Europe that has been found in this study. Taking into account the findings of a small but generally increasing trend in CRD across Europe, this may suggest that the general megatrend of increasing productivism (Debonne et al., 2022) is outpaced by the efforts of farmers to mitigate economic risks by diversifying their crops. At the same time, the positive correlation between farmer age ratio and the trend in CRD does pose a potential risk to the still currently increasing trend of higher diversities across Europe, as this is possibly indicates that CRD is decreasing among young farmers.

Although not fully aligning with the distribution of organic farms across Europe mapped by (Debonne et al., 2022), the Random Forest analysis for predicting and explaining the variability of CRD has shown a strong negative relationship with the fraction under organic cultivation. This is not in agreement with Barbieri et al., 2017, who found that organic systems usually have greater CRD, as due to the limits of fertilizer use in organic systems, organic farmers are forced to adopt alternative methods for pest control and nutrient management. However, Chongtham et al., 2017 found that the main factor affecting the choice of organic farmers which crops to grow, is market price and demand. As organic farming can entail high production costs with relatively low yields, organic farmers may specialize in only one or a few crops, to maintain economic viability.

Therefore, organic farmers may be more specialized in one crop when compared to conventional farmers. Another plausible explanation for why a negative correlation between CRD and organic farming is found, is that the prevalence of organic farms tends to be larger in marginal regions, e.g. mountainous areas or unfavorable climates (Helfenstein et al., 2024; Willer et al., 2024), where CRD is generally lower because of arable farming being less suitable. Examining this more properly would require comparing CRD of organic versus inorganic farms in the same region, which is not possible with the available aggregated data.

The climatic optima for the highest CRD that resulted from the Random Forest models align with the lowest CRDs being found in regions with the largest climatic extremes. This indicates that climate is a factor that can limit the level of CRD that can be practiced. In regions with challenging or extreme climates, the range of crops which can be cultivated is more narrow (Golla, 2021), e.g. due to high levels of soil degradation (Melesse, 2007) and an unfavorable water balance. Also, in dry climates, extended periods of fallow or the cultivation of forage crops for livestock may be necessary (Gorooei et al., 2023), which limits the number of different crops which can be included in rotation systems. The length of the growing season also impacts to which extent CR systems can be diversified, because of greater flexibility and time for sequential cropping.

Subsidy intensity appears to be the most important socioeconomic variable incorporated in this research for explaining variability in CRD. The partial dependence of subsidy intensity on the Random Forest model shows that CRD increases strongly with increasing subsidy intensity, up to a certain point after which it slightly decreases and stabilizes. Without, or with low subsidies, farmers may not have the economic ability to adapt more diverse crop rotation systems, as changing or diversification of crop rotation systems usually requires economic investments (Santos & Gomes, 2022). These investments can include the acquisition of new seeds, modifying farming equipment and accessing new markets. When subsidies increase, farmers may be more incentivized to diversify their cropping systems. However, high subsidy levels may also be associated with stricter regulations or conditions that push farmers toward specific cropping patterns, reducing flexibility in farmers' choices between different systems (Shrestha et al., 2020; Theriault & Smale, 2021).

The results in section 3.4 have shown discrepancies between crops which are found to be most commonly cultivated in diverse rotational systems versus crops which are increasingly grown in diverse systems. Crops which are found to be most commonly in diverse systems are pulses, rapeseed, root crops and sunflowers, whereas positive trends are mainly found for areas with a high share of pulses and cereals in rotation, and negative trends are found mainly for maize rapeseed and root crops. Ballot et al., 2023 identified the main crop sequence patterns across Europe, and shows that rotational schemes of these four crop types in combination with cereals are found where in general, high CRDs were found as well. The reason that cereals did not correlate with higher CRD is because cereals are also commonly grown in monocultures. However, the significant positive trend in CRD in areas with a high share of cereals grown may indicate that rotations with cereals are getting increasingly diversified, or that cereals are less commonly grown in monocultures.

4.2 Main factors driving SOC changes across Europe

Although the Random Forest models for ΔSOC and $\Delta\text{SOC}_{\text{rel}}$ had a similar model performance and similar variable importances, the model for $\Delta\text{SOC}_{\text{rel}}$ had much more significantly contributing variables than the model for ΔSOC . This is likely explained by the smaller variance of $\Delta\text{SOC}_{\text{rel}}$ values due to the stricter limits for points to be removed for $\Delta\text{SOC}_{\text{rel}}$ than for ΔSOC . Due to the $\Delta\text{SOC}_{\text{rel}}$ having a smaller spread, it means there is less overall variation in the data that is unrelated to predictor variables. This can make it easier for the permutation process to detect significant relationships, because noise has been reduced.

Both absolute and relative changes in SOC were found to have strongly local variations, indicating that large-scale differences are insignificant. De Rosa et al., 2024 found that humid and cold climates better facilitate SOC stock growth, which potentially explains why climate was found to be an important predictor of changes in SOC in the RF models. However, the poor model performance implies that interpreting partial dependences of isolated model parameters should be done with caution. The RF models showed a strong and significant feature importance for OC and N, indicating that initial OC and N content are important variables for predicting SOC stock changes. The findings that soil and climate factors are most important in regulating SOC stock changes aligns with research by Edlinger et al., 2023.

Variables excluded from the RF models were soil texture (sand, silt and clay fractions) and Fe/Al-oxalate content (due to a high number of NAs). It is recommended that future research includes these variables, as these influence SOC stabilization and thus may be important in predicting SOC stock changes (Ichinose et al., 2025; Van De Vreken et al., 2016). Possibly, models could be made with a subset of the data, to overcome the issue of a high number of NAs.

Previous research has shown that grassland has a bigger potential of carbon sequestration than cropland (Conant et al., 2001; De Rosa et al., 2024), substantiating the hypothesis that points with a larger share of ley involved in rotation also show relatively larger increases in SOC stocks. However, no significant correlation between ley and changes in SOC stocks was found in this research. Previous research also found that incorporating (fodder) legumes in rotation contributes to increasing SOC (Ouda et al., 2018; Yang et al., 2024). Although a significant positive relationship with ΔSOC and $\Delta\text{SOC}_{\text{rel}}$ was found, it is only very small. Sunflower showed the largest positive correlation with ΔSOC and $\Delta\text{SOC}_{\text{rel}}$, in disagreement with (Bowman et al., 2000), who found that cultivating sunflower leads to depletion of SOC due to the requirement of tillage. Simultaneously, sunflowers are relatively high residue producing crops, because the residues are neither used as feed for livestock nor suitable for use as fuel due to a low energy value per unit mass (Babu et al., 2014). Given that crop residue retention is one of the most promising methods for enhancing agricultural SOC stocks (Fu et al., 2021; Karlen et al., 2019), this might explain why sunflowers were found to enhance SOC levels in this study.

4.3 Effect of CRD on changes in SOC stocks

Although previous research has found from small-scale field experiments that increasing CRD is beneficial to soil health and increasing SOC stocks (Chahal et al., 2021; Maiga et al., 2019; Schmer et al., 2020), this effect was not found in this study. Directly plotting SOC changes against changes in CRD lead to a low and insignificant correlation, although it should be taken into account that the random effect of variability between different soil types and climates is not accounted for. Even though CRD contributed significantly as an explanatory variable in the Random Forest model for $\Delta\text{SOC}_{\text{rel}}$, the poor model performance did not allow for interpreting relationships with a high level of certainty. The partial dependence of CRD in the model for $\Delta\text{SOC}_{\text{rel}}$ showed the exact opposite relationship than what was hypothesized, namely that $\Delta\text{SOC}_{\text{rel}}$ is highest for either very low or very high CRD. The disagreement with literature, low model performance and considerably small variable importance in the model makes this result an insufficient basis to draw any conclusions.

No evidence to support the hypothesis that increasing CRD would, until a certain level, increase SOC stocks on a European level could therefore be found. There is a high likelihood that the reason for the lack of a relationship being found is due to the quality of the OC measurements in the LUCAS database. The high variability between years, together with the large potential measurement errors compared to the expected changes, likely caused a large amount of noise that the Random Forest models were unable to effectively distinguish from true patterns in the data.

4.4 Overcoming limitations in using LUCAS data for SOC analyses

Using the 3-step matching technique for adding FADN farm management data to the LUCAS database worked successfully in almost all cases. For LUCAS points that could not be allocated any farm management data, this meant that not enough data from farms cultivating the crop specified in the LUCAS data within the same country was available. Given that 99.4% of LUCAS points could be allocated FADN data within the first data matching step, and under the assumption that farm management indicators are rather consistent for the same crop, in the same altitude class, in the same NUTS2-region, it is expected that this is a reliable method to combine farm management data from the FADN to the LUCAS database. However, in order to fully assess the accuracy, it may be necessary to make a comparison with management observations from a variety of farms on which LUCAS samples are taken. Direct farm-level data (in contrast to aggregations) from the FADN database cannot be used for this, since information on the exact locations of the farms is not available due to privacy-sensitivity.

Apart from the assumptions that have to be made in order to match FADN data to LUCAS, another factor that limits the usability is the data quality of LUCAS itself. The main principle of LUCAS is to build a harmonized large-scale dataset in which measurements are done through identical methods, even though they are carried out by a large number of different surveyors. However, an extensive LUCAS data evaluation carried out by Hiederer, 2020 showed that although most parameters are generally sufficiently reliable for large-scale analysis, discrepancies are found between e.g. the limits of detection of the laboratory analysis and file formats in which data was reported. Hiederer, 2020 particularly casts a doubt about the reliability of the 2012 survey which was carried out in Romania and Bulgaria. It is also underscored that OC data has remarkable extremes on both ends (i.e. it is heavy-tailed), and large variability between repeated samples. This

was also observed in this research, and indicates that there is considerable variance in changes of SOC stocks.

In order to account for unrealistic changes, the removal of points with a ΔSOC or $\Delta\text{SOC}_{\text{rel}}$ of -3 to 3 g C kg⁻¹ y⁻¹ or beyond -50 – 100% across the entire time period, respectively, likely decreased the amount of noise in the dataset. However, these limits may still have been too broad. Reijneveld et al., 2009 quantified SOC stocks and changes in the Netherlands over 20 years, and found that SOC stocks in arable land changed approximately 0.08 g kg⁻¹ y⁻¹ with a standard error of 0.02 g kg⁻¹ y⁻¹. In heavy peat soils, which are prone to high levels of SOC degradation, they hardly found any decreases of more than -2 g kg⁻¹ y⁻¹. Important to note is that these peat soils were used as permanent grassland. The findings of Reijneveld et al., 2009 underscore that yearly changes in SOC are small, especially compared to the measurement error. This increases the measurement uncertainty (De Rosa et al., 2024). Beillouin et al., 2023 found after a global meta-analysis that exclusively agricultural land management can cause relative changes of SOC stocks between approximately -25 – 100% (after removing an extreme outlier). More extreme changes are therefore suggested to always be at least in part due to extreme changes such soil transportation or land use change, which are factors which the Random Forest models in this study do not account for.

To further reduce noise and increase the reliability of the data, future research may need to consider including more strict limitations on points to be included in the analysis. To limit the risk of exclusion of reliable points that coincidentally happen to be outliers, and to limit the risk of the dataset being severely thinned out, comparisons could also be made to e.g. bulk density (Iheshiulo et al., 2023; Ouda et al., 2018) and nitrogen content (Bi et al., 2023; Manzoni & Cotrufo, 2024) to check for unrealistic OC values, as these variables are strongly related to OC. This may also be applied to the analysis of SOC stocks in general, where changes in bulk density and N can potentially function as auxiliary data to more accurately estimate changes in SOC stocks.

It is however is not only sampling and laboratory analysis that create a level of uncertainty, but also in-field observations. For example, to ensure that only cropland was included in the analysis, all LUCAS points with a permanent grassland land cover class were removed. Yet, it is plausible that temporary grasslands (ley) may have occasionally have been misclassified as permanent grassland, as grasslands being in place for a duration of 5 years still must be classified as ley (Nevens & Reheul, 2003; Upcott et al., 2023), while the duration the grassland has been in place can be difficult to estimate in-field. Remote sensing may be a usable tool in assessing the duration grasslands have been in place (Andreatta et al., 2022; Yin et al., 2018), which may be recommendable for deciphering temporary from permanent grasslands for future LUCAS surveys, in contrast to in-field observations.

5. Conclusion

In this study, a large-scale analysis on the spatial and temporal variability of CRD in Europe was carried out, and the effect of increasing CRD on SOC stocks was investigated. Data from the LUCAS and FADN databases was successfully combined, with 98.4% of LUCAS datapoints being allocated farm management data. From the FADN data, it was found that current levels of CRD in Europe are highest in Central Europe and lowest in the Alps and along the Mediterranean Coast. Levels of CRD have been largely increasing between 2009-2020, showing an increasing trend across Europe. The increasing trend suggests that the general megatrend of increasing intensification and specialization is outpaced by farmers' efforts to mitigate economic risks and improve sustainability by diversifying their cropping systems. The Random Forest model for explaining variability in CRD had a high model performance, with the most important predictors being the share of arable area under organic farming, farm size and various climatic factors. In contrast, the Random Forest models for explaining SOC stock changes had a low model performance, likely caused by high variability and noise within the OC measurements in LUCAS. Soil OC and N content were found to be most important significant predictors, suggesting a high likelihood that the capacity to increase SOC stocks is largely dependent on the initial OC and N content. The low model performances and lack of direct correlations between ΔCRD and ΔSOC that could be established means that the effects of varying levels of species diversity in rotational systems on SOC stocks still remains unclear. In future research, OC data in the LUCAS database should be further refined in order to reduce noise and uncertainty. This will further facilitate clearing up what still remains unknown about the effects of CRD on SOC stocks, so that concrete policy and management recommendations can be made to help secure healthy soils and sufficient food production, even in a changing climate.

References

- Andreatta, D., Gianelle, D., Scotton, M., Vescovo, L., & Dalponte, M. (2022). Detection of grassland mowing frequency using time series of vegetation indices from Sentinel-2 imagery. *GIScience & Remote Sensing*, 59(1), 481–500. <https://doi.org/10.1080/15481603.2022.2036055>
- Babu, S., Rana, D. S., Yadav, G. S., Singh, R., & Yadav, S. K. (2014). A Review on Recycling of Sunflower Residue for Sustaining Soil Health. *International Journal of Agronomy*, 2014(1), 601049. <https://doi.org/10.1155/2014/601049>
- Baldwin-Kordick, R., De, M., Lopez, M. D., Liebman, M., Lauter, N., Marino, J., & McDaniel, M. D. (2022). Comprehensive impacts of diversified cropping on soil health and sustainability. *Agroecology and Sustainable Food Systems*, 46(3), 331–363. <https://doi.org/10.1080/21683565.2021.2019167>
- Ballot, R., Guilpart, N., & Jeuffroy, M.-H. (2023). The first map of crop sequence types in Europe over 2012–2018. *Earth System Science Data*, 15(12), 5651–5666. <https://doi.org/10.5194/essd-15-5651-2023>
- Barbieri, P., Pellerin, S., & Nesme, T. (2017). Comparing crop rotations between organic and conventional farming. *Scientific Reports*, 7(1), 13761. <https://doi.org/10.1038/s41598-017-14271-6>
- Beck, H. E., McVicar, T. R., Vergopolan, N., Berg, A., Lutsko, N. J., Dufour, A., Zeng, Z., Jiang, X., van Dijk, A. I. J. M., & Miralles, D. G. (2023). High-resolution (1 km) Köppen-Geiger maps for 1901–2099 based on constrained CMIP6 projections. *Scientific Data*, 10(1), 724. <https://doi.org/10.1038/s41597-023-02549-6>
- Beillouin, D., Corbeels, M., Demenois, J., Berre, D., Boyer, A., Fallot, A., Feder, F., & Cardinael, R. (2023). A global meta-analysis of soil organic carbon in the Anthropocene. *Nature Communications*, 14(1), 3700. <https://doi.org/10.1038/s41467-023-39338-z>
- Bi, X., Chu, H., Fu, M., Xu, D., Zhao, W., Zhong, Y., Wang, M., Li, K., & Zhang, Y. (2023). Distribution characteristics of organic carbon (nitrogen) content, cation exchange capacity, and specific

- surface area in different soil particle sizes. *Scientific Reports*, 13(1), 12242.
<https://doi.org/10.1038/s41598-023-38646-0>
- Bodner, G., Mentler, A., & Keiblinger, K. (2021). Plant Roots for Sustainable Soil Structure Management in Cropping Systems. In *The Root Systems in Sustainable Agricultural Intensification* (pp. 45–90). John Wiley & Sons, Ltd.
<https://doi.org/10.1002/9781119525417.ch3>
- Boincean, B., & Dent, D. (2019). Crop Rotation. In B. Boincean & D. Dent (Eds.), *Farming the Black Earth: Sustainable and Climate-Smart Management of Chernozem Soils* (pp. 89–124). Springer International Publishing. https://doi.org/10.1007/978-3-030-22533-9_5
- Bowles, T. M., Mooshammer, M., Socolar, Y., Calderón, F., Cavigelli, M. A., Culman, S. W., Deen, W., Drury, C. F., Garcia, A. G. Y., Gaudin, A. C. M., Harkcom, W. S., Lehman, R. M., Osborne, S. L., Robertson, G. P., Salerno, J., Schmer, M. R., Strock, J., & Grandy, A. S. (2020). Long-Term Evidence Shows that Crop-Rotation Diversification Increases Agricultural Resilience to Adverse Growing Conditions in North America. *One Earth*, 2(3), 284–293.
<https://doi.org/10.1016/j.oneear.2020.02.007>
- Bowman, R. A., Nielsen, D. C., Vigil, M. F., & Aiken, R. M. (2000). EFFECTS OF SUNFLOWER ON SOIL QUALITY INDICATORS AND SUBSEQUENT WHEAT YIELD. *Soil Science*, 165(6), 516.
- Bybee-Finley, K. A., Muller, K., White, K. E., Cavigelli, M. A., Han, E., Schomberg, H. H., Snapp, S., Viens, F., Correndo, A. A., Deiss, L., Fonteyne, S., Garcia, A. G. y, Gaudin, A. C. M., Hooker, D. C., Janovicek, K., Jin, V., Johnson, G., Karsten, H., Liebman, M., ... Bowles, T. M. (2024). Rotational complexity increases cropping system output under poorer growing conditions. *One Earth*, 0(0). <https://doi.org/10.1016/j.oneear.2024.07.008>
- Chahal, I., Hooker, D. C., Deen, B., Janovicek, K., & Van Eerd, L. L. (2021). Long-term effects of crop rotation, tillage, and fertilizer nitrogen on soil health indicators and crop productivity in a temperate climate. *Soil and Tillage Research*, 213, 105121.
<https://doi.org/10.1016/j.still.2021.105121>

- Chatterjee, A., Cooper, K., Klaustermeier, A., Awale, R., & Cihacek, L. J. (2016). Does Crop Species Diversity Influence Soil Carbon and Nitrogen Pools? *Agronomy Journal*, 108(1), 427–432. <https://doi.org/10.2134/agronj2015.0316>
- Chongtham, I. R., Bergkvist, G., Watson, C. A., Sandström, E., Bengtsson, J., & Öborn, I. (2017). Factors influencing crop rotation strategies on organic farms with different time periods since conversion to organic production. *Biological Agriculture & Horticulture*, 33(1), 14–27. <https://doi.org/10.1080/01448765.2016.1174884>
- Conant, R. T., Paustian, K., & Elliott, E. T. (2001). Grassland Management and Conversion into Grassland: Effects on Soil Carbon. *Ecological Applications*, 11(2), 343–355. [https://doi.org/10.1890/1051-0761\(2001\)011\[0343:GMACIG\]2.0.CO;2](https://doi.org/10.1890/1051-0761(2001)011[0343:GMACIG]2.0.CO;2)
- Costa, H., Almeida, D., Vala, F., Marcelino, F., & Caetano, M. (2018). Land Cover Mapping from Remotely Sensed and Auxiliary Data for Harmonized Official Statistics. *ISPRS International Journal of Geo-Information*, 7(4), Article 4. <https://doi.org/10.3390/ijgi7040157>
- Cutforth, L. B., Francis, C. A., Lynne, G. D., Mortensen, D. A., & Eskridge, K. M. (2001). Factors affecting farmers' crop diversity decisions: An integrated approach. *American Journal of Alternative Agriculture*, 16(4), 168–176. <https://doi.org/10.1017/S0889189300009164>
- De Rosa, D., Ballabio, C., Lugato, E., Fasiolo, M., Jones, A., & Panagos, P. (2024). Soil organic carbon stocks in European croplands and grasslands: How much have we lost in the past decade? *Global Change Biology*, 30(1), e16992. <https://doi.org/10.1111/gcb.16992>
- Debonne, N., Bürgi, M., Diogo, V., Helfenstein, J., Herzog, F., Levers, C., Mohr, F., Swart, R., & Verburg, P. (2022). The geography of megatrends affecting European agriculture. *Global Environmental Change*, 75, 102551. <https://doi.org/10.1016/j.gloenvcha.2022.102551>
- Degani, E., Leigh, S. G., Barber, H. M., Jones, H. E., Lukac, M., Sutton, P., & Potts, S. G. (2019). Crop rotations in a climate change scenario: Short-term effects of crop diversity on resilience and ecosystem service provision under drought. *Agriculture, Ecosystems & Environment*, 285, 106625. <https://doi.org/10.1016/j.agee.2019.106625>

- Edlinger, A., Garland, G., Banerjee, S., Degruner, F., García-Palacios, P., Herzog, C., Pescador, D. S., Romdhane, S., Ryo, M., Saghai, A., Hallin, S., Maestre, F. T., Philippot, L., Rillig, M. C., & van der Heijden, M. G. A. (2023). The impact of agricultural management on soil aggregation and carbon storage is regulated by climatic thresholds across a 3000 km European gradient. *Global Change Biology*, 29(11), 3177–3192. <https://doi.org/10.1111/gcb.16677>
- Engel, R. E., Miller, P. R., McConkey, B. G., & Wallander, R. (2017). Soil Organic Carbon Changes to Increasing Cropping Intensity and No-Till in a Semiarid Climate. *Soil Science Society of America Journal*, 81(2), 404–413. <https://doi.org/10.2136/sssaj2016.06.0194>
- European Commission. (2019). *European Regional Competitiveness Index*. https://ec.europa.eu/regional_policy/information-sources/maps/regional-competitiveness/2019_en
- European Commission. (2024, March 13). *FADN - European Commission*. https://agriculture.ec.europa.eu/data-and-analysis/farm-structures-and-economics/fadn_en
- Eurostat. (2020). *Agri-environmental indicator—Tillage practices*. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Agri-environmental_indicator_-_tillage_practices
- Eurostat. (2022a). *LUCAS 2022 (Land Use / Cover Area Frame Survey)—Technical reference document C3 Classification (Land cover & Land use)*. <https://ec.europa.eu/eurostat/documents/205002/13686460/C3-LUCAS-2022.pdf>
- Eurostat. (2022b). *Statistical regions in the European Union and partner countries—NUTS and statistical regions 2021*. <https://ec.europa.eu/eurostat/documents/3859598/15193590/KS-GQ-22-010-EN-N.pdf>
- Eurostat. (2024). *Glossary: Livestock unit (LSU)*. [https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Livestock_unit_\(LSU\)](https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Livestock_unit_(LSU))

- FAO. (2019). *TAPE Tool for Agroecology Performance Evaluation 2019 – Process of development and guidelines for application. Test version.*
<https://openknowledge.fao.org/server/api/core/bitstreams/8ad4bb1b-c06d-4260-835e-564698493149/content>
- Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37(12), 4302–4315.
<https://doi.org/10.1002/joc.5086>
- Fowler, A. F., Basso, B., Millar, N., & Brinton, W. F. (2023). A simple soil mass correction for a more accurate determination of soil carbon stock changes. *Scientific Reports*, 13(1), 2242.
<https://doi.org/10.1038/s41598-023-29289-2>
- Fu, B., Chen, L., Huang, H., Qu, P., & Wei, Z. (2021). Impacts of crop residues on soil health: A review. *Environmental Pollutants and Bioavailability*, 33(1), 164–173.
<https://doi.org/10.1080/26395940.2021.1948354>
- Golla, B. (2021). Agricultural production system in arid and semi-arid regions. *International Journal of Agricultural Science and Food Technology*. <https://doi.org/10.17352/2455-815X.000113>
- Gorooei, A., Gaiser, T., Ayneband, A., Rahn timer, A., & Kamali, B. (2023). The Effect of Farming Management and Crop Rotation Systems on Chlorophyll Content, Dry Matter Translocation, and Grain Quantity and Quality of Wheat (*Triticum aestivum* L.) Grown in a Semi-Arid Region of Iran. *Agronomy*, 13(4), Article 4. <https://doi.org/10.3390/agronomy13041007>
- Greenwell, B., M. (2017). pdp: An R Package for Constructing Partial Dependence Plots. *The R Journal*, 9(1), 421. <https://doi.org/10.32614/RJ-2017-016>
- Guiasu, R. C., & Guiasu, S. (2012). The Weighted Gini-Simpson Index: Revitalizing an Old Index of Biodiversity. *International Journal of Ecology*, 2012(1), 478728.
<https://doi.org/10.1155/2012/478728>

- Haddaway, N. R., Hedlund, K., Jackson, L. E., Kätterer, T., Lugato, E., Thomsen, I. K., Jørgensen, H. B., & Isberg, P.-E. (2017). How does tillage intensity affect soil organic carbon? A systematic review. *Environmental Evidence*, 6(1), 30. <https://doi.org/10.1186/s13750-017-0108-9>
- Helfenstein, J., Hepner, S., Kreuzer, A., Achermann, G., Williams, T., Bürgi, M., Debonne, N., Dimopoulos, T., Diogo, V., Fjellstad, W., Garcia-Martin, M., Hernik, J., Kizos, T., Lausch, A., Levers, C., Liira, J., Mohr, F., Moreno, G., Pazur, R., ... Herzog, F. (2024). Divergent agricultural development pathways across farm and landscape scales in Europe: Implications for sustainability and farmer satisfaction. *Global Environmental Change*, 86, 102855. <https://doi.org/10.1016/j.gloenvcha.2024.102855>
- Heuscher, S. A., Brandt, C. C., & Jardine, P. M. (2005). Using Soil Physical and Chemical Properties to Estimate Bulk Density. *Soil Science Society of America Journal*, 69(1), 51–56. <https://doi.org/10.2136/sssaj2005.0051a>
- Hiederer, R. (2020). *Data evaluation of LUCAS soil survey laboratory data: Survey 2009, 2012 and 2015*. European Commission. <https://data.europa.eu/doi/10.2760/791714>
- Hollister, J. W., Shah, T., Nowosad, J., Robitaille, A. L., Beck, M. W., & Johnson, M. (2023). *elevatr: Access elevation data from various APIs (Version 0.99.0)* [Dataset]. Zenodo. <https://doi.org/10.5281/zenodo.8335450>
- Hufnagel, J., Reckling, M., & Ewert, F. (2020). Diverse approaches to crop diversification in agricultural research. A review. *Agronomy for Sustainable Development*, 40(2), 14. <https://doi.org/10.1007/s13593-020-00617-4>
- Ichinose, Y., Matsui, K., Fukumasu, J., Matsuura, S., Takata, Y., & Wagai, R. (2025). How do reactive aluminum and iron phases control soil organic carbon and phosphate adsorption capacity in agricultural topsoils across Japan? *Soil Science and Plant Nutrition*, 71(1), 38–52. <https://doi.org/10.1080/00380768.2024.2410322>
- Iheshiulo, E. M.-A., Larney, F. J., Hernandez-Ramirez, G., St. Luce, M., Liu, K., & Chau, H. W. (2023). Do diversified crop rotations influence soil physical health? A meta-analysis. *Soil and Tillage Research*, 233, 105781. <https://doi.org/10.1016/j.still.2023.105781>

- Isbell, F., Adler, P. R., Eisenhauer, N., Fornara, D., Kimmel, K., Kremen, C., Letourneau, D. K., Liebman, M., Polley, H. W., Quijas, S., & Scherer-Lorenzen, M. (2017). Benefits of increasing plant diversity in sustainable agroecosystems. *Journal of Ecology*, 105(4), 871–879. <https://doi.org/10.1111/1365-2745.12789>
- Jägermeyr, J., Müller, C., Ruane, A. C., Elliott, J., Balkovic, J., Castillo, O., Faye, B., Foster, I., Folberth, C., Franke, J. A., Fuchs, K., Guarin, J. R., Heinke, J., Hoogenboom, G., Iizumi, T., Jain, A. K., Kelly, D., Khabarov, N., Lange, S., ... Rosenzweig, C. (2021). Climate impacts on global agriculture emerge earlier in new generation of climate and crop models. *Nature Food*, 2(11), 873–885. <https://doi.org/10.1038/s43016-021-00400-y>
- Karlen, D. L., Schmer, M. R., Kaffka, S., Clay, D. E., Wang, M. Q., Horwath, W. R., Kendall, A., Keller, A., Pieper, B. J., Unnasch, S., Darlington, T., Vocasek, F., & Chute, A. G. (2019). Unraveling Crop Residue Harvest Effects on Soil Organic Carbon. *Agronomy Journal*, 111(1), 93–98. <https://doi.org/10.2134/agronj2018.03.0207>
- Kassambara, A. (2023). *rstatix: Pipe-Friendly Framework for Basic Statistical Tests* (Version 0.7.2) [R]. <https://cran.r-project.org/web/packages/rstatix/index.html>
- Kim, Y. J., & Cribbie, R. A. (2018). ANOVA and the variance homogeneity assumption: Exploring a better gatekeeper. *British Journal of Mathematical and Statistical Psychology*, 71(1), 1–12. <https://doi.org/10.1111/bmsp.12103>
- Kluger, D. M., Owen, A. B., & Lobell, D. B. (2022). Combining randomized field experiments with observational satellite data to assess the benefits of crop rotations on yields. *Environmental Research Letters*, 17(4), 044066. <https://doi.org/10.1088/1748-9326/ac6083>
- Kollas, C., Kersebaum, K. C., Nendel, C., Manevski, K., Müller, C., Palosuo, T., Armas-Herrera, C. M., Beaudoin, N., Bindi, M., Charfeddine, M., Conradt, T., Constantin, J., Eitzinger, J., Ewert, F., Ferrise, R., Gaiser, T., Cortazar-Atauri, I. G. de, Giglio, L., Hlavinka, P., ... Wu, L. (2015). Crop rotation modelling—A European model intercomparison. *European Journal of Agronomy*, 70, 98–111. <https://doi.org/10.1016/j.eja.2015.06.007>

- Kruse, T., Dechezleprêtre, A., Saffar, R., & Robert, L. (2022). Measuring environmental policy stringency in OECD countries: An update of the OECD composite EPS indicator. *OECD Economics Department Working Papers*, 1703. <https://doi.org/10.1787/90ab82e8-en>
- Kumar, M. (2016). Impact of climate change on crop yield and role of model for achieving food security. *Environmental Monitoring and Assessment*, 188(8), 465. <https://doi.org/10.1007/s10661-016-5472-3>
- Kutsch, W. L., Aubinet, M., Buchmann, N., Smith, P., Osborne, B., Eugster, W., Wattenbach, M., Schrumpf, M., Schulze, E. D., Tomelleri, E., Ceschia, E., Bernhofer, C., Béziat, P., Carrara, A., Di Tommasi, P., Grünwald, T., Jones, M., Magliulo, V., Marloie, O., ... Ziegler, W. (2010). The net biome production of full crop rotations in Europe. *Agriculture, Ecosystems & Environment*, 139(3), 336–345. <https://doi.org/10.1016/j.agee.2010.07.016>
- Lal, R. (2004a). Agricultural activities and the global carbon cycle. *Nutrient Cycling in Agroecosystems*, 70(2), 103–116. <https://doi.org/10.1023/B:FRES.0000048480.24274.0f>
- Lal, R. (2004b). Soil carbon sequestration to mitigate climate change. *Geoderma*, 123(1), 1–22. <https://doi.org/10.1016/j.geoderma.2004.01.032>
- Li, G., Yu, C., Shen, P., Hou, Y., Ren, Z., Li, N., Liao, Y., Li, T., & Wen, X. (2024). Crop diversification promotes soil aggregation and carbon accumulation in global agroecosystems: A meta-analysis. *Journal of Environmental Management*, 350, 119661. <https://doi.org/10.1016/j.jenvman.2023.119661>
- Liu, X., Tan, S., Song, X., Wu, X., Zhao, G., Li, S., & Liang, G. (2022). Response of soil organic carbon content to crop rotation and its controls: A global synthesis. *Agriculture, Ecosystems & Environment*, 335, 108017. <https://doi.org/10.1016/j.agee.2022.108017>
- Liu, Y., Chen, S., Yu, Q., Cai, Z., Zhou, Q., Bellingrath-Kimura, S. D., & Wu, W. (2023). Improving digital mapping of soil organic matter in cropland by incorporating crop rotation. *Geoderma*, 438, 116620. <https://doi.org/10.1016/j.geoderma.2023.116620>

- Lugato, E., Panagos, P., Bampa, F., Jones, A., & Montanarella, L. (2014). A new baseline of organic carbon stock in European agricultural soils using a modelling approach. *Global Change Biology*, 20(1), 313–326. <https://doi.org/10.1111/gcb.12292>
- Machefer, M., Zampieri, M., van der Velde, M., Dentener, F., Claverie, M., & d'Andrimont, R. (2024). Earth Observation based multi-scale analysis of crop diversity in the European Union: First insights for agro-environmental policies. *Agriculture, Ecosystems & Environment*, 374, 109143. <https://doi.org/10.1016/j.agee.2024.109143>
- Maiga, A., Alhameid, A., Singh, S., Polat, A., Singh, J., Kumar, S., & Osborne, S. (2019). Responses of soil organic carbon, aggregate stability, carbon and nitrogen fractions to 15 and 24 years of no-till diversified crop rotations. *Soil Research*, 57(2), 149–157. <https://doi.org/10.1071/SR18068>
- Manzoni, S., & Cotrufo, M. F. (2024). Mechanisms of soil organic carbon and nitrogen stabilization in mineral-associated organic matter – insights from modeling in phase space. *Biogeosciences*, 21(18), 4077–4098. <https://doi.org/10.5194/bg-21-4077-2024>
- Marini, L., St-Martin, A., Vico, G., Baldoni, G., Berti, A., Blecharczyk, A., Malecka-Jankowiak, I., Morari, F., Sawinska, Z., & Bommarco, R. (2020). Crop rotations sustain cereal yields under a changing climate. *Environmental Research Letters*, 15(12), 124011. <https://doi.org/10.1088/1748-9326/abc651>
- McDaniel, M. D., Tiemann, L. K., & Grandy, A. S. (2014). Does agricultural crop diversity enhance soil microbial biomass and organic matter dynamics? A meta-analysis. *Ecological Applications*, 24(3), 560–570. <https://doi.org/10.1890/13-0616.1>
- Melesse, T. L. (2007). *Conservation Tillage Systems and Water Productivity—Implications for Smallholder Farmers in Semi-Arid Ethiopia*. Delft University of Technology.
- Merlos, F., & Hijmans, R. J. (2020). The scale dependency of spatial crop species diversity and its relation to temporal diversity. *Proceedings of the National Academy of Sciences*, 117(42), 26176–26182. <https://doi.org/10.1073/pnas.2011702117>

- Mortensen, D. A., & Smith, R. G. (2020). Confronting Barriers to Cropping System Diversification. *Frontiers in Sustainable Food Systems*, 4. <https://doi.org/10.3389/fsufs.2020.564197>
- Nevens, F., & Reheul, D. (2003). Permanent grassland and 3-year leys alternating with 3 years of arable land: 31 years of comparison. *European Journal of Agronomy*, 19(1), 77–90. [https://doi.org/10.1016/S1161-0301\(02\)00021-7](https://doi.org/10.1016/S1161-0301(02)00021-7)
- Novelli, L. E., Caviglia, O. P., & Piñeiro, G. (2017). Increased cropping intensity improves crop residue inputs to the soil and aggregate-associated soil organic carbon stocks. *Soil and Tillage Research*, 165, 128–136. <https://doi.org/10.1016/j.still.2016.08.008>
- Nunes, M. R., Karlen, D. L., Veum, K. S., Moorman, T. B., & Cambardella, C. A. (2020). Biological soil health indicators respond to tillage intensity: A US meta-analysis. *Geoderma*, 369, 114335. <https://doi.org/10.1016/j.geoderma.2020.114335>
- Nunes, M. R., van Es, H. M., Schindelbeck, R., Ristow, A. J., & Ryan, M. (2018). No-till and cropping system diversification improve soil health and crop yield. *Geoderma*, 328, 30–43. <https://doi.org/10.1016/j.geoderma.2018.04.031>
- Olesen, J. E., Trnka, M., Kersebaum, K. C., Skjelvåg, A. O., Seguin, B., Peltonen-Sainio, P., Rossi, F., Kozyra, J., & Micale, F. (2011). Impacts and adaptation of European crop production systems to climate change. *European Journal of Agronomy*, 34(2), 96–112. <https://doi.org/10.1016/j.eja.2010.11.003>
- Omay, A. B., Rice, C. W., Maddux, L. D., & Gordon, W. B. (1997). Changes in Soil Microbial and Chemical Properties under Long-term Crop Rotation and Fertilization. *Soil Science Society of America Journal*, 61(6), 1672–1678. <https://doi.org/10.2136/sssaj1997.03615995006100060019x>
- Orgiazzi, A., Ballabio, C., Panagos, P., Jones, A., & Fernández-Ugalde, O. (2018). LUCAS Soil, the largest expandable soil dataset for Europe: A review. *European Journal of Soil Science*, 69(1), 140–153. <https://doi.org/10.1111/ejss.12499>

- Ouda, S., Zohry, A., & Noreldin, T. (2018). Crop Rotation Maintains Soil Sustainability. In S. Ouda, A. E.-H. Zohry, & T. Noreldin (Eds.), *Crop Rotation: An Approach to Secure Future Food* (pp. 55–76). Springer International Publishing. https://doi.org/10.1007/978-3-030-05351-2_4
- Parré, J. L., Chagas, A. L. S., & Arends-Kuenning, M. P. (2024). The effect of farm size and farmland use on agricultural diversification: A spatial analysis of Brazilian municipalities. *Agricultural and Food Economics*, 12(1), 27. <https://doi.org/10.1186/s40100-024-00323-9>
- Piet, L. (Ed.). (2016). *Recent trends in the distribution of farm sizes in the EU*. <https://doi.org/10.22004/ag.econ.245075>
- Pocock, S. J. (2006). The simplest statistical test: How to check for a difference between treatments. *BMJ*, 332(7552), 1256. <https://doi.org/10.1136/bmj.332.7552.1256>
- Reijneveld, A., Van Wensem, J., & Oenema, O. (2009). Soil organic carbon contents of agricultural land in the Netherlands between 1984 and 2004. *Geoderma*, 152(3–4), 231–238. <https://doi.org/10.1016/j.geoderma.2009.06.007>
- Rezaei, E. E., Webber, H., Asseng, S., Boote, K., Durand, J. L., Ewert, F., Martre, P., & MacCarthy, D. S. (2023). Climate change impacts on crop yields. *Nature Reviews Earth & Environment*, 4(12), 831–846. <https://doi.org/10.1038/s43017-023-00491-0>
- Ricciardi, V., Mehrabi, Z., Wittman, H., James, D., & Ramankutty, N. (2021). Higher yields and more biodiversity on smaller farms. *Nature Sustainability*, 4(7), 651–657. <https://doi.org/10.1038/s41893-021-00699-2>
- Santos, D. F. L., & Gomes, S. M. (2022). Diversification of crops: Assessment of managerial flexibility and economic impact on sugarcane medium-sized farms. *Frontiers in Sustainable Food Systems*, 6. <https://doi.org/10.3389/fsufs.2022.987214>
- Schmer, M. R., Jin, V. L., Wienhold, B. J., Becker, S. M., & Varvel, G. E. (2020). Long-term rotation diversity and nitrogen effects on soil organic carbon and nitrogen stocks. *Agrosystems, Geosciences & Environment*, 3(1), e20055. <https://doi.org/10.1002/agg2.20055>

- Shrestha, P., Karim, R. A., Sieverding, H. L., Archer, D. W., Kumar, S., Nleya, T., Graham, C. J., & Stone, J. J. (2020). Life cycle assessment of wheat production and wheat-based crop rotations. *Journal of Environmental Quality*, 49(6), 1515–1529. <https://doi.org/10.1002/jeq2.20158>
- Smith, M. E., Vico, G., Costa, A., Bowles, T., Gaudin, A. C. M., Hallin, S., Watson, C. A., Alarcón, R., Berti, A., Blecharczyk, A., Calderon, F. J., Culman, S., Deen, W., Drury, C. F., Garcia, A. G. y, García-Díaz, A., Plaza, E. H., Jonczyk, K., Jäck, O., ... Bommarco, R. (2023). Increasing crop rotational diversity can enhance cereal yields. *Communications Earth & Environment*, 4(1), 1–9. <https://doi.org/10.1038/s43247-023-00746-0>
- Stein, S., & Steinmann, H.-H. (2018). Identifying crop rotation practice by the typification of crop sequence patterns for arable farming systems – A case study from Central Europe. *European Journal of Agronomy*, 92, 30–40. <https://doi.org/10.1016/j.eja.2017.09.010>
- Stekhoven, D. J., & Bühlmann, P. (2012). MissForest—Non-parametric missing value imputation for mixed-type data. *Bioinformatics*, 28(1), 112–118. <https://doi.org/10.1093/bioinformatics/btr597>
- Theriault, V., & Smale, M. (2021). The unintended consequences of the fertilizer subsidy program on crop species diversity in Mali. *Food Policy*, 102, 102121. <https://doi.org/10.1016/j.foodpol.2021.102121>
- Tilman, D., Balzer, C., Hill, J., & Befort, B. L. (2011). Global food demand and the sustainable intensification of agriculture. *Proceedings of the National Academy of Sciences*, 108(50), 20260–20264. <https://doi.org/10.1073/pnas.1116437108>
- Tilman, D., Cassman, K. G., Matson, P. A., Naylor, R., & Polasky, S. (2002). Agricultural sustainability and intensive production practices. *Nature*, 418(6898), 671–677. <https://doi.org/10.1038/nature01014>
- Tomczak, E., & Tomczak, M. (2014). The need to report effect size estimates revisited. An overview of some recommended measures of effect size. *Biblioteka Akademii Wychowania Fizycznego w Poznaniu*. <https://www.wbc.poznan.pl/dlibra/publication/413565/edition/325867>

- Tsui, C.-C., Tsai, C.-C., & Chen, Z.-S. (2013). Soil organic carbon stocks in relation to elevation gradients in volcanic ash soils of Taiwan. *Geoderma*, 209–210, 119–127. <https://doi.org/10.1016/j.geoderma.2013.06.013>
- Upcott, E. V., Henrys, P. A., Redhead, J. W., Jarvis, S. G., & Pywell, R. F. (2023). A new approach to characterising and predicting crop rotations using national-scale annual crop maps. *Science of The Total Environment*, 860, 160471. <https://doi.org/10.1016/j.scitotenv.2022.160471>
- Van De Vreken, P., Gobin, A., Baken, S., Van Holm, L., Verhasselt, A., Smolders, E., & Merckx, R. (2016). Crop residue management and oxalate-extractable iron and aluminium explain long-term soil organic carbon sequestration and dynamics. *European Journal of Soil Science*, 67(3), 332–340. <https://doi.org/10.1111/ejss.12343>
- Wei, T., & Simko, V. (2024). *R package “corrplot”: Visualization of a Correlation Matrix. (Version 0.95)*. <https://github.com/taiyun/corrplot>
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org/>
- Willer, H., Trávníček, J., & Schlatter, B. (2024). *The World of Organic Agriculture. Statistics and Emerging Trends 2024*. Research Institute of Organic Agriculture FiBL, Frick, and IFOAM – Organics International, Bonn. <https://www.organic-world.net/yearbook/yearbook-2024.html>
- Williams, A., Birt, H. W. G., Raghavendra, A., & Dennis, P. G. (2023). Cropping System Diversification Influences Soil Microbial Diversity in Subtropical Dryland Farming Systems. *Microbial Ecology*, 85(4), 1473–1484. <https://doi.org/10.1007/s00248-022-02074-w>
- Williams, T. G., Bürgi, M., Debonne, N., Diogo, V., Helfenstein, J., Levers, C., Mohr, F., Stratton, A. E., & Verburg, P. H. (2024). Mapping lock-ins and enabling environments for agri-food sustainability transitions in Europe. *Sustainability Science*, 19(4), 1221–1242. <https://doi.org/10.1007/s11625-024-01480-y>

- Yang, X., Xiong, J., Du, T., Ju, X., Gan, Y., Li, S., Xia, L., Shen, Y., Pacenka, S., Steenhuis, T. S., Siddique, K. H. M., Kang, S., & Butterbach-Bahl, K. (2024). Diversifying crop rotation increases food production, reduces net greenhouse gas emissions and improves soil health. *Nature Communications*, 15(1), 198. <https://doi.org/10.1038/s41467-023-44464-9>
- Yin, H., Prishchepov, A. V., Kuemmerle, T., Bleyhl, B., Buchner, J., & Radeloff, V. C. (2018). Mapping agricultural land abandonment from spatial and temporal segmentation of Landsat time series. *Remote Sensing of Environment*, 210, 12–24. <https://doi.org/10.1016/j.rse.2018.02.050>
- Zheng, F., Liu, X., Ding, W., Song, X., Li, S., & Wu, X. (2023). Positive effects of crop rotation on soil aggregation and associated organic carbon are mainly controlled by climate and initial soil carbon content: A meta-analysis. *Agriculture, Ecosystems & Environment*, 355, 108600. <https://doi.org/10.1016/j.agee.2023.108600>
- Zomer, R. J., Xu, J., & Trabucco, A. (2022). Version 3 of the Global Aridity Index and Potential Evapotranspiration Database. *Scientific Data*, 9(1), 409. <https://doi.org/10.1038/s41597-022-01493-1>

Appendix I: Intermediate Model Results

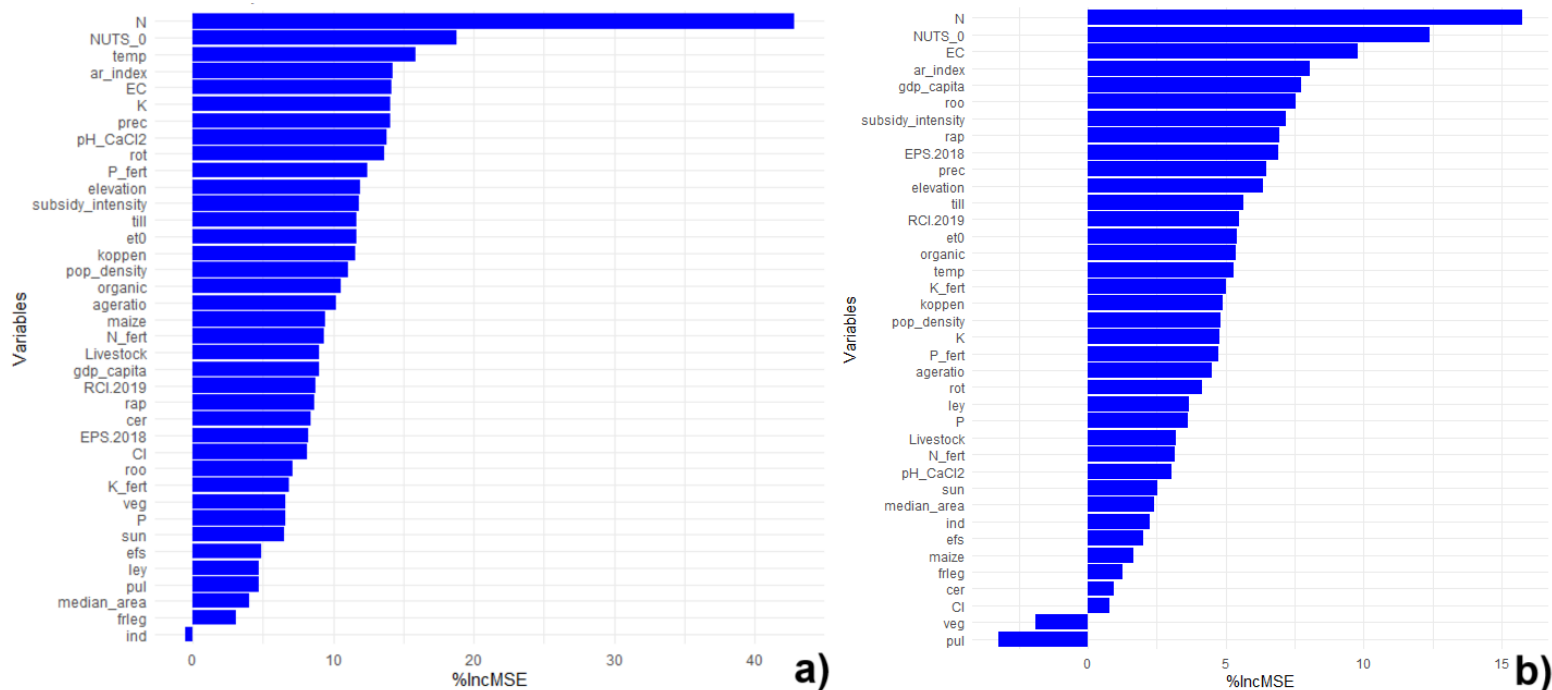


Figure 13: a) Relative importance of variables in an intermediate RF model for ΔSOC , b) Relative importance of variables in an intermediate RF model for ΔSOC_{rel} .

Figures 13a and 13b show the relative importance plots of intermediate results of the Random Forest models for ΔSOC and ΔSOC_{rel} , respectively. It should be noted that the models and their input data requirements have changed numerous times during the research, and that these models were the last intermediate versions before they were finalized. The plots did not yet have the improved variable labelling and categorical coloring. Also, these models were not permuted due to time limitations, so variable significance was not yet assessed. The main differences of these models compared to the final models, is that they did not have points with a ΔSOC_{rel} beyond -50 and 100 % across the entire time period removed yet, and also soil OC content was not included as an explanatory variable. The R^2 of the models was 0.15 and 0.07, respectively. RMSE and ME were not computed for these models.

When comparing these models to the final models used in this study, some clear and large changes can be observed. Where e.g. N, NUTS_0 and aridity index remained some of the most important variables in the final models, variables such as EC and share of root crops (for ΔSOC_{rel}) have much less variable importance in the final models. Overall, the variable importances of the highlighted intermediate model for ΔSOC_{rel} changed most strongly by far, which can be explained by only points with extreme ΔSOC_{rel} values being removed in the final models. Figures 13a and 13b show the partial dependence plots of the 12 most important variables of the intermediate Random Forest models for ΔSOC and ΔSOC_{rel} , respectively. It can be clearly noted that where some variables' partial dependences are similar to the final models (e.g. precipitation and K), some partial dependences are very different or even inverse (e.g. N, pH and rot (CRD)). Apart from some variables' partial dependence differing strongly to those of the final models, there are also major differences between these models, for example for elevation, aridity index and subsidy intensity. The large differences between the intermediate and final results underscore the very high sensitivity of the models to (minor) changes in input data.

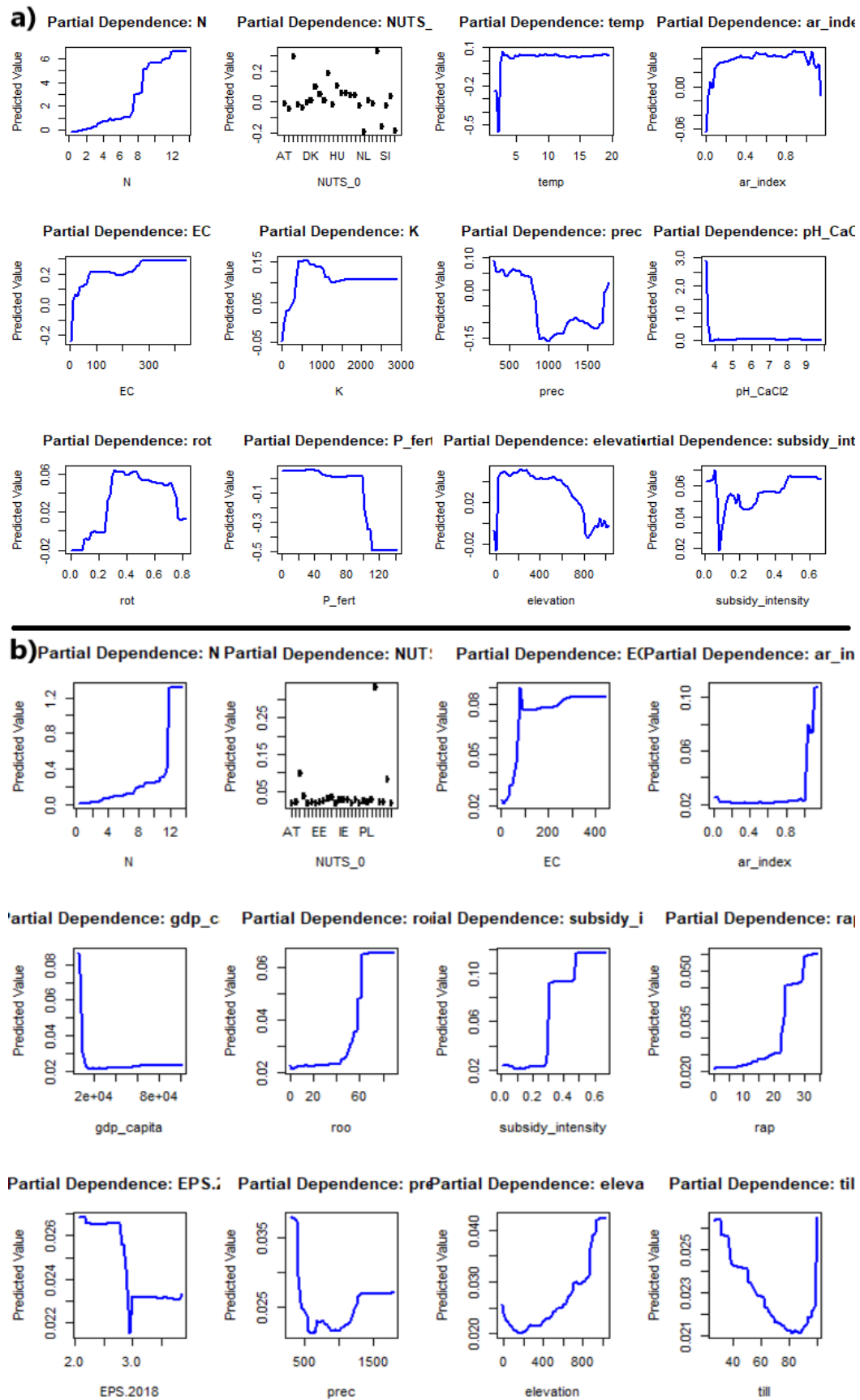


Figure 14: a) Partial dependence plots of the 12 most important variables in the RF model for ΔSOC_{rel} and ΔSOC_{rel}

Appendix II: Use of Artificial Intelligence

Artificial Intelligence (AI) has been used as an aide in writing code, solving issues in *R* and occasionally finding papers based on prompts that specifically defined the target. Any code or other information provided by AI has been carefully checked. AI has not been used for generating any written text in this thesis. The author declares that the use of AI has been within the guidelines set by Wageningen University and the Soil Geography & Landscape group.