Applied Geochemistry 193 (2025) 106526

Contents lists available at ScienceDirect

Applied Geochemistry

journal homepage: www.elsevier.com/locate/apgeochem

Compositional analysis of the relationships between the organic matter content and chemical and physical properties of soil

Matthias Templ [a],*, Christoph Hofer [b]

[a] School of Business, University of Applied Sciences and Arts Northwestern Switzerland, Riggenbachstrasse 16, Olten, 4600, Solothurn, Switzerland
[b] Institute of Data Analysis and Process Design, Zurich University of Applied Sciences, Technikumstrasse 81, Winterthur, 8401, Zurich, Switzerland

ARTICLE INFO

Editorial handling by Tao Wen

Keywords:
Soil organic matter content
Compositional data analysis
Regression analysis

ABSTRACT

Soil organic matter (SOM) plays a crucial role in soil fertility, carbon sequestration, and ecosystem sustainability, making its accurate analysis essential for environmental and agricultural management. However, studying the relationships between soil organic matter content (SOMC) and its influencing factors remains challenging due to the compositional nature of soil constituents. This study addresses key methodological challenges in analyzing the relationships between SOMC and soil texture, chemical composition, and bulk density using compositional data analysis. Specifically, we solve methodological issues related to integrating compositional and non-compositional variables in regression modeling and apply, for the first time, compositional data analysis to a mix of compositions, including the SOMC composition. The study explores the multivariate dependencies of the log-ratio coordinates—transformations that map compositional data from the constrained simplex space to real space—of major chemical elements in the soil and their relationship to log-ratio coordinates of SOMC. To appropriately account for the compositional nature of both the chemical element composition and soil texture, compositional data analysis methods are employed. Additionally, since outliers are common in soil data, all estimations are carried out using robust estimation methods. The application focuses on topsoil in the canton of Zurich (Switzerland), providing new insights into these relationships. Some findings contrast with previous studies that did not adopt a compositional approach, revealing, for example, a weak positive association between calcium and SOMC, a positive effect of phosphorus, and a decreasing dominance of organic matter in soil texture with increasing bulk density. Furthermore, free and open-source software has been extended to enable linear regression modeling that integrates both compositional and non-compositional explanatory variables, offering a practical solution to these methodological challenges in soil science.

1. Introduction

Soil is a complex, irreplaceable, and indispensable natural resource essential for sustaining ecosystems and providing humans with food, energy, and fiber. Soil organic matter content (SOMC) is a key component of soil, particularly in the upper layers, as it plays a crucial role in storing water and nutrients while binding minerals. Soil organic matter content originates from plant and microbial residues and undergoes a series of chemical transformations mediated by soil organisms (Bot and Benites, 2005; Zuazagoitia and Villarroel, 2016). Over time, these transformations lead to the formation of humus, a relatively stable organic fraction that can persist in the soil for extended periods (Sokolov et al., 2021).

Beyond its role in soil fertility, soil organic carbon (SOC), a major component of SOMC, plays a crucial role in carbon storage and climate regulation. SOC acts as a significant carbon sink, helping to mitigate climate change by sequestering atmospheric CO_2. The stability of SOC depends on various factors, including mineral interactions and microbial activity, which influence carbon turnover and long-term storage potential (Sparks et al., 2023).

The interactions between soil organic matter content and minerals drive key processes influencing soil fertility, structure, and carbon sequestration. These processes include the stabilization of organic matter through mineral associations, the decomposition of organic material by microbial activity, and the regulation of carbon cycling through biochemical transformations (Sparks et al., 2023). However, many studies investigating these relationships do not account for the compositional nature of soil constituents. Given that both soil texture (sand, silt, clay, and organic matter) and the chemical composition of soils are inherently compositional, improper statistical treatment, i.e., a non-compositional analysis, can lead to biased results (Filzmoser et al.,

* Corresponding author.
E-mail address: matthias.templ@fhnw.ch (M. Templ).

https://doi.org/10.1016/j.apgeochem.2025.106526
Received 15 April 2024; Received in revised form 26 June 2025; Accepted 10 August 2025
Available online 23 September 2025

2018). The higher the contribution of, for example, clay and silt, the lower the relative proportion of organic content. Many studies have analyzed soil organic matter content through mapping and single-variable approaches (Yuan et al., 2021), but such analyses may introduce bias because they fail to account for the compositional constraints of soil texture.

The constant sum constraint is a fundamental characteristic of compositional data, where the components of a composition (such as percentages, proportions, or parts of a whole) always sum up to a constant value. This constraint imposes certain mathematical and statistical limitations that make the analysis of compositional data different from unconstrained data (Pawlowsky-Glahn and Egozcue, 2006). The so-called simplex, say S_D, defines the sample space of a D-part composition, which differs from the real space $\mathbb{R}^D$ (Egozcue and Pawlowsky-Glahn, 2018). Consequently, standard statistical methods applied directly to compositional data can lead to spurious correlations (Pearson, 1897) and biased interpretations due to the inherent relative nature of the components (Pawlowsky-Glahn et al., 2007; Templ and Gonzalez-Rodriguez, 2024a).

Compositional data analysis (CoDa) provides appropriate statistical methods to address these challenges by employing log-ratio transformations (Aitchison, 1986). These transformations map compositional data to an unconstrained space, enabling valid statistical inference while respecting the relative nature of the data. Without such transformations, an increase in one component may misleadingly appear as an increase in its absolute quantity, when in fact it may result from a decrease in other components. This issue is particularly relevant in soil science, where soil texture (sand, silt, clay, and organic matter) and chemical compositions are inherently compositional (Filzmoser et al., 2018).

1.1. Compositional analysis in soil science

Compositional methods have been widely applied in geochemistry and mineralogy but remain underutilized in studies of soil organic matter content. Many studies have investigated soil organic matter (SOM) without considering compositional constraints, often using standard predictive modeling approaches based on soil properties such as clay content (Lazzaretti et al., 2020). These approaches fail to account for the relative nature of soil constituents, potentially leading to biased conclusions.

Other studies have attempted to analyze the relationships between SOM and chemical elements but have applied non-compositional methods (Dai et al., 2019; Yuan et al., 2021; Voltr et al., 2021), limiting their ability to capture true dependencies within the soil matrix.

Some research has incorporated compositional methods but with incomplete application. For example, Keskinen et al. (2022), Tarvainen et al. (2019) utilized pivot log-ratio transformations for soil chemical compositions, yet a comprehensive compositional regression framework integrating both SOM and soil texture composition remains underexplored.

Recent advances in soil science have introduced machine learning (ML) methods, including deep learning (Fernandes et al., 2019), to predict organic matter content. However, ML models often do not account for the compositional nature of soil data, resulting in suboptimal predictions due to incompatible loss functions. Even deep learning models require an appropriate transformation of the data into real space to improve prediction performance (Templ, 2021a,c).

Additionally, many applied studies rely on the additive log-ratio (ALR) transformation, particularly in mapping applications such as SoilGrids (Poggio et al., 2021). However, ALR coordinates are non-orthogonal, which can distort statistical interpretations and complicate regression modeling. We argue that the isometric log-ratio (ILR) or centered log-ratio (CLR) transformations provide a more appropriate approach for modeling compositional dependencies in soil science.

1.2. Addressing the missing link

The existing literature has not yet fully explored the interactions between soil organic matter content, soil chemical composition, soil texture, and bulk density using a rigorous compositional framework. Previous research has often considered only partial aspects of soil composition, neglecting the compositional constraints that affect statistical modeling. Additionally, soil texture, defined by its components—sand, silt, clay, and soil organic matter—has been frequently treated without considering its inherently compositional nature. Similarly, the role of major chemical elements such as aluminum (Al), calcium (Ca), iron (Fe), potassium (K), magnesium (Mg), manganese (Mn), sodium (Na), phosphorus (P), silicon (Si), and titanium (Ti) in relation to SOMC remains underexplored when analyzed outside of a compositional data framework.

Moreover, the treatment of explanatory variables as compositions has been largely overlooked. Our study addresses this gap by developing a compositional regression framework that allows for the simultaneous integration of compositional explanatory variables (e.g., chemical element compositions, SOMC and soil texture) and non-compositional variables (e.g., pH, temperature, bulk density). This approach ensures that the relative nature of soil constituents is appropriately accounted for, reducing potential biases and improving the interpretation of soil property relationships. Software was still missing to treat compositional and non-compositional explanatory variables simultaneously.

In addition, robust estimation in the presence of outliers isn't well reflected in literature.

1.3. Objectives and research questions

This study is not primarily about prediction but rather about interpreting the relationships between soil organic matter content and its influencing factors from a pedological perspective. By 'pedological perspective', we refer explicitly to applying compositional data analysis to interpret soil properties in a manner consistent with soil science principles, focusing on relative relationships between compositional parts and their influence on SOMC. This approach contrasts with predictive modeling aimed at universal applicability, instead prioritizing accurate representation and interpretation of compositional relationships in local or regional contexts.

Our goal is to establish interpretable relationships between SOMC, soil texture, chemical composition, and bulk density using compositional data analysis. Specifically, we aim to:

- Investigate the relationships between soil organic matter content composition and major chemical elements using exploratory analysis and compositional regression.
- Examine the influence of bulk density on SOMC to understand how physical soil properties interact with organic matter.
- Analyze the relationships between SOMC and soil texture by considering SOM as part of the soil texture composition alongside sand, silt, and clay.
- Extend compositional regression methods to include both compositional and non-compositional explanatory variables to provide a more comprehensive framework for analyzing soil data.
- Compare results with traditional non-compositional approaches to demonstrate the importance of proper compositional treatment in soil science.

To address the concern regarding linear models for potentially non-linear relationships, we emphasize that compositional transformations allow for a linearized representation of the relationships in an appropriate coordinate system. While raw data relationships may be non-linear, transforming compositional data using log-ratio coordinates can effectively linearize the structure, making standard regression models applicable.

1.4. Contributions

This study makes both a methodological and applied contribution to soil science and compositional data analysis:

- *Methodological Contribution*: We extend existing compositional regression frameworks to incorporate multiple compositional explanatory variables while allowing for the inclusion of non-compositional explanatory variables. This advancement provides a generalized statistical approach for analyzing soil compositions.
- *Applied Contribution*: We apply this framework to analyze topsoil in the canton of Zurich, demonstrating how compositional methods yield new insights into the relationships between SOMC, chemical elements, and bulk density.

By addressing these challenges, our work establishes a foundation for more robust soil data analysis and provides a methodological bridge to future machine learning applications in compositional data science.

2. Materials and methods

2.1. Geographical region

Samples from 512 locations were collected from the soil monitoring network of the canton Zurich. An aggregated database extract, covering six measurement periods between 1995 and 2020, forms the basis for these samples (Soil Protection Agency of the Canton of Zurich, 2021). Sample sites are divided into four distinct groups: (1) forests (all varieties, 185 sites), (2) croplands (agricultural, 217 sites), (3) permanent grasslands (non-agricultural, 105 sites), and (4) other (gardens and landscaped areas, 5 sites). The locations are illustrated in Fig. 2. Due to the scarcity of the category labeled ‘others’, not all our analyses take it into account.

2.2. Sampling

The sampling process of the topsoil (0–20 cm) from one site can be broken down into five steps: (1) composite samples (usually 25 never fewer than 16) over a support area of 100 m^2 (400 m^2 in forest), (2) drying of samples in a drying cabinet at 40 °C, (3) crushing the sample with a jaw crusher, (4) sieving to obtain the fine soil fraction (<2 mm), and (5) conducting various chemical and physical measurements. In this study, only samples from the topsoil are considered, as the focus is on the processes occurring in this layer.

In Table 1, all the variables used in this study are listed. These variables are grouped as follows:

- **Soil texture composition:** Sand, silt, clay, and soil organic matter, forming the fundamental components influencing soil structure and water retention.
- **Major chemical elements:** Aluminum, calcium, iron, potassium, magnesium, manganese, sodium, phosphorus, silicon, and titanium, which play a critical role in soil chemistry and fertility. The concentrations of the chemical elements at the sampling points are derived from the arithmetic mean of the measured values from up to six measurement periods conducted between 1995 and 2020.
- **Non-compositional variables:** Bulk density, pH measured in 0.01 M calcium chloride solution based on Agroscope (2020), altitude measured in the field with GPS, long-term average annual precipitation, and temperature (Soil Protection Agency of the Canton of Zurich, 2021), providing additional environmental context.

In this study, we employ the ten most prevalent chemical elements found in the soil. X-ray fluorescence spectroscopy (XRF) is used to measure the amount of chemical elements in dry fine soil particles that are smaller than 2 millimeters. The air-dry fine soil, which is less than two millimeters in size, is further divided into the following fraction based on the Department of Agriculture (1999): clay (smaller than 0.002 mm), silt (between 0.002 mm and 0.05 mm), and sand (between 0.05 mm and 2 mm). The combination of these different fine soil fractions and the organic matter make up the soil texture used in this paper. The organic matter is determined by the organic carbon, which is determined by dry ashing of the fine soil, and not by the particle size.

We also consider two different bulk densities: natural bulk density and bulk density for particles < 2 mm. The first is the density measured with the Burger cylinder from the original sample before the soil is sieved and dried. The second is the density measured after the soil has been sieved and dried, which is referred to as the bulk density of the fine soil.

3. Statistical methods

Compositional data analysis involves specialized statistical techniques to appropriately handle the constant sum constraint of compositions. To mitigate the issues posed by this constraint, log-ratio transformations such as centered log-ratio (CLR) and isometric log-ratio (ILR) are applied (Aitchison, 1986). These transformations allow the data to be analyzed in an unconstrained space, enabling valid statistical inference while preserving the relative nature of the data.

In this study, we apply CoDa methods to analyze soil texture and chemical composition in relation to soil organic matter content (SOMC). Specifically, we use exploratory analysis with principal component analysis (PCA) on log-ratio-transformed compositions and apply linear regression models incorporating compositional explanatory variables. Robust estimation techniques are used to address potential outliers, which are common in soil datasets.

By employing compositional regression, we ensure that both soil texture and chemical composition are appropriately modeled as constrained variables, preventing the biases inherent in traditional statistical methods that ignore the compositional nature of the data.

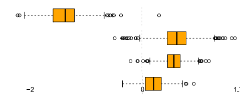
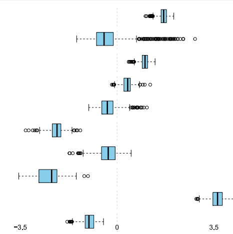
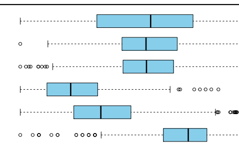
The following part first introduces the compositional transformations used in this article, namely the centered log-ratio and the pivot log-ratio transformation. Note, that the additive log-ratio (ALR) transformation is often used in compositional data analysis, but it has notable drawbacks. First, ALR coordinates are non-orthogonal, meaning statistical analyses performed in this transformed space can be difficult to interpret and may lead to misleading results (Filzmoser and Hron, 2019). Second, the choice of the denominator part in the transformation is arbitrary, which can introduce variability and inconsistencies in model comparisons. Third, ALR does not preserve subcompositional coherence (Pawlowsky-Glahn et al., 2007), meaning results obtained for a subset of components may not be consistent with those from the full composition. Finally, alternative transformations such as the isometric log-ratio (ILR) provide an orthonormal basis, making statistical modeling more robust and interpretable (Egozcue et al., 2003; Coenders and Pawlowsky-Glahn, 2020).

In the following, we introduce two compositional transformations used in this study. Centered log-ratio (CLR) coefficients are primarily employed for visual representations, such as boxplots and principal component analysis (PCA). In contrast, pivot log-ratio (PLR) coordinates are particularly useful for regression analysis and also play a role in PCA when robust estimation is applied (Filzmoser et al., 2009).

3.1. Centered log-ratio coefficients

Centered log-ratio (clr) coefficients are widely used in compositional data analysis. This approach is essential when dealing with data components such as proportions, percentages, or parts of a whole that are relative to each other and sum to a constant (usually 1 or 100%). The clr transformation involves taking the natural logarithm of each component of the composition divided by the geometric mean

Table 1
Soil texture composition, chemical element composition and *external*/non-compositional variables. Chemical elements are measured by XRF. $\bar{x}_{(geom)}$ represents the geometric mean, which serves as a compositional measure of the central tendency of a distribution. The centered log-ratio (CLR) coefficients are presented to represent the relative contributions of the variables to the overall composition of the soil texture and the chemical element composition. Since CLR coefficients have a zero-sum constraint, their variance is inherently limited. Therefore, the median absolute deviation (MAD) serves as a more robust measure of dispersion. The relative range of both compositions is indicated (from −2 to 1.7 and −3.5 to 3.5) as well as the 0 line, pointing out that the relative contribution of the corresponding variable to the total composition is equal to the geometric mean of all components in the composition.

Description	Unit	$\bar{x}_{(geom)}$	CLR coefficients	$\mathrm{MAD}_{(CLR)}$
SOMC	g/100 g air dried fine soil	4.8		0.28
sand	g/100 g air dried fine soil	36		0.25
silt	g/100 g air dried fine soil	34		0.17
clay	g/100 g air dried fine soil	23		0.21
Aluminum	g per 100 g dry soil	4.50		0.14
Calcium	g per 100 g dry soil	0.65		0.44
Iron	g per 100 g dry soil	2.28		0.14
Potassium	g per 100 g dry soil	1.22		0.17
Magnesium	g per 100 g dry soil	0.62		0.30
Manganese	g per 100 g dry soil	0.10		0.22
Sodium	g per 100 g dry soil	0.36		0.36
Phosphorus	g per 100 g dry soil	0.40		0.40
Silicon	g per 100 g dry soil	31.74		0.25
Titanium	g per 100 g dry soil	0.31		0.23
Description	**Unit**	$\bar{x}$	**values**	**MAD**
pH	measured with pH meter	5.5		1.25
natural bulk density (burger cyl.)	kg per m^3	1.2		0.17
bulk density (fine soil < 2 mm)	kg per m^3	1.1		0.17
altitude	meters above sea level	566		133
precipitation	average precip. in ml	166		14.68
temperature	average temp. in °C	8.3		0.59

of all components. Mathematically, for a composition with components $x_1, x_2, \ldots, x_n$, the clr of component i is given by Aitchison (1986):

$$\mathrm{clr}(x_i) = \log\left(\frac{x_i}{g(x)}\right) \tag{1}$$

where $g(x)$ is the geometric mean of all components $x_1, x_2, \ldots, x_n$.

The clr coefficient is a measure of the relative magnitude of each component in comparison to the geometric mean of all components. A coefficient of 0 indicates that the component's proportion is equal to the geometric mean, while positive values signify proportions higher than the geometric mean and negative values signify proportions lower than the geometric mean. This transformation allows for the application of standard statistical techniques such as linear regression or principal component analysis, and enables meaningful comparisons between components, overcoming the limitation of the constant sum constraint. However, the interpretation of clr coefficients is dependent on the composition of the data set, and the clr transformation treats all parts of the composition equally.

3.2. Pivot log-ratio coordinates

The clr coefficients map a composition $\mathbf{x}$ from S^D to a $(D-1)$-dimensional hyperplane in $\mathbb{R}^D$ (Aitchison, 1986). The isometric logratio (ilr) coordinates are used to construct an orthonormal basis in this hyperplane and express the composition as a vector $\mathbf{z}$ in $\mathbb{R}^{D-1}$. This avoids the singularity issue that occurred with clr coefficients.

One particular choice of a ilr basis – the pivot coordinates – leads to Fišerová and Hron (2011), Filzmoser et al. (2018)

$$\mathrm{pivotCoord}(\mathbf{x}) = \mathbf{z} = (z_1, \ldots, z_{D-1})'$$

with

$$z_j = \sqrt{\frac{D-j}{D-j+1}} \ln \frac{x_j}{\sqrt[D-j]{\prod_{k=j+1}^{D} x_k}} \quad \text{for} \quad j = 1, \ldots, D-1. \tag{2}$$

The rationale for this notation is clear: one element (in this case, x_1) is chosen as the pivot, which is only present in the first coordinate. This selection is also significant for the entire coordinate system. Reformulation for a $n \times D$ matrix $\mathbf{X}$ of compositional data, with $i = 1, \ldots, n$ compositions $\mathbf{x}_i' = (x_{i1}, \ldots, x_{iD})$ in the rows of $\mathbf{X}$, the elements of the $n \times (D-1)$ matrix of pivot coordinates $\mathbf{Z}$ are given by

$$z_{ij} = \sqrt{\frac{D-j}{D-j+1}} \ln \frac{x_{ij}}{\sqrt[D-j]{\prod_{k=j+1}^{D} x_{ik}}}. \tag{3}$$

The pivot coordinates have the property that the part x_1 only appears at the coordinate z_1. This is not the case for other parts; x_2, for example, appears in z_1 and in z_2. The appeal of highlighting a component in a coordinate system is that z_1 encapsulates all the relative data (logarithms) related to x_1 and y_1,

$$\begin{aligned} z_1 &= \sqrt{\frac{D-1}{D}} \ln \frac{x_1}{\sqrt[D-1]{\prod_{k=2}^{D} x_k}} \\ &= \sqrt{\frac{1}{D(D-1)}} \left(\ln \frac{x_1}{x_2} + \ln \frac{x_1}{x_3} + \ldots + \ln \frac{x_1}{x_D} \right), \end{aligned} \tag{4}$$

and can thus be interpreted as the relative dominance of x_1 within the given composition. In other words, z_1 expresses the level of dominance of part x_1 relative to the other parts, taking into account the geometric mean in the denominator. Positive values of z_1 indicate that the first part dominates the composition compared to an "average part" (formed by the geometric mean), while $z_1 < 0$ implies the opposite. A value of $z_1 = 0$ suggests a balanced state between x_1 and the average behavior of the other parts in the composition. No other part can be interpreted in this way, so the definition of pivot coordinates in (2) is specifically designed in favor of an interpretation for the first part. This property is used in Sections 3.4 and 3.5. However, the first part of a composition does not necessarily have a special role in a practical data set. It can still be of interest to gain an interpretation for a single part within a given composition that is not the first part. To do this, one can rearrange the parts of the composition so that the part of interest is in the first position, and then the pivot coordinates (see Eq. (2)) can be calculated for the rearranged composition.

If the interest is in part x_l, where $l \in \{1, \ldots, D\}$, the original composition $\mathbf{x} = (x_1, \ldots, x_D)'$ is rearranged to

$$\mathbf{x}^{(l)} = (x_l, x_1, \ldots, x_{l-1}, x_{l+1}, \ldots, x_D)' =: (x_1^{(l)}, x_2^{(l)}, \ldots, x_l^{(l)}, x_{l+1}^{(l)}, \ldots, x_D^{(l)})'.$$

The pivot coordinates corresponding to Eq. (2) for the permuted composition are (see also Filzmoser et al., 2018)

$$z_j^{(l)} = \sqrt{\frac{D-j}{D-j+1}} \ln \frac{x_j^{(l)}}{\sqrt[D-j]{\prod_{k=j+1}^{D} x_k^{(l)}}} \quad \text{for} \quad j = 1, \ldots, D-1, \tag{5}$$

which define the coordinates $\mathbf{z}^{(l)} = (z_1^{(l)}, \ldots, z_{D-1}^{(l)})'$. It is clear that $z_j^{(1)} = z_j$, for $j = 1, \ldots, D-1$. Here, only part x_l is moved to the first position, while the order of the remaining parts is kept the same. This is because the focus is on $z_1^{(l)}$, which contains all the relative information about the part x_l. This notation is used in Sections 3.4 and 3.5.

3.3. Principal componenent analsysis

Principal component analysis (PCA) is a method used to obtain new orthogonal projections of the original data. This is done by creating a weighted sum (linear combination) of the variables in the data set such that the projected values have the highest variance. The first principal component (the first score vector) explains most of the data, while the second principal component is the weighted sum of the variables with two constraints. It explains the second highest variance of projected data and is orthogonal to the first principal component. This principle continues for all D principal components. PCA is often used to characterize the components of the chemical composition of soil samples (Filzmoser et al., 2018) and is commonly employed as an exploratory tool to interpret multivariate dependencies in a data set. Biplots of the first two principal components are often plotted to show the projected observations along with the sign and magnitude of each variable's contribution to the first two principal components. Selecting any other two principal components in a biplot provides additional information. As demonstrated in Templ and Templ (2020, 2021) spurious results are obtained when PCA is applied to standardized, non-standardized, logarithmized or untransformed data. In this study, PCA was applied after a compositional treatment of the data (centered log ratio coefficients). To reduce the influence of outliers, robust principal component analysis and, respectively, a robust biplot are used according to Filzmoser et al. (2009). The rationale for employing this method is that using the pivot coordinate representation has a drawback in that the newly derived variables cannot be directly interpreted in terms of the initial input variables. In the study by Filzmoser et al. (2009), the authors propose an alternative approach in which they estimate the principal components based on pivot log-ratio coordinates. However, they subsequently transformed the results back into centered log-ratio coefficients, which can be interpreted in the biplot. Furthermore, when considering both compositions, the concentrations of chemical elements and the composition of material matter, the approach of Templ et al. (2011) is used that implements the analysis of the principal components of multi-compositions.

3.4. Regression analysis with compositional explanatory variables

In our research, the response variable is always a composition, representing the proportion of soil organic matter in the material fraction. The other composition—major chemical element concentrations—can serve as explanatory variables. While SOMC is part of the broader soil texture composition (sand, silt, clay, and SOMC), we also analyze the chemical composition separately, using specific elements from this composition as explanatory variables.

First, the D-part composition $\mathbf{x} = (x_1, \ldots, x_D)'$ is expressed in a D coordinate system (2). If n measurements of explanatory variables are taken together with those of the real response variable Y, the resulting models can be written as follows:

$$Y_i = b_0^{(l)} + b_1^{(l)} z_{i1}^{(l)} + \cdots + b_{D-1}^{(l)} z_{i,D-1}^{(l)} + \varepsilon_i, \; i = 1, \ldots, n; \; l = 1, \ldots, D. \tag{6}$$

The R^2, F-statistic and inference and value of b_0 remain the same for all $l = 1, \ldots, D$ (Filzmoser et al., 2018), however, the t statistics, the normalized t statistics (the p values) and standard errors of the remaining regression coefficients do not. The parameters $b_1^{(l)}$ are of particular interest due to the interpretation of the first coordinates in (2). The remaining coordinates $z_2^{(l)}, \ldots, z_{D-1}^{(l)}$ are necessary to represent the remaining parts $x_2^{(l)}, \ldots, x_D^{(l)}$, and thus cannot be excluded from the model. The parameters $b_0, b_1^{(1)}, \ldots, b_1^{(D)}$ along with their associated characteristics (standard errors, t-statistics and p-values) are usually presented in a single table, as if they all came from one regression model. It is important to remember, however, that the results are from D separate regression models, so it would not be sensible to try to make inferences about all of the $b_1^{(1)}, \ldots, b_1^{(D)}$ parameters at the same time. In conclusion, for each regression coefficient, one must construct a model that uses a pivot coordinate representation with the associated variable as the pivotal element. One should note that the estimation of b_0 is also independent on the choice of l (Filzmoser et al., 2018).

3.5. Regression analysis with a mix of compositional and non-compositional explanatory variables

We look at the situation where the explanatory variables consist of a combination of compositional and non-compositional variables. The non-compositional variables are left unaltered. This is easy to conceptualize. For example, when there is one external categorical variable with $c_1, \ldots, c_m$ categories we get

$$\begin{aligned} Y_i =& b_0^{(l)} + b_1^{(l)} z_{i1}^{(l)} + \cdots + b_{D-1}^{(l)} z_{i,D-1}^{(l)} + b_D I(x_{i,D} = c_2) + \cdots \\ &+ b_{D+m-1} I(x_{i,D+m-1} = c_m) + \varepsilon_i, \end{aligned} \tag{7}$$

with $i = 1, \ldots, n;\; l = 1, \ldots, D$ and I an indicator function that equals 1 (if the condition in the brackets is true) or 0. Note that reference category c_1 is included in b_0.

Consequently, similar to the transformation of variables with non-compositional data, where some explanatory variables may be altered (e.g. with the logarithm), transformations are also applied here. This involves:

(a) a series of pivot transformations to ensure that one compositional variable is always in a pivotal position, so that the corresponding regression parameter can be accurately inferred.
(b) the non-compositional variables are treated independently and do not require a series of transformations. The results are the same for each coefficient and the corresponding inference (t statistics) independently of which of the l' variables is placed in the pivotal position in (a). The reason why they remain the same is simple because any pivot transformation l' contains the same information as for any pivot transformation $l \in \{1, \ldots, D\}$ and any of these pivot transformations represents an orthonormal basis in the hyperplane formed by the clr coefficients (Filzmoser et al., 2018). In other words, any pivot transformation with any choice of $l \in \{1, \ldots, D\}$ is orthonormal (log-ratio) coordinates that contain all relative information.

Developing this in software is a challenging task that had not been previously accomplished. The new implementation has been made accessible with this article and it is accessible through the R package robCompositions (Templ et al., 2011).

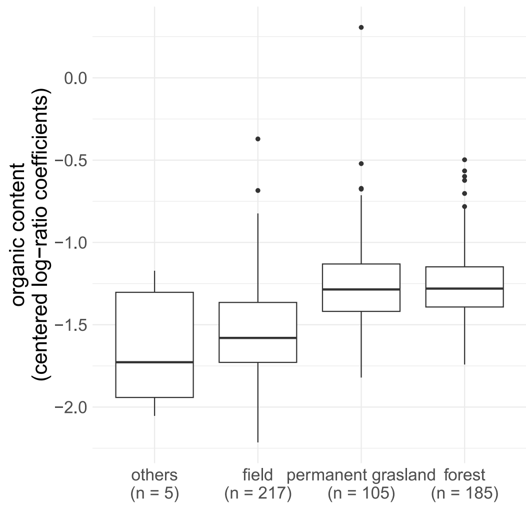


Fig. 1. Centered log-ratio coefficients showing the relative content of soil organic matter from the soil texture (sand, silt, clay, and soil organic matter).

3.6. Robust regression analysis with compositional response and non-compositional explanatory variables

In our second analysis, we used only the soil organic matter content after conversion to centered log-ratio coefficients as the response, which simplifies and reduces the approach to standard regression. For all fitted models, robust regression with MM-estimation, as proposed by Maronna et al. (2006), is employed to mitigate the impact and minimize the leverage effect of outliers. Robust MM-regression is a statistical technique used in regression analysis to handle data with outliers. It combines the ideas of two other robust regression methods: M-estimation (for efficiency) and S-estimation (for high breakdown point). The approach first applies the initial robust S-estimator with a high breakdown point to resist the influence of outliers in the data. This initial estimate is then refined using an M-estimator approach to achieve high statistical efficiency for the final model.

To assess the reliability of the joint regression model combining physical and chemical soil properties, we examined multicollinearity using variance inflation factors (VIFs). No problematic collinearity levels were detected. Furthermore, by applying pivot log-ratio coordinates for the chemical elements, we mitigated spurious correlations due to the compositional nature of the data, ensuring that the regression coefficients can be interpreted more reliably.

4. Results

4.1. Univariate and basic explorative analysis

Before discussing the main objectives of our research, we highlight the relationship of soil organic matter content in different fields. The soil texture (sand, silt, clay and organic content) is represented in centered log-ratio coefficients using Eq. (1). Therefore, the coefficients of the soil organic matter represent the log-ratio of its proportion relative to the geometric mean of all components in the composition. Specifically, for the soil organic matter, the coefficient in the clr transformation indicates how its proportion deviates from the geometric mean of sand, silt, clay, and soil organic matter. If the coefficient for soil organic matter is positive, it signifies that the soil organic matter is relatively more abundant compared to the geometric mean of all components.

The organic content coefficients are shown in Fig. 1, which shows that permanent grasslands and forests, due to their combination of plant diversity, continual input of organic matter, natural soil processes, and minimal disturbance of the soil, have a higher relative organic content than agricultural croplands that are managed more intensively. Because soil organic matter is typically not dominant in a soil texture, it results in centered log-ratio coefficients that are negative, indicating proportions of organic matter that are lower than the average of the whole composition.

Fig. 2 shows the relative content of soil organic matter content in soil texture. Hereby, the centered log-ratio coefficients are shown, while in the supplementary material also the untransformed SOMC and the log10 transformed SOMC is shown. The relative SOMC is generally higher in permanent grasslands and forests, but also regional differences are visible, e.g. higher SOMC for forest sites near the gold coast (next to the Zurich lake) and lower SOMC values in the gradient between Uster and Hinwil in the North-West of the gold coast.

4.2. Organic content versus main chemical elements

The aim is to gain a better understanding of the relationship between the composition of the main chemical elements and the organic content of the composition of the soil texture (sand, silt, clay, and soil organic matter). The centered log-ratio coefficients (Eq. (1)) of the organic matter content in the soil texture are used. First, the distributions of the centered log-ratio coefficients of the main chemical elements are shown in Fig. 3. It is visible that silicon has a dominance over the other main chemical elements. On the contrary, manganese and phosphorus have the lowest median and are dominated by the other components. The observations of calcium, sodium, magnesium and phosphorus have a wide range, while the observations of aluminum, iron, potassium, and titanium are more limited. Furthermore, for several chemical elements, such as phosphorus, titanium, and calcium, we observe a slight shift in the centered log-ratio coefficients for forested areas compared to cropland and permanent grasslands. This suggests that the effect of land use on organic content may not be negligible.

The concentrations of Ti, Si, Na and K are greater for observations with low organic content. Hereby, the quartiles are calculated from the centered log-ratio coefficients – built from the composition sand, silt, clay and organic content – of organic content (low organic content). Conversely, the concentrations of Mg and Ca are higher for observations with an organic content above the third quartile (high organic content).

In summary, there are notable distinctions between the forest category and the cropfield and permanent grassland samples in terms of the relative concentrations of the main elements. In addition, for some chemical elements, the relative concentration is contingent on the relative organic content. Note that these boxplots illustrate general tendencies rather than statistically validated effects. The rigorous assessment of land-use impact is conducted later in the manuscript through formal modeling approaches (e.g., Table 6), where we quantitatively evaluate relationships.

A multivariate analysis with a principal component analysis will provide further information. The biplot in Fig. 4 displays the first two principal components that were derived from a robust principal component analysis.

In the upper left graphic, robust PCA was used to input the main element concentrations of all observations. The convex hulls do not show much difference, as the scores for all land-use groups are spread out. However, there are more scores with low relative organic concentration and positive values on the second principal component. This component is characterized by high weights for P and Na and low weights for Mg and Ca. Further interpretation can be done by looking at subsets based on land-use. For permanent grasland observations, those with low organic content are mostly characterized by high relative concentrations of Na and low concentrations of Mn and Ca. On the other hand, compositions with high organic content are distinguished by high relative concentrations of Ca and Mg. This is more or less also visible for land-use field. A different picture emerges for land-use

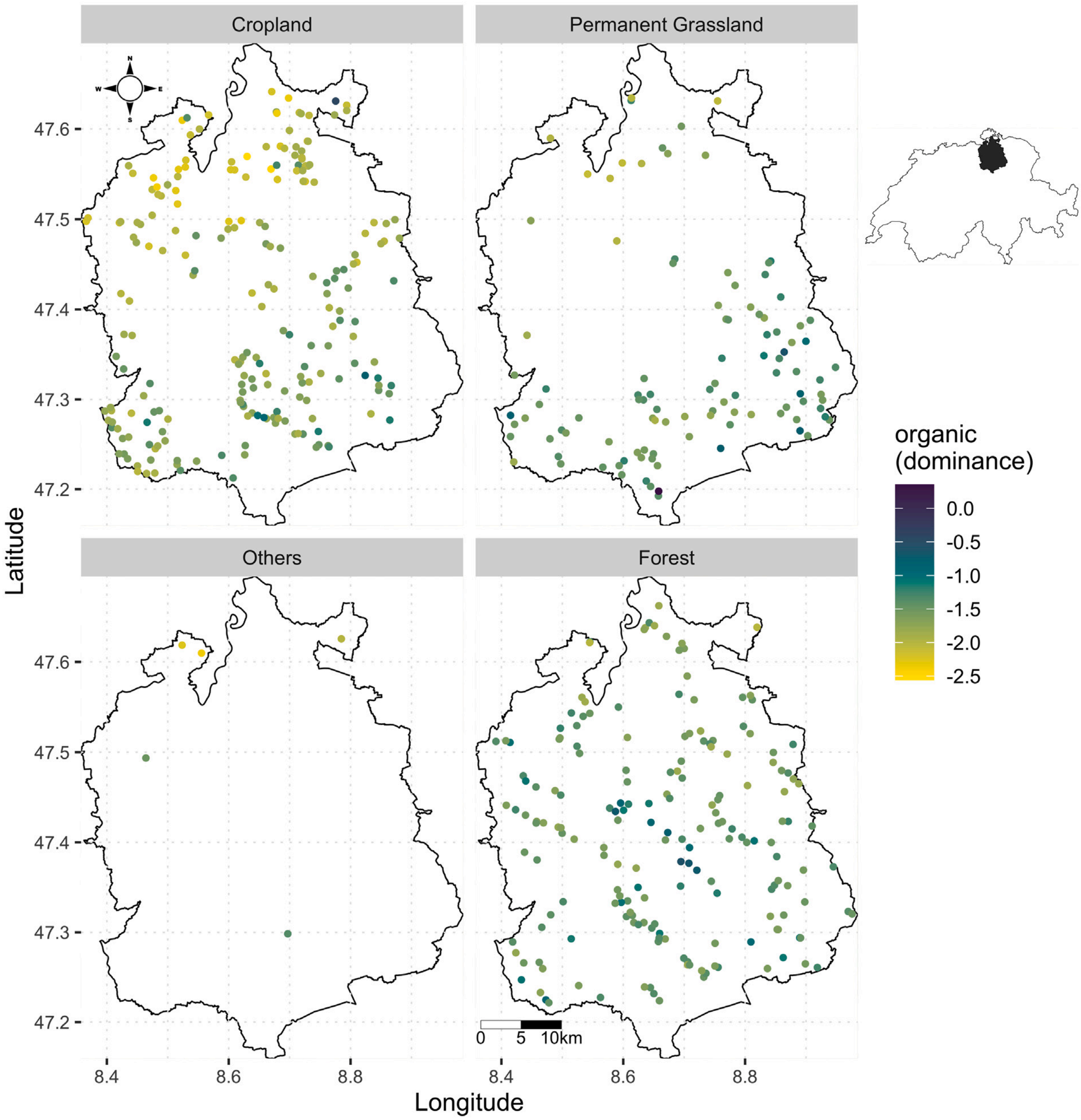


Fig. 2. Spatial distribution of the centered log-ratio coordinates of soil organic matter content, i.e., the relative soil organic matter in the soil texture at the sampling sites. The black border represents the boundaries of the Canton of Zurich. In the top right corner, the national borders of Switzerland are shown for orientation, with the Canton of Zurich as a black area.

in forests, where low relative soil organic matter concentrations are mostly associated with higher relative concentrations of Ti and K (as well as Na and Si).

This pattern can be attributed to several factors. Mineral-dominated soils may have a higher proportion of inorganic minerals, and since soil organic matter is lower, the relative concentration of these inorganic elements appears higher. Additionally, organic matter plays a crucial role in retaining nutrients in the soil; lower organic content reduces the soil's capacity to hold onto nutrient-rich compounds, making mineral elements more prominent. Moreover, various forms of organic matter can act as electron donors in mineral reduction processes (Xu and Tsang, 2024).

Titanium (Ti), in particular, is generally considered an immobile element (Brimhall and Dietrich, 1987). Its association with low organic matter soils could indicate that these soils have undergone extensive leaching, resulting in the depletion of more mobile nutrients while leaving behind relatively immobile elements like Ti. This further supports the idea that organic matter not only influences nutrient retention but also affects the overall geochemical composition of the soil.

In contrast, sites with high organic content are those with high relative concentrations of P, Mg, Ca, and Fe. A primary explanation is nutrient enrichment, as organic matter, such as decaying plant material, is rich in nutrients and releases elements such as P, Mg, and Ca into the soil as it decomposes (Zhang et al., 2022). This process is closely tied to microbial activity, which is typically elevated in organic-rich soils and plays a key role in nutrient cycling and mobilization (Wu et al., 2024). In particular, microbes can influence the mobility of elements such as P and Fe, either facilitating their release into the soil solution or contributing to their stabilization through mineral–organic interactions.

Fig. 3. Distribution of the centered log-ratio coefficients of the main chemical elements for low (below the 0.25 quantile), middle and large (above the 0.75 quantile) organic content and land-use.

Additionally, organic matter is known for its high cation exchange capacity (CEC), which allows it to retain and exchange positively charged nutrient ions such as Mg, Ca, and Fe more effectively (Solly et al., 2020). The decomposition of organic matter can also influence soil pH, further affecting nutrient availability. Altogether, these interacting processes help explain the higher relative abundance of key nutrient elements observed in high-OM soils.

In the following, a regression model is robustly fit for all three types of land-use. Robustness was crucial, since there are several extreme outliers in the data, e.g., for model 1, the weights of two observations were set to 0 while the others received weights mostly between 0.8 and 1.

In Table 2 one can see that at the 5% significance level, potassium, manganese, sodium, and phosphorus have a significant effect on the relative content of soil organic matter in the soil. For example, doubling the dominance of phosphorus over the averaged contributions of the other chemical elements would increase the dominance of organic matter over averaged contributions of the other parts of the soil texture (clay, slit and sand) by a factor of $1.36 = e^{0.31}$. On the other hand, potassium, manganese, and sodium have a significant negative influence on the response. Doubling the dominance of sodium over the average contributions of the other chemical elements would now produce a decrease in the dominance of organic matter over the average contributions of the other parts of the soil texture by a factor of $0.92 = e^{-0.08}$.

Table 2
Model 1. The relative soil organic matter content of land-use forests was modeled against the log-ratio coordinates of the primary chemical elements ($R^2 = 0.603$).

term	estimate	std.error	statistic	p.value
	Estimate	Std. Error	t value	Pr(> $\|t\|$)
(Intercept)	-1.99	0.57	-3.48	0.0006
Al_{pivot}	0.18	0.33	0.55	0.5779
Ca_{pivot}	0.05	0.03	1.55	0.1221
Fe_{pivot}	0.36	0.22	1.60	0.1109
K_{pivot}	-0.44	0.11	-3.77	0.0002
Mg_{pivot}	0.04	0.09	0.44	0.6581
Mn_{pivot}	-0.36	0.05	-7.19	1.75e-11
Na_{pivot}	-0.08	0.03	-2.25	0.0255
P_{pivot}	0.31	0.04	6.91	8.65e-11
Si_{pivot}	0.07	0.10	0.69	0.4856
Ti_{pivot}	-0.13	0.14	-0.92	0.3587

The magnitude of the influence on the response is different for samples in permanent grassland land use sites, see Table 3. For example, the dominance of aluminum (Al) over the averaged contributions of

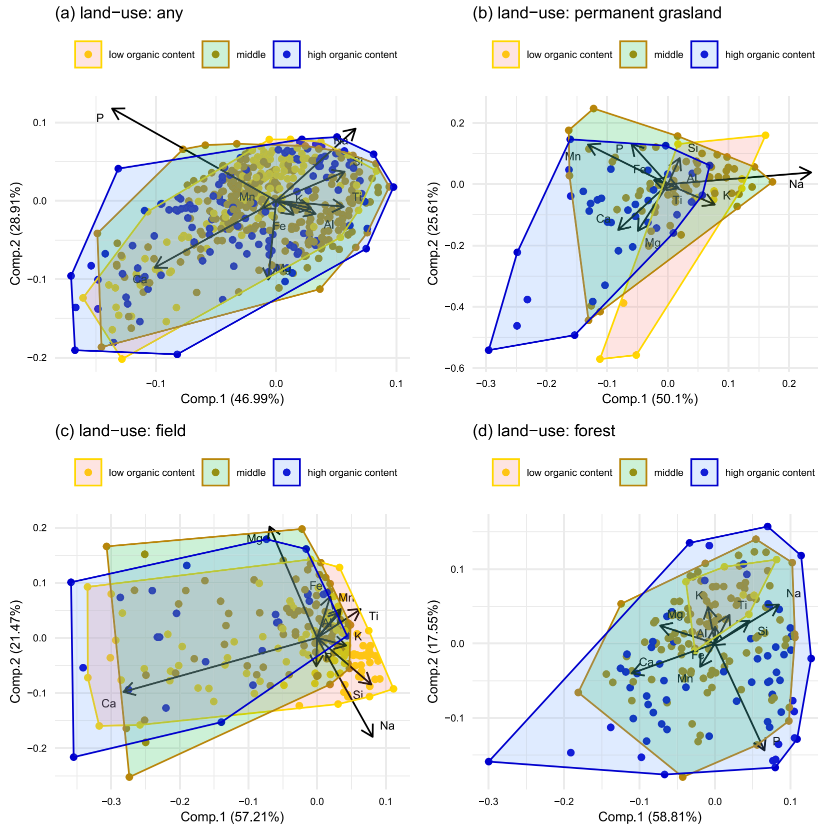


Fig. 4. Biplots of the first two principal components obtained with robust principal component analysis. The color of scores and convex hull distinguish between the three relative soil organic matter content groups (low, middle and high). (a) For any land-use, (b) for land-use permanent grasland, (c) for land-use field, and (d) for land-use forest.

Table 3
Model 2. The relative soil organic matter content of land-use permanent grasland was modeled against the log-ratio coordinates of the primary chemical elements ($R^2 = 0.69$).

term	estimate	std.error	statistic	p.value
(Intercept)	-4.011	0.89	-4.48	2.01e-05
Al_{pivot}	1.82	0.42	4.26	4.76e-05
Ca_{pivot}	0.14	0.09	1.49	0.1387
Fe_{pivot}	-0.22	0.32	-0.70	0.4846
K_{pivot}	-0.75	0.21	-3.55	0.0005
Mg_{pivot}	-0.10	0.11	-0.90	0.3669
Mn_{pivot}	-0.18	0.08	-2.09	0.0385
Na_{pivot}	-0.17	0.07	-2.44	0.0161
P_{pivot}	0.25	0.14	1.74	0.0842
Si_{pivot}	-0.09	0.15	-0.60	0.5487
Ti_{pivot}	-0.67	0.19	-3.48	0.0007

the other chemical elements increases the dominance of soil organic matter over the averaged contributions of the other parts of the soil texture (clay, slit, and sand) by a factor of $6.17 = e^{1.82}$. Aluminum is a common element in many soils, especially clay minerals. In more acidic soils, aluminum becomes more soluble and can be more dominant in the soil composition. This increased solubility and mobility of aluminum can influence the soil's chemical environment. There is an impact on organic matter, especially in acidic conditions, where aluminum is more active, where there can be an indirect effect on the breakdown and stabilization of organic matter. Acidic environments can slow the decomposition of organic matter, potentially leading to its accumulation. Therefore, a dominance of aluminum, indicating more acidic conditions, could be associated with an increase in soil organic matter (Scheel et al., 2008), as our results show.

Table 4 shows a different picture for land-use field. Iron (Fe) now has a significant effect, as do aluminum (Al), potassium (K), sodium (Na), phosphorus (P), silicon (Si) and titanium (Ti). Iron is correlated with organic matter content (Chen et al., 2020). It can play a role in the preservation of organic matter by binding with it and protecting it from rapid decomposition. The presence of iron could indicate the stability of organic matter in the soil. Potassium is an essential nutrient for plant growth. Its presence in the soil can indicate soil fertility and its capacity to support plant life (Grzebisz et al., 2012), which in turn can contribute to organic matter content through plant residues and root exudates. Sodium can affect the structure of the soil, particularly clay-rich soils. Although not directly related to organic matter, it can influence soil water retention and aeration (Tang et al., 2021), indirectly affecting organic matter decomposition. Phosphorus is another key nutrient for plants and microorganisms. High levels of phosphorus can support a vibrant soil ecosystem, contributing to the cycle of organic matter, while phosphorus deficits can lead to a decrease in plant activity (Hoque et al., 2023) and therefore to a decrease in soil organic matter in the long term. Silicon is beneficial for plant health, and numerous studies have demonstrated its beneficial effects (Pavlovic et al., 2021), particularly in strengthening cell walls (Sheng and Chen, 2020). Healthier plants can contribute more organic residues to the soil, thus increasing the organic content. Titanium is typically found in soil minerals and is not directly involved in the dynamics of organic matter. However, its presence may be indicative of certain types of soil that may have particular characteristics related to organic matter content. It also applies that titanium has beneficial effects on plant growth and healthiness (Lyu et al.) and thus the same argument as for Silicon may apply.

Our regression results indicate that certain chemical elements, such as potassium and titanium, are significant predictors of SOMC in both field and grassland soils. This suggests that some compositional relationships between soil organic matter and chemical elements are

Table 4
Model 3. The relative soil organic matter content of land-use field was modeled against the log-ratio coordinates of the primary chemical elements ($R^2 = 0.73$).

term	estimate	std.error	statistic	p.value
(Intercept)	-2.08	0.63	-3.30	0.0011
Al_{pivot}	2.16	0.27	7.88	1.80e-13
Ca_{pivot}	0.03	NA	NA	NA
Fe_{pivot}	-0.74	0.19	-3.89	0.0001
K_{pivot}	-1.05	0.13	-7.73	4.56e-13
Mg_{pivot}	0.04	0.06	0.75	0.4537
Mn_{pivot}	0.05	0.09	0.60	0.5453
Na_{pivot}	-0.24	0.07	-3.43	0.0007
P_{pivot}	0.51	0.06	7.61	9.18e-13
Si_{pivot}	-0.31	0.11	-2.74	0.0065
Ti_{pivot}	-0.42	0.11	-3.80	0.0001

consistent across land-use types, potentially reflecting underlying geochemical processes that are not strictly land-use dependent. For instance, K is a key nutrient in plant growth and organic matter turnover, which may explain its significance in both land-use types, while Ti, being a relatively immobile element, may reflect broader soil mineralogical characteristics rather than direct biological influences. These shared relationships indicate that while land use affects soil composition, certain elemental interactions with SOMC persist regardless of land-use differences.

When comparing the three models, we can see that iron (Fe) influences the response differently in each model (Model 1: significantly positive effect; Model 2: no significant effect; Model 3: significantly negative effect). Furthermore, only aluminum (Al), calcium (Ca), potassium (K), sodium (Na), phosphorus (P) and titanium (Ti) have the same influence on the response in terms of sign in all three models. Although the impacts of the other factors have already been discussed, calcium plays a role in promoting a conducive environment for the accumulation (Shabtai et al., 2023) and the preservation of organic matter in the soil, and calcium regulates the accumulation of organic carbon in the soil by mediating microbial communities (Dou et al., 2023). This occurs directly through its influence on the health of plants and microorganisms, as well as indirectly through its influence on the pH and structure of the soil.

4.3. Organic content versus bulk density

Organic matter is considered part of the fine soil fraction, while bulk density determined using the burger cylinder method reflects the density of the entire soil sample, including coarse fragments (often referred to as the soil skeleton).

The objective is to gain a better understanding of the relationships between the bulk densities of the soil and the centered log-ratio coordinates of the organic matter content. Two types of bulk densities are gauged during the data collection procedure.

First, the compositional PCA is applied to the composition of the mineral fraction (sand, silt, clay, and soil organic matter), see Fig. 5. Therefore, the bulk densities are provided as *external* (non-compositional) variables. It is clearly visible that the lower the relative content of soil organic matter content the higher the bulk density at fine soil level as well as measured with the burger cylinder. In addition, the higher the relative content of silt in the material fraction, the higher the bulk densities.

Moreover, robust linear regression is applied and the fit of the coefficients is shown in Table 5.

The fine soil bulk density is significant on the 5% significance level. Additionally, this model implies that double the value of the fine soil bulk density decreases the dominance of organic matter over the average contributions of the other parts of the soil texture (clay, silt, and sand) by a factor of $0.13 = e^{-2.064}$.

4.4. Relationships of organic content in the soil texture

The objective is to acquire a more profound comprehension of the connections and associations between the log-ratios of the mineral components. Note that in a 3-D ternary diagram, it is already visible that the relationship of soil organic matter content is weak with other soil textures (sand, silt, clay).

We follow the consistent approach for a compositional analysis. Soil texture components (sand, silt, clay, and organic matter) form a compositional system, meaning their values are inherently constrained. To account for this, we use log-ratio transformations (here pairwise log-ratios), which allow us to analyze the relative distribution of organic matter across different soil fractions while ensuring statistical validity. This approach avoids spurious correlations and provides a more robust interpretation of soil texture effects on organic matter content.

When further analysis of all pairwise log-ratios of the soil texture is carried out, we see that two effects are significant but all three models do not have much explanatory power, see Table 6. All models are fitted with robust MM-regression because of large outliers/leverage points present.

- Model 1 (log(organic/sand) explained by log(silt/clay)). The effect of silt/clay to organic/sand is positive ($\beta_1 = 0.83, R^2 = 0.2$), which means that the higher the log-ratios of silt/sand the higher the log-ratios of organic/sand. Compositions with high values in silt and thus low values in sand will likely result in a high organic content compared to the sand content. This is the strongest relationship of the three models. This can be explained because silt and clay particles are smaller and have a larger surface area compared to sand. This larger surface area allows these finer particles to retain more water and nutrients, which are beneficial for organic matter decomposition and accumulation. Furthermore, increased moisture retention in clay and silt-rich soils creates a favorable environment for microbial activity, which is crucial for the breakdown of organic matter and nutrient cycling. Additionally, sandy soils tend to dry out faster and retain less organic matter; i.e., sandy soils generally have a lower capacity to hold nutrients and water. These effects of clay and sand on soil composition aligns with standard principles in soil science.
- Model 2 (log(organic/clay) explained by log(silt/sand)). The effect of silt/sand to organic/clay is not significant
- Model 3 (log(organic/silt) explained by log(clay/sand)). The effect of clay/sand to organic/silt is positive ($\beta_1 = 0.2475, R^2 = 0.0855$), which means that the higher the clay/sand log-ratio the higher the log-ratios on organic/silt. Soil samples with a high clay and small sand content will likely have a large organic content compared to the silt content. The high water and nutrient retention properties of clay, combined with reduced leaching and increased microbial activity, contribute to a higher organic matter content in soils where clay is prevalent and silt and sand are in lower proportions (Matus, 2021).

We first investigate the dependencies between all pairwise log-ratios with the PCA. We then apply three different models for the linear regression analysis with the different pairwise log-ratios.

4.5. Organic content in light of environmental parameters, bulk density, chemical elements and remaining soil texture components

Models are examined to elucidate the different factors that influence organic matter in soil, see Table 7. The variable of interest in the model is the dominance of soil organic matter in the soil texture, specifically represented by the pivot log-ratio coordinates of soil organic matter in the soil texture.

The bulk density is an important explanatory variable with a negative coefficient, which means that the dominance of soil organic matter

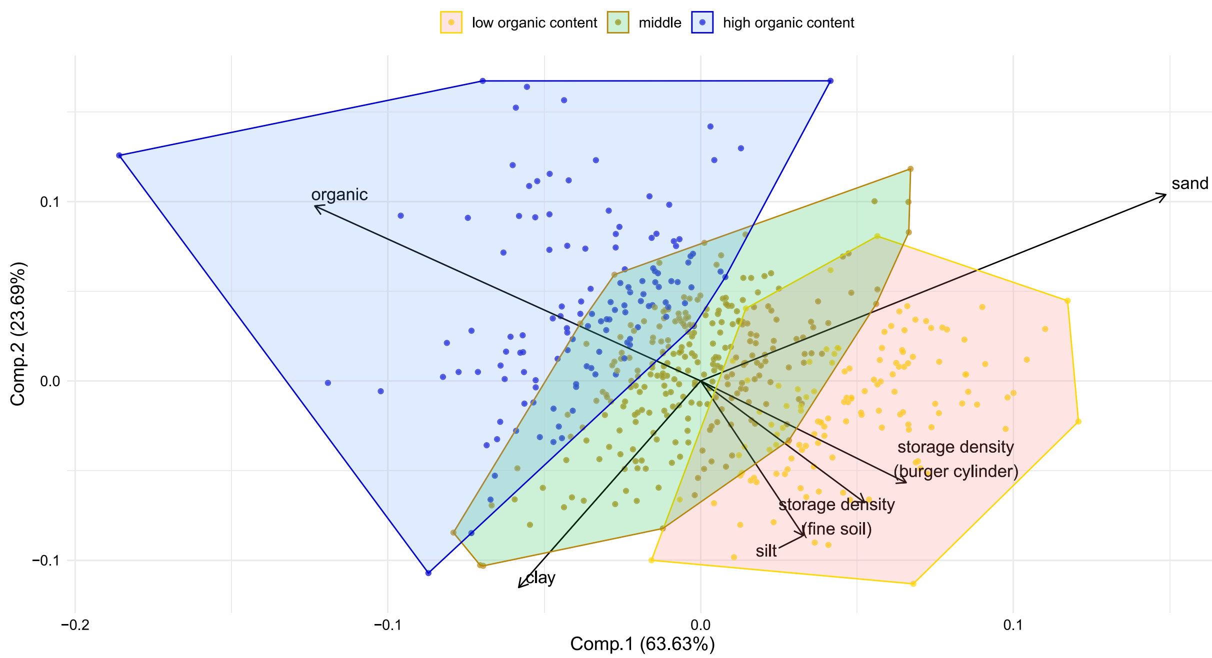


Fig. 5. Compositional biplot for soil texture and bulk density.

Table 5
Prediction of the relative (centered log-ratio coefficients) soil organic matter content with bulk densities of the fine soil and of the burger cylinder ($R^2 = 0.657$) using robust MM-regression.

term	estimate	std.error	statistic	p.value
(Intercept)	0.3159	0.0673	4.70	<3.4e-06
bulk density (burger cylinder)	-0.9493	0.1219	-7.79	<3.9e-14
bulk density (fine soil)	-0.5047	0.1217	-4.15	<3.9e-05

Table 6
Effects on the organic/sand ratio from clay/silt ratio and type of land use.

	term	estimate	p.value
1	(Intercept)	−1.72	0.00
2	clay.silt	1.27	0.00
3	landusepermanent grasland	0.17	0.13
4	landuseforest	−0.01	0.91
5	clay.silt:landusepermanent grasland	−0.24	0.35
6	clay.silt:landuseforest	−1.04	0.00

Table 7
Effects on the dominance of organic content in material composition. The response variable in the model is the pivot coordinates of soil organic matter content i.e. the dominance of organic content to sand, silt, and clay. The model is applied to all soil samples, and to forest, cropland, and grassland sites only. Gray color indicates the presence of significant effects. The values are approximated to two decimal places. The related (robustly fitted) R^2's are 0.773 (all), 0.694 (forest), 0.842 (cropland), 0.864 (grassland). Because of the presence of leverage points, robust MM-regression was applied.

	all		forest		cropland		grassland	
term	coef	p-val	coef	p-val	coef	p-val	coef	p-val
(Intercept)	**-2.649**	**0.000**	**-3.142**	**0.000**	-1.237	0.299	-3.119	0.076
bulk density	**-0.820**	**0.000**	0.218	0.290	**-0.530**	**0.001**	**-1.785**	**0.000**
soil density	**-0.425**	**0.001**	**-0.727**	**0.000**	**-0.434**	**0.021**	**0.910**	**0.002**
altitude	**0.001**	**0.000**	**0.001**	**0.000**	0.000	0.307	0.000	0.999
precipitation	-0.001	0.282	-0.001	0.526	**0.003**	**0.019**	0.001	0.335
temperature	**0.137**	**0.000**	**0.190**	**0.000**	0.050	0.407	-0.028	0.673
pH	**-0.068**	**0.000**	-0.006	0.850	-0.014	0.669	**0.123**	**0.008**
Al	**0.903**	**0.000**	0.265	0.480	**1.163**	**0.000**	**1.509**	**0.002**
Ca	**0.058**	**0.003**	0.054	0.309	0.037	0.197	0.125	0.054
Fe	-0.016	0.900	0.221	0.320	-0.278	0.086	-0.464	0.137
K	**-0.530**	**0.000**	-0.303	0.075	**-0.503**	**0.003**	**-0.495**	**0.017**
Mg	**0.091**	**0.011**	0.062	0.522	0.013	0.819	**-0.274**	**0.002**
Mn	**-0.151**	**0.000**	**-0.394**	**0.000**	0.029	0.703	0.004	0.965
Na	-0.037	0.365	-0.089	0.150	-0.063	0.356	-0.135	0.063
P	**0.119**	**0.000**	**0.297**	**0.000**	**0.292**	**0.000**	0.153	0.268
Si	-0.050	0.531	-0.092	0.502	**-0.365**	**0.001**	0.014	0.912
Ti	**-0.386**	**0.000**	-0.020	0.903	**-0.325**	**0.006**	-0.438	0.215

decreases for samples with a higher bulk density, except for forest sites only where the effect is not significant. Bulk density also has negative effects in all models except grassland. If altitude increases by one meter, the dominance of organic content will increase by 0.001, or in other words, if the altitude increases by 100 m, the dominance will increase by 0.1, so even significant for all sites the effect size is rather small (values on the dominance of organic content range between −2.56 and 0.35). Temperature has a rather large and significant effect size in forests. Increasing the temperature by 1 degree Celsius increases the dominance of organic content in soil texture by 0.19 in average. The behavior of pH varies in grasslands, having a positive impact, while exhibiting negative effects in other environments. The interpretation of the dominances of chemical elements in the chemical element composition is similar to Table 2 to 4, but since more parameters are estimated, fewer coefficients are significant, and slight variations to previous results exist because other external effects are considered in the model as well. Generally, Al, Ca, P have positive effects, while Fe, Na (both not significant), K, but also Mn and Ti have negative effects.

5. Discussion

5.1. General considerations

The application of compositional data analysis in soil science remains limited, despite its relevance for understanding soil organic matter and soil texture relationships. Our study addresses this gap by providing a compositional regression framework that jointly models chemical compositions and soil texture in a statistically coherent manner. While previous studies (Keskinen et al., 2022; Tarvainen et al., 2019) applied compositional methods for chemical element compositions, they did not incorporate soil texture composition, which plays a critical role in understanding soil organic matter distribution.

To facilitate comparisons, we provide supplementary figures and tables showcasing non-compositional analyses alongside our compositional approach. This enables a direct evaluation of potential biases that arise when compositional constraints are ignored, reinforcing the importance of log-ratio transformations for valid statistical inference in soil science. The inclusion of compositional methods in soil studies is necessary to improve accuracy and to better interpret the role of organic matter within different soil matrices.

5.2. Extension of existing methods and software

Robust estimation techniques have rarely been applied in compositional regression modeling, despite their known advantages in reducing the influence of outliers, which are commonly encountered in soil data. To address this gap, we present a significant methodological advancement by extending the framework originally proposed by Hron and Thompson (2012), which allowed robust regression modeling for compositional data but did not accommodate simultaneous inclusion of compositional and non-compositional explanatory variables. Our extended approach, detailed in Section 3.5, now enables researchers to seamlessly incorporate both external (non-compositional) variables, such as climatic or geographic information, and compositional predictors, such as soil texture or chemical element compositions, into a unified robust regression model.

This methodological extension is implemented in the well-established R package `robCompositions`, ensuring practical accessibility to soil scientists, geochemists, and researchers from various fields dealing with similar complex data structures. The extended software capabilities enhance flexibility and facilitate comprehensive modeling and interpretation, thus broadening the potential applications in compositional data analysis and significantly contributing to reproducible research practices.

5.3. Univariate and basic explorative analysis

It is important not to study one compositional part independently from other parts of the composition, and this begins already by showing, for example, a map of one compositional part, discussed already in McKinley et al. (2016). In our case, the organic content is not independent of the other parts of the soil texture. Fig. 2 shows the distribution of the relative organic content of the whole composition.

The literature shows only the soil organic matter content without considering the compositional nature of the soil texture components. Often parallel boxplots are used to show the organic content for different land use, for example Sutfin et al. (2021) or Richardson et al. (2023) and Cammarano et al. for different agricultural used land sites, Liu et al. for different kinds of soil types, or Cha et al. (2020), Bogunovic et al. and Rehm (2018) for different land uses (forest, grasslands, afforestation sites, cropland. The distributions of *average* organic content in the whole soil texture in Fig. 1 differ considerably compared to Cha et al. (2020), Bogunovic et al., Rehm (2018) and also to the non-compositional versions (original value and log10 transformed values) in the supplementary file, because we analyze the relative organic content in relation to and not independently of the content of sand, silt and clay. The relative organic content in grassland and forest land use soil samples is about the same, while they differ considerably without taking into account the compositional nature of the soil texture, which is in contradiction to the literature given above.

5.4. Organic content versus bulk density and organic content in the soil texture

Our results, indicating that an increase in the proportion of silt in soil texture leads to higher bulk densities, contradict numerous other studies, for example, Chaudhari et al. (2018), while some studies do not observe a notable positive relationship between silt content and bulk density (Özdemir et al., 2022). Differences from our results might relate to the fact that we obtained unbiased results from a compositional treatment of the data while the results from others did not consider silt to be part of a whole composition; i.e., if the silt concentration increases, automatically the concentration on sand, clay and organic matter must decrease.

Our results show that the relationship of soil organic matter content is weak with other soil textures (sand, silt, clay). This is in line with some of the literature, e.g., Costa et al. (2013), who used non-compositional treatment of data, but note that the dependencies and correlations differ from a compositional and non-compositional analysis. Other literature showed that, e.g., silt and SOMC are strongly correlated, but the results might be considerably biased because the compositional nature of the soil texture did not account for. Also, the linear relationship of SOMC to sand and silt is considered quite high in the literature, and it is the case with our data without a compositional regression approach, e.g., Thabit et al. (2023) obtained an R^2 value of 0.85, biased because of interrelationships resulting from the measurement process (part of a whole), while when treating the soil texture as composition and fitting the model with compositional regression, the R^2 reduces to about 0.2.

5.5. Organic content versus main chemical elements

Showing the concentrations of chemical elements one should be again aware that these are compositions. While many authors report raw numbers (such as the arithmetic mean of concentrations), e.g. Jamil Maia et al. (2022), unbiased results are only obtained when considering the compositional nature of a composition (Filzmoser et al., 2018). Observations in forests, fields, and permanent grasslands reveal distinct centered log-ratio coefficients for the main chemical elements. This indicates that the impact of land use on organic content cannot be overlooked. Additionally, the relative content of the chemical elements varies depending on the relative organic content in the soil texture, for example, having lower values on Ti, Si and Na in soil samples from croplands and grasslands when the relative organic content is high. Note that there are considerable differences between the compositional results and results obtained with original or log-transformed SOMC values (cf. Fig. 3 and corresponding figures in the supplementary material).

Principal component analysis and the corresponding biplots are a standard method to investigate multivariate relationships of chemical elements, and it is well known to use compositional principal component analysis for this (Filzmoser et al., 2009). Ignoring the compositional nature of the data can lead to biased correlations and loadings that direct only into a half-space when visualized in a biplot. Examples can be seen in Clunes et al. (2022), where all loadings show in almost the same direction, and Templ and Gonzalez-Rodriguez (2024b), who illustrate the differences between compositional and non-compositional principal component analysis. This distinction is also reflected in our results: Fig. 4 (compositional PCA) differs considerably from the non-compositional versions presented in the supplementary material.

In our compositional biplots, we highlight the scores regarding low, middle, and high organic content, after applying a centered log-ratio transformation to the soil texture components. This extended multivariate view reveals clear relationships between chemical elements and relative organic content. Summarizing Fig. 4, for grasslands, Sodium (Na), Potassium (K), and Aluminum (Al) are oriented towards low organic content. The reasons for this pattern are complex and may involve multiple mechanisms.

One possible explanation is that Na and K are relatively mobile elements in soils. Sodium is often present in mineral forms such as sodium feldspar and sodium-bearing clays. In soils with lower organic matter content, there may be less organic material available to complex with and retain sodium ions, allowing them to remain more abundant in the soil mineral fraction. Similarly, potassium is primarily present in soil minerals and exchangeable forms, and its retention depends on both cation exchange capacity (CEC) and mineral weathering. The reviewer raises an important point about the role of leaching, as precipitation can flush Na and K from the topsoil. However, compositional PCA accounts for the relative nature of these elements, meaning that while absolute concentrations might decrease, their relative dominance compared to other elements in the soil matrix may still be evident in the compositional structure. This suggests that soils with low SOMC may retain relatively higher mineral-bound Na and K, despite potential aqueous-phase leaching.

Aluminum, a major constituent of soil minerals, follows a similar pattern. In low-organic-matter soils, there are fewer organic ligands available to bind with Al, meaning that Al may remain relatively enriched in its mineral-bound form. This observation aligns with previous studies that have reported a stronger association of Al with mineral surfaces rather than organic complexes in soils with low organic content. On the contrary, calcium (Ca) is strongly associated with high relative organic content, which is in line with Shabtai et al. (2023), who investigated the relationship between Ca and SOMC. Higher organic matter content enhances cation retention and complexation, preventing leaching of calcium ions. In contrast to Na and K, which may be more susceptible to loss through leaching, Ca is often more effectively stabilized in organic-rich soils through complexation with organic acids and microbial activity. The higher biological activity in organic-rich soils may also enhance mineral weathering processes, making Ca more bioavailable and more likely to be incorporated into the organic matter fraction.

Our compositional biplots provide a multivariate perspective that differs from traditional non-compositional analyses. While previous studies, such as Rowley et al. (2023), reported strong correlations between Ca and SOMC, their results may be biased due to the non-compositional treatment of soil data, which can lead to artificially inflated correlations (Templ and Templ, 2021). We emphasize that compositional PCA provides a structurally different interpretation, as it accounts for relative changes within the soil matrix rather than treating individual element concentrations as independent variables.

Note that in our regression models, the explanatory variables are represented as pivot log-ratio coordinates, meaning that the coefficients reflect the dominance of each element relative to the geometric mean of the other chemical elements, rather than their absolute concentrations. The observed negative regression coefficients for Si and Ti indicate that as these elements become more dominant in the chemical composition, soil organic matter content (SOMC) tends to decrease. This does not necessarily imply that Si and Ti themselves directly reduce SOMC but rather that soils where these elements are relatively more abundant tend to have lower organic matter content.

This pattern may be linked to mineralogical and geochemical differences. Higher relative contributions of Si and Ti often indicate mineral-dominated or weathered soils, which typically contain fewer organic components. Additionally, Ti is an immobile element, and its relative enrichment can reflect nutrient leaching processes, where more mobile elements associated with organic matter have been depleted over time. Similarly, Si is a major component of quartz-rich sandy soils, which generally have lower organic matter retention due to their lower capacity for organic matter stabilization. Thus, the negative relationship between the dominance of Si and Ti and SOMC likely reflects broader geochemical and soil development processes, rather than direct biological effects.

5.6. Organic content in light of environmental parameters, bulk density, chemical elements and remaining soil texture components

The effect of land-use on the content of soil organic matters was already discussed in Masoudi et al. (2023) but for different agricultural tillage practices, and for forest, grassland, farmland and unused land in Li et al. (2021). Generally, we also observe that the soil organic matter content depends on land use, by which we investigated the relative contribution of soil organic matter to the entire fraction of the material of sand, silt, clay and soil organic matter. When doing so, the results change and are different from Li et al. (2021) and also different from the results of our data when analyzing the absolute values. Our results are different because they are no longer biased, taking into account the compositional nature of the composition. For example, the soil organic matter content is about the same in permanent grasslands and forests, but very different in Li et al. (2021).

The SOMC measured by hot water extractable carbon was shown by Voltr et al. (2021) with a linear regression models that HWEC content is significantly and negatively affected – amongst others – by phosphorus content (−30%) and potassium (−7%). However, this is in contradiction with our results (the phosphorus content had a positive effect) and we have to mention that Voltr et al. (2021) did not use compositional methods, probably giving an error.

The content of soil organic matter was predicted by Lazzaretti et al. (2020) using the clay content, but negating that both are part of a composition. On the one hand, this has a serious effect, and we did not receive a good model for predicting the relative content of soil organic matter at all. Nevertheless, our used models were unbiased. In addition, Fernandes et al. (2019) employed an artificial neural network to forecast soil organic matter content, without considering the compositional nature of the compositions. However, artificial deep neural networks are not as effective and can be biased when applied to compositional data, as demonstrated in Templ (2021b), at least for all standard loss functions that rely on an Euclidean vector space.

Note that when treating the soil texture composition and the chemical element composition not as compositional, then the robust MM regression did not converge and also gave very different results (see supplementary file).

5.7. Remark on the use of different analysis methods

We observed differences in the interpretation of elemental relationships with SOMC between PCA and regression analyses, notably for calcium (Ca). In the PCA, Ca strongly associates with high SOMC, indicating its relative prominence in organic-rich soils. However, in regression analyses using pivot log-ratio coordinates, Ca did not exhibit significant effects consistently. This discrepancy arises primarily due to methodological differences: PCA captures overall co-variations among elements and organic matter content, reflecting broader multivariate patterns. In contrast, regression analyses estimate conditional effects of each element on SOMC, controlling explicitly for the presence of other elements. Thus, even though Ca prominently co-occurs with higher SOMC in multivariate space, its conditional effect, when controlling for other soil components, may not be statistically robust.

Similarly, the apparent reversal in relationships for other elements such as aluminum (Al) can be attributed to these methodological differences. While PCA reflects relative abundance and compositional dominance, regression analyses focus on specific controlled influences of each element on SOMC. Hence, we recommend that readers interpret PCA findings as indicative of multivariate compositional patterns, whereas regression results offer insights into the direct conditional impact of each element.

Therefore, we caution against overinterpreting individual PCA patterns if regression analysis indicates weak or non-significant conditional effects. Instead, consistent findings across methods—such as the negative conditional impact of bulk density or the consistent relative association of phosphorus (P) with higher SOMC—provide robust interpretations. Future studies should explicitly report these methodological differences to guide clear interpretation of compositional data analyses.

5.8. Future work related to pH content

While our analysis has primarily focused on compositional relationships in soil organic matter content (SOMC), future research could extend this approach to investigate the role of soil pH, particularly in low-pH environments where soil organic carbon (SOC) is known to be less soluble. Lower pH levels can influence SOC retention through multiple mechanisms, including reduced microbial decomposition rates and increased interactions with mineral surfaces. For example, Clarholm and Skyllberg (2013) demonstrated that in acidic forest soils, low pH increases the sorption of dissolved organic matter to mineral surfaces, enhancing SOM stabilization. The influence of pH on microbial decomposition rates has been further elucidated by Malik et al. (2018), who found that soil pH is a key driver of microbial community composition and function, thereby indirectly affecting SOM turnover rates (Malik et al., 2018). Their study emphasized the need for incorporating pH-dependent microbial processes in Earth system models to improve predictions of soil carbon dynamics.

Since both soil organic carbon content and mineral interactions exhibit relative constraints within the broader soil system, a compositional data analysis approach would be well suited to disentangle these dependencies. By applying log-ratio transformations, it becomes possible to identify how the relative balance of SOM fractions, mineral-associated organic matter, and chemical element compositions shifts across different pH levels. This would allow for a more robust quantification of pH effects while avoiding spurious correlations that arise in standard statistical analyses due to the closed-sum constraint of soil constituents.

However, such an analysis presents certain challenges. In many datasets, soils with extremely low pH values tend to be underrepresented, leading to smaller sample sizes and potential issues with statistical power. Additionally, interactions between pH, organic matter, and chemical elements may introduce further complexity. Future work should consider targeted sampling strategies or the integration of larger datasets to ensure meaningful comparisons across different pH regimes. Addressing these challenges would enhance our ability to generalize compositional trends across a wider range of soil conditions.

5.9. Limitations

While our results reveal consistent relationships between SOMC and certain chemical elements in our study area, we acknowledge that these relationships may not be universally consistent across different regions due to variations in parent material, climate, land-use history, and soil formation processes. Differences in mineralogical composition can strongly influence the availability and mobility of elements, while climatic factors such as precipitation and temperature affect organic matter decomposition rates, leaching processes, and cation exchange dynamics. As a result, the observed associations between SOMC and chemical elements may not necessarily hold under different environmental and geological conditions. Future research should consider comparative studies across diverse soil systems, incorporating regional differences in soil genesis and geochemistry to better understand how these factors shape the compositional relationships of soil organic matter''.

6. Conclusion

It is well established in the literature on compositional data analysis that ignoring the compositional structure of soil texture and chemical elements leads to biased estimates, as standard statistical methods do not account for the constrained nature of such data (Filzmoser et al., 2018; Templ and Templ, 2021). Our findings reinforce this by demonstrating substantial differences between compositional and non-compositional analyses, as illustrated in the supplementary material. Additionally, as soil data often contain outliers that could distort results, our study applied robust statistical estimation techniques to mitigate their impact.

One of the key contributions of this work is extending the regression framework proposed by Hron and Thompson (2012) to include both compositional and external explanatory variables, addressing a crucial gap in practical applications. Many real-world analyses require incorporating environmental factors such as temperature, altitude, precipitation, pH, and bulk density alongside compositional variables. This advancement enhances the flexibility of compositional regression and is now freely available through the R package robCompositions (Templ et al. 2011), providing a practical tool for researchers.

While machine learning (ML) techniques are increasingly applied in soil science, their direct use with compositional data remains problematic. Many ML approaches implicitly assume Euclidean data structures and fail to account for the relative nature of compositional data, leading to potential biases. Our study highlights the importance of applying transformations that properly map compositional data into real space before integrating them into ML models. This ensures that methods such as deep learning do not suffer from misinterpretations due to incompatibility with the compositional sample space (Templ, 2021a,c). By embedding compositional principles within the modeling process, our work provides a more statistically coherent foundation for future applications in ML-based soil analysis.

Beyond methodological advancements, our study contributes to a deeper understanding of soil organic matter and its interaction with chemical and physical soil properties. The findings support previous work emphasizing the role of calcium in organic matter retention, while compositional PCA offers a more structured view of multivariate relationships compared to traditional correlation-based approaches. Future research should further explore compositional ML methods and apply our approach across different regions to assess the robustness of the observed relationships under varying geological and climatic conditions.

By addressing key methodological challenges and expanding the applicability of compositional regression, our study enhances statistical modeling in soil science. It provides a foundation for more robust, compositionally aware analyses that can guide both traditional statistical approaches and the integration of ML techniques in a more rigorous manner.

CRediT authorship contribution statement

Matthias Templ: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Christoph Hofer:** Visualization, Resources, Investigation, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

We express our gratitude to the Soil Protection Agency of the Canton of Zurich, for sharing the data and engaging in discussions with us.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.apgeochem.2025.106526.

Data availability

The authors do not have permission to share data.

References

Agroscope, 2020. Schweizerische Referenzmethoden der Forschungsanstalten Agroscope: pH-Wert einer $CaCl_2$-Suspension (1:2.5). Agroscope, Zürich, http://www.agroscope.ch/referenzmethoden.

Aitchison, J., 1986. The Statistical Analysis of Compositional Data. Chapman & Hall, London, http://dx.doi.org/10.1007/978-94-009-4109-0.

Bogunovic, I., Viduka, A., Magdic, I., Telak, L.J., Francos, M., Pereira, P., Agricultural and forest land-use impact on soil properties in Zagreb periurban area (Croatia). Agronomy 10, 2020. http://dx.doi.org/10.3390/agronomy10091331.

Bot, A., Benites, J., 2005. The importance of soil organic matter: key to drought-resistant soil and sustained food production, volume 80, Food and Agriculture Organization of the United Nations, Bibliografija po poglavjih.

Brimhall, G.H., Dietrich, W.E., 1987. Constitutive mass balance relations between chemical composition, volume, density, porosity, and strain in metasomatic hydrochemical systems: Results on weathering and pedogenesis. Geochim. Cosmochim. Acta 51, 567–587. http://dx.doi.org/10.1016/0016-7037(87)90070-6, https://www.sciencedirect.com/science/article/pii/0016703787900706.

Cammarano, D., Holland, J., Ronga, D., Spatial and temporal variability of spring barley yield and quality quantified by crop simulation model. Agronomy 10, 2020. http://dx.doi.org/10.3390/agronomy10030393.

Cha, S., Kim, C.-B., Kim, J., Lee, A., Park, K.-H., Koo, N., Kim, Y., 2020. Land-use changes and practical application of the land degradation neutrality (ldn) indicators: a case study in the subalpine forest ecosystems, republic of korea. For. Sci. Technol. 16, 8–17. http://dx.doi.org/10.1080/21580103.2019.1700831.

Chaudhari, P., Ahire, D., Ahire, V., Chkravarty, M., Maity, S., 2018. Soil bulk density as related to soil texture, organic matter content and available total nutrients of coimbatore soil. Int J Sci Res Publ 2, 1–8.

Chen, C., Hall, S.J., Coward, E., Thompson, A., 2020. Iron-mediated organic matter decomposition in humid soils can counteract protection. Nat. Commun. 11 (2255), http://dx.doi.org/10.1038/s41467-020-16071-5.

Clarholm, M., Skyllberg, U., 2013. Translocation of metals by trees and fungi regulates ph, soil organic matter turnover and nitrogen availability in acidic forest soils. Soil Biol. Biochem. 63, 142–153.

Clunes, J., Dörner, J., Zúñiga Ugalde, F., Bravo, A., Jara, R., 2022. Did we underestimate silt and clay content in the textural analysis? Chil. J. Agric. Anim. Sci. 38, 94–103. http://dx.doi.org/10.29393/CHJAAS38-9DWJF50009.

Coenders, G., Pawlowsky-Glahn, V., 2020. On interpretations of tests and effect sizes in regression models with a compositional predictor. SORT Stati. Oper. Res. Trans. 44, 201–220. http://dx.doi.org/10.2436/20.8080.02.100.

Costa, A., Albuquerque, J., Costa, A., Pértile, P., Silva, F., 2013. Water retention and availability in soils of the state of santa catarina-brazil: effect of textural classes, soil classes and lithology. Rev. Bras. de Cienc. Do Solo 37, 1535–1548. http://dx.doi.org/10.1590/S0100-06832013000600010.

Dai, G., Zhu, E., Liu, Z., Wang, Y., Zhu, S., Wang, S., Ma, T., Jia, J., Wang, X., Hou, S., Fu, P., Peterse, F., Feng, X., 2019. Compositional characteristics of fluvial particulate organic matter exported from the world's largest alpine wetland. J. Geophys. Res.: Biogeosciences 124, 2709–2727. http://dx.doi.org/10.1029/2019JG005231.

Department of Agriculture, U.S., 1999. Soil taxonomy: A basic system of soil classification for making and interpreting soil surveys. In: Agricultural Handbook, vol. 436, U.S. Government Printing Office, Washington, DC.

Dou, X., Zhang, J., Zhang, C., Ma, D., Chen, L., Zhou, G., Li, J., Duan, Y., 2023. Calcium carbonate regulates soil organic carbon accumulation by mediating microbial communities in northern china. Catena 231, http://dx.doi.org/10.1016/j.catena.2023.107327.

Egozcue, J.J., Pawlowsky-Glahn, V., 2018. Modelling compositional data. the sample space approach. In: Sagar, B. Daya, Cheng, Q., Agterberg, F. (Eds.), Handbook of Mathematical Geosciences: Fifty Years of IAMG. Springer International Publishing, Cham, pp. 81–103. http://dx.doi.org/10.1007/978-3-319-78999-6_4.

Egozcue, J.J., Pawlowsky-Glahn, V., Mateu-Figueras, G., Barceló-Vidal, C., 2003. Isometric logratio transformations for compositional data analysis. Math. Geol. 35, 279–300. http://dx.doi.org/10.1023/A:1023818214614.

Fernandes, M., Coelho, A., Fernandes, C., da Silva, M., Dela Marta, C., 2019. Estimation of soil organic matter content by modeling with artificial neural networks. Geoderma 350, 46–51. http://dx.doi.org/10.1016/j.geoderma.2019.04.044.

Filzmoser, P., Hron, K., 2019. Comments on: Compositional data: the sample space and its structure. TEST 28, 639–643. http://dx.doi.org/10.1007/s11749-019-00671-5.

Filzmoser, P., Hron, K., Reimann, C., 2009. Principal component analysis for compositional data with outliers. Environmetrics 20, 621–632. http://dx.doi.org/10.1002/env.966.

Filzmoser, P., Hron, K., Templ, M., 2018. Applied compositional data analysis. With worked examples in r. In: Springer Series in Statistics. Springer Publishing, Cham, Switzerland, http://dx.doi.org/10.1007/978-3-319-96422-5.

Fišerová, E., Hron, K., 2011. On interpretation of orthonormal coordinates for compositional data. Math. Geosci. 43, 455–468.

Grzebisz, W., Szczepaniak, W., Potarzycki, J., ukowiak, R.Ł., 2012. Sustainable management of soil potassium – a crop rotation oriented concept. In: Issaka, R.N. (Ed.), Soil Fertility. IntechOpen, Rijeka, http://dx.doi.org/10.5772/53185.

Hoque, T.S., Sarkar, D., Datta, R., Kibria, M.G., Ullah, R., Ahmed, N., Hossain, M.A., Masood, A., Anjum, N.A., 2023. Sustainable management of phosphorus in agriculture for environmental conservation. In: Anjum, A.P.N.A., Masood, D.A., Umar, D.S., Khan, D.N.A. (Eds.), Phosphorus in Soils and Plants. IntechOpen, Rijeka, http://dx.doi.org/10.5772/intechopen.113086.

Hron, P.F.K., Thompson, K., 2012. Linear regression with compositional explanatory variables. J. Appl. Stat. 39, 1115–1128. http://dx.doi.org/10.1080/02664763.2011.644268.

Jamil Maia, A., Cabral Nascimento, R., Jacques Agra Bezerra da Silva, Y., Williams Araújo do Nascimento, C., de Sousa Mendes, W., Germano Veras Neto, J., Coelho de Araújo Filho, J., Tiecher, T., Jacques Agra Bezerra da Silva, Y., 2022. Near-infrared spectroscopy for prediction of potentially toxic elements in soil and sediments from a semiarid and coastal humid tropical transitional river basin. Microchem. J. 179, 107544. http://dx.doi.org/10.1016/j.microc.2022.107544.

Keskinen, R., Hillier, S., Liski, E., Nuutinen, V., Nyambura, M., Tiljander, M., 2022. Mineral composition and its relations to readily available element concentrations in cultivated soils of finland. Acta Agric. Sc Andin. Sect. B Soil Plant Sci. 72, 751–760. http://dx.doi.org/10.1080/09064710.2022.2075790.

Lazzaretti, B., da Silva, L. Souza, Drescher, G., Dotto, A., Britzke, D., Nörnberg, J., 2020. Prediction of soil organic matter and clay contents by near-infrared spectroscopy - nirs. CiÊncia Rural. 50, http://dx.doi.org/10.1590/0103-8478cr20190506.

Li, X., Jin, X., Song, Y., 2021. Effects of different land use patterns on soil organic matter and soil cycle process. In: IOP Conference Series: Earth and Environmental Science. vol. 804, 022101. http://dx.doi.org/10.1088/1755-1315/804/2/022101.

Liu, Y., Shi, Z., Zhang, G., Chen, Y., Li, S., Hong, Y., Shi, T., Wang, J., Liu, Y., Application of spectrally derived soil type as ancillary data to improve the estimation of soil organic carbon by using the chinese soil vis-nir spectral library. Remote. Sens. 10, 2018. http://dx.doi.org/10.3390/rs10111747.

Lyu, S., Wei, X., Chen, J., Wang, C., Wang, X., Pan, D., Titanium as a beneficial element for crop production. Front. Plant Sci. 8, 2017. http://dx.doi.org/10.3389/fpls.2017.00597.

Malik, A.A., Puissant, J., Buckeridge, K.M., Goodall, T., Jehmlich, N., Chowdhury, S., Gweon, H.S., Peyton, J.M., Mason, K.E., van Agtmaal, M., et al., 2018. Land use driven change in soil ph affects microbial carbon cycling processes. Nat. Commun. 9, 3591.

Maronna, R., Martin, R., Yohai, V., 2006. Robust Statistics: Theory and Methods. John Wiley & Sons, New York, ISBN 978-0-470-01092-1.

Masoudi, M., Centeri, C., Karlik, M., Jakab, G., 2023. Extracted samples and in situ soil investigations to assess the effects of different land use and tillage management practices on soil organic matter composition. Land Degrad. Dev. 34, 2560–2572. http://dx.doi.org/10.1002/ldr.4629.

Matus, F.J., 2021. Fine silt and clay content is the main factor defining maximal c and n accumulations in soils: a meta-analysis. Sci. Rep. 11, 6438. http://dx.doi.org/10.1038/s41598-021-84821-6.

McKinley, J.M., Hron, K., Grunsky, E.C., Reimann, C., de Caritat, P., Filzmoser, P., van den Boogaart, K.G., Tolosana-Delgado, R., 2016. The single component geochemical map: Fact or fiction? J. Geochem. Explor. 162, 16–28. http://dx.doi.org/10.1016/j.gexplo.2015.12.005.

Özdemir, N., Demir, Z., Bülbül, E., 2022. Relationships between some soil properties and bulk density under different land use. Soil Stud. 11, 43–50. http://dx.doi.org/10.21657/soilst.1218353.

Pavlovic, J., Kostic, L., Bosnic, P., Kirkby, E.A., Nikolic, M., 2021. Interactions of silicon with essential and beneficial elements in plants. Front. Plant Sci. 12, http://dx.doi.org/10.3389/fpls.2021.697592.

Pawlowsky-Glahn, V., Egozcue, J.J., 2006. Compositional Data and their Analysis: An Introduction. vol. 264, (1–10), Geological Society, London, Special Publications, https://api.semanticscholar.org/CorpusID:130177935.

Pawlowsky-Glahn, V., Egozcue, J., Tolosana-Delgado, J., 2007. Lecture notes on compositional data analysis, http://www.sediment.uni-goettingen.de/staff/tolosana/extra/CoDa.pdf. available online.

Pearson, K., 1897. Mathematical contributions to the theory of evolution. on a form of spurious correlation which may arise when indices are used in the measurement of organs. In: Proceedings of the Royal Society of London, vol. 60, pp. 489–502.

Poggio, L., de Sousa, L., Batjes, N., Heuvelink, G., Kempen, B., Ribeiro, E., Rossiter, D., 2021. SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. Soil 7, 217–240, https://www.proquest.com/scholarly-journals/soilgrids-2-0-producing-soil-information-globe/docview/2540446093/se-2. copyright -©2021. This work is published under https://creativecommons.org/licenses/by/4.0/ (the License). Notwithstanding the ProQuest Terms and Conditions, you may use this content in accordance with the terms of the License; Zuletzt aktualisiert - 2025-01-06; SubjectsTermNotLitGenreText - Environmental.

Rehm, R., 2018. Small scale variability in soil hydraulic properties in a headwater catchment of the Indian western ghats geographien Südasiens. http://dx.doi.org/10.11588/xabooks.389.

Richardson, G.S., Ruark, M.D., Radatz, T., Radatz, A., Cooley, E., Silva, E.M., Augarten, A.J., Zhu, J., Zegler, C.H., 2023. The influence of inherent soil factors and agricultural management on soil organic matter. Ecosphere 14, e4459. http://dx.doi.org/10.1002/ecs2.4459.

Rowley, M.C., Nico, P.S., Bone, S.E., Marcus, M.A., Pegoraro, E.F., Castanha, C., Kang, K., Bhattacharyya, A., Torn, M.S., Peña, J., 2023. Association between soil organic carbon and calcium in acidic grassland soils from point reyes national seashore, ca. Biogeochemistry 165, 91–111. http://dx.doi.org/10.1007/s10533-023-01059-2.

Scheel, T., Jansen, B., Van Wijk, A.J., Verstraten, J.M., Kalbitz, K., 2008. Stabilization of dissolved organic matter by aluminium: a toxic effect or stabilization through precipitation? Eur. J. Soil Sci. 59, 1122–1132. http://dx.doi.org/10.1111/j.1365-2389.2008.01074.x.

Shabtai, I.A., Wilhelm, R.C., Schweizer, S.A., Höschen, C., Buckley, D.H., Lehmann, J., 2023. Calcium promotes persistent soil organic matter by altering microbial transformation of plant litter. Nat. Commun. 14, http://dx.doi.org/10.1038/s41467-023-42291-6, 6609.

Sheng, H., Chen, S., 2020. Plant silicon-cell wall complexes: Identification, model of covalent bond formation and biofunction. Plant Physiol. Biochem. 155, 13–19. http://dx.doi.org/10.1016/j.plaphy.2020.07.020.

Soil Protection Agency of the Canton of Zurich, 2021. Personal communication. Database query on September 30 2021.

Sokolov, D.A., Dmitrevskaya, I.I., Pautova, N.B., Lebedeva, T.N., Chernikov, V.A., Semenov, V.M., 2021. A study of soil organic matter stability using derivatography and long-term incubation methods. Eurasian Soil Sci. 54, 487–498. http://dx.doi.org/10.1134/S1064229321040141.

Solly, E.F., Weber, V., Zimmermann, S., Walthert, L., Hagedorn, F., Schmidt, M.W.I., 2020. A critical evaluation of the relationship between the effective cation exchange capacity and soil organic carbon content in swiss forest soils. Front. For. Glob. Chang. 3.

Sparks, D., Singh, B., Siebecker, M., 2023. Environmental Soil Chemistry. Elsevier Science.

Sutfin, N.A., Wohl, E., Fegel, T., Day, N., Lynch, L., 2021. Logjams and channel morphology influence sediment storage, transformation of organic matter, and carbon storage within mountain stream corridors. Water Resour. Res. 57, e2020. http://dx.doi.org/10.1029/2020WR028046, WR028046.

Tang, S., She, D., Wang, H., 2021. Effect of salinity on soil structure and soil hydraulic characteristics. Can. J. Soil Sci. 101, 62–73. http://dx.doi.org/10.1139/cjss-2020-0018.

Tarvainen, T., Sapon, S., Jarva, J., 2019. Applying heatmaps in interpretation of geochemical baseline data on urban soils in finland. J. Geochem. Explor. 205, 106345. http://dx.doi.org/10.1016/j.gexplo.2019.106345.

Templ, M., 2021a. Artificial neural networks to impute rounded zeros in compositional data. In: Filzmoser, P., Hron, K., Martin-Fernandez, J., Palarea-Albaladejo, J. (Eds.), Advances in Compositional Data Analysis: Festschrift in Honour of Vera Pawlowsky-Glahn. Springer International Publishing, Cham, pp. 163–187.

Templ, M., 2021b. Artificial neural networks to impute rounded zeros in compositional data. In: Filzmoser, P., Hron, K., Martín-Fernández, J.A., Palarea-Albaladejo, J. (Eds.), Advances in Compositional Data Analysis: Festschrift in Honour of Vera Pawlowsky-Glahn. Springer International Publishing, Cham, pp. 163–187.

Templ, M., 2021c. Can we ignore the compositional nature of compositional data by using deep learning aproaches?. In: C. Perna, F. Schirripa Spagnolo (Ed.), Book of Short Papers SIS 2021. Pearson, London, pp. 243–248, Surface Inspection Summit Europe, Aachen, Germany, 9-10 2021.

Templ, M., Gonzalez-Rodriguez, J., 2024a. Advancing forensic research: An examination of compositional data analysis with an application on petrol fraud detection. Sci. & Justice 64, 9–18. http://dx.doi.org/10.1016/j.scijus.2023.11.003.

Templ, M., Gonzalez-Rodriguez, J., 2024b. Advancing forensic research: An examination of compositional data analysis with an application on petrol fraud detection. Sci. Justice 64, 9–18. http://dx.doi.org/10.1016/j.scijus.2023.11.003.

Templ, M., Hron, K., Filzmoser, P ., 2011. robCompositions: An R-Package for Robust Statistical Analysis of Compositional Data. John Wiley & Sons, Ltd, pp. 341–355. http://dx.doi.org/10.1002/9781119976462.ch25.

Templ, M., Templ, B., 2020. Analysis of chemical compounds in beverages- guidance for establishing a compositional analysis. Food Chem. 325, 1–7. http://dx.doi.org/10.1016/j.foodchem.2020.126755.

Templ, M., Templ, B., 2021. Statistical analysis of chemical element compositions in food science: problems and possibilities. Molecules 26, http://dx.doi.org/10.3390/molecules26195752.

Thabit, F.N., El-Shater, A.-H., Soliman, W., 2023. Role of silt and clay fractions in organic carbon and nitrogen stabilization in soils of some old fruit orchards in the nile floodplain. J. Soil Sci. Plant Nutr. 23, 2525–2544. http://dx.doi.org/10.1007/s42729-023-01209-3.

Voltr, V., Menšík, L., Hlisnikovský, L., Hruška, M., Pokorný, E., Pospíšilová, L., 2021. The soil organic matter in connection with soil properties and soil inputs. Agronomy 11, http://dx.doi.org/10.3390/agronomy11040779.

Wu, H., Cui, H., Fu, C., Li, R., Qi, F., Liu, Z., Yang, G., Xiao, K., Qiao, M., 2024. Unveiling the crucial role of soil microorganisms in carbon cycling: A review. Sci. Total Environ. 909, 168627. http://dx.doi.org/10.1016/j.scitotenv.2023.168627.

Xu, Z., Tsang, D.C., 2024. Mineral-Mediated Stability of Organic Carbon in Soil and Relevant Interaction Mechanisms. Eco-Environment & Health, http://dx.doi.org/10.1016/j.eehl.2023.12.003.

Yuan, Y., Li, B., Yu, W., Gao, X., 2021. Estimation and mapping of soil organic matter content at a national scale based on grid soil samples, a soil map and dem data. Ecol. Inform. 66, 101487. http://dx.doi.org/10.1016/j.ecoinf.2021.101487.

Zhang, X., Liu, S., Li, J., Wu, Y., Luo, H., Jiang, Z., Huang, X., 2022. Nutrient enrichment decreases dissolved organic carbon sequestration potential of tropical seagrass meadows by mediating bacterial activity. Ecol. Indic. 145, http://dx.doi.org/10.1016/j.ecolind.2022.109576, 109576.

Zuazagoitia, D., Villarroel, J., 2016. Studying the importance of soil organic matter: An educational proposal for secondary education. Educación Química 27, 37–42. http://dx.doi.org/10.1016/j.eq.2015.09.007.