

# Investigation of the relationships between selenium fractions and soil properties by canonical correlation

Abdullah Özkan<sup>1</sup>, Abdullah Nuri Özsoy<sup>2</sup>, Veli Uygur<sup>3\*</sup>

<sup>1</sup>Department of Biomedical Engineering, Faculty of Engineering and Natural Sciences, Iskenderun Technical University, 31200 Hatay, Türkiye

<sup>2</sup>Department of Animal Sciences, Agricultural Faculty, Isparta University of Applied Sciences, 32200 Isparta, Türkiye

<sup>3</sup> Department of Soil Science and Plant Nutrition, Agricultural Faculty, Isparta University of Applied Sciences, 32200 Isparta, Türkiye

## How to cite

Ozkan, A., Ozsoy, A.N. & Uygur, V. (2023). Investigation of the relationships between selenium fractions and soil properties by canonical correlation. *Soil Studies*, 12(2), 92-101. <http://doi.org/10.21657/soilst.1408065>

## Article History

Received 20 October 2023

Accepted 08 November 2023

First Online 21 December 2023

## \*Corresponding Author

Tel.: +90 (246) 214 60 00

E-mail: [veliuygur@isparta.edu.tr](mailto:veliuygur@isparta.edu.tr)

## Keywords

Amik Plain

Soil properties

Selenium

Sequential extraction

Canonical correlation

## Abstract

The availability and distribution of an element are determined by soil conditions and the geochemical fractions. In this study, canonical correlation (CCA) was used to examine the associations between the soil parameters (X set) of the Amik plain and the fractions of selenium (Se) (Y set). A statistically significant correlation ( $P < 0.01$ ) of  $0.391 \pm 0.06$  was observed between the first canonical variable pair (FCVP). The clay, sand, and silt variables exhibited the greatest standardized canonical coefficients, with values of 74.55, 75.83, and 94.77, respectively. Additionally, the B3 percent had a standardized canonical coefficient of 1.029. The analysis of the FCVP showed that the original variables of the X set and/or its canonical variable (U1) had the strongest linear connection (0.770) with organic matter (OM) and the B3 percentage in the Y set (0.603). The OM (0.3012) exhibited the most substantial linear structural correlations between the canonical variables of the Y set and the original variables of the X set. However, the canonical redundancy analysis revealed that in the FCVP, the variance of the X set was described by its canonical variables to the extent of 20.97%, whereas the Y set was explained by 21.96%. Hence, it can be inferred that the use of canonical correlation is effective in elucidating the relationship between soil parameters and various components of plant nutrition.

## Introduction

Selenium (Se) is one of the trace elements that are absolutely needed for humans and animals (Rayman, 2000), owing to its key roles in the defence system as an antioxidant (Tapiero et al., 2003) and its crucial functions in cancer prevention (Reid et al., 2008). Hence, it is imperative for both animals and humans to consume sufficient quantities of this essential element for the maintenance of good health. Nevertheless, it is worth noting that in several nations, the average daily selenium consumption per person

falls below the World Health Organization's suggested threshold of 40 µg (Navarro-Alarcon and Cabrera-Vique, 2008). One of the primary factors contributing to this circumstance is the inadequate uptake of this element by plants, which fails to satisfy the dietary needs of animals and people via consumption. The extent of this phenomenon is contingent upon the aggregate quantity of selenium (Se) present in the soil, as well as the chemical and geochemical proportions of various selenium species. However, it is important to

note that the characteristics, accessibility, conversions, and absorption of selenium (Se) by plants exhibit variability in response to both biotic and abiotic variables during the process of soil development, therefore being influenced by soil qualities and agricultural methodologies.

The bioavailability and geochemical fractionation of selenium in soils are influenced by many continuous processes, including sorption/desorption, precipitation-dissolution, and ligand synthesis (Alfthan et al., 2015). The overall concentration of selenium (Se) in a given soil is also influenced by the various transport mechanisms taking place inside the solid or solution phase. The transformations between different geochemical fractions in the field of interest are primarily influenced by factors such as pH and redox potential (Nakamaru and Altansuvd, 2014; Xing et al., 2015; Fan et al., 2018), organic matter (Xing et al., 2015), soil texture and mineralogy (Antoniadis et al., 2017; Munier-Lamy et al., 2007), Fe/Al oxides extractable with dithionite-citrate-bicarbonate (Araujo et al., 2019; Fan and Zhao, 2018; Hyun et al., 2006), microbial characteristics (Munier-Lamy et al., 2007), Se content (Winkel et al., 2015; Saha et al., 2017), topographic features (Xu et al., 2018), and the presence of competitive anions such as sulphates and phosphates (Lee et al., 2011). The impact of competitive anions is particularly apparent in soils that receive excessive amounts of fertilizer, particularly phosphorus. In the soil environment, the presence of inorganic Se is seen in several forms, including selenite ( $\text{Se}^{-2}$ ), elemental selenium [Se (0)], and selenate ( $\text{Se}^{4+}$ ,  $\text{Se}^{6+}$ ), which are determined by factors such as pH and redox potential. Selenium is present in well-drained soils in the forms of  $\text{SeO}_4^{2-}$ ,  $\text{SeO}_3^{2-}$ , and  $\text{HSeO}_3^-$ . In soils with neutral and alkaline pH levels, the major species of Se is  $\text{Se}^{6+}$ . This particular form of Se is often highly soluble, mobile, and readily available to plants (Mikkelsen et al., 1989). In contrast, previous studies have shown that the application of humic acid and the presence of inadequate drainage in neutral soils might induce inter-fractional transitions (Fan et al., 2019). According to many studies (Tolu et al., 2014; Eswayah et al., 2017; Li et al., 2017; Wang et al., 2019), it has been shown that the presence of organic matter in soils, either naturally rich or organically amended, leads to the reduction of hazardous Se-oxyanions into the elemental fraction that exhibits limited bioavailability. In contrast, the adsorption characteristics of Se are influenced by several factors such as the presence of different iron/aluminum oxides, organic matter, and pH levels. Nevertheless, the allocation of Se across different fractions is subject to variation, which is contingent upon the elemental form of Se and the duration of time that Se ion has been adsorbed, as shown by Fan and Zhao (2018).

The comprehension of Se concentration and geochemical fractions in the soil system cannot be

achieved by simplistic one-dimensional interactions between dependent and independent variables. The lack of clear differentiation between dependent and independent variables is typically seen while analysing the connections between nutrient availability, soil fractions, and soil attributes within the soil system. Hence, in order to conduct a more comprehensive analysis of the connections, it is advisable to use multivariate statistical methods that do not differentiate between dependent and independent variables (Shin, 1996).

Canonical correlation analysis (CCA) is a statistical technique that assesses the degree of association between two groups of variables, which may be either dependent or independent (Sharma, 1996; Ozsoy, 2019). Canonical correlation analysis is a multivariate statistical approach that has similarities with principal component analysis (PCA) and factor analysis. However, CCA distinguishes itself as the most suitable and robust method in situations when there are several sets of dependent and/or independent variables.

The objective of this research was to examine the impact of soil characteristics on the alterations in selenium (Se) fractions in soils within the Amik plain region. Since Amik plain has key properties for accumulation/leaching of Se and partitioning to different Se fraction it may be regarded as good data sets to test the relations by multivariate statistical approach. This investigation, therefore, used canonical correlation analysis, using two distinct sets of variables: soil properties and Se fractions to extract the coherence between assigned two data set.

## Materials and Methods

**Sampling and descriptive analysis:** A total of 179 soil samples were collected from the Amik plain, specifically at the intersections of designated transects of 2 x 2 kilometres. Assigned sampling coordinates and the drainage-web map were merged in ArcMAP environment (Figure 1). The soil parameters examined in this study included textural components by hydrometer method (sand, silt, and clay), pH and electrical conductivity (EC) in 1:2.5 soil: distilled water suspension, organic matter content by wet oxidation method of Walkley-Black, and calcium carbonate equivalent by a manometric method. These parameters were analysed using the conventional procedures commonly used in Türkiye, as described by Kacar (2009) and site properties of the Amik plain were given elsewhere (Özkan et al., 2022). Table 1 presents the descriptive statistics for the aforementioned attributes.

**Selenium fractionation method:** The sequential extraction technique is a method used to separate and isolate different components or fractions of a sample in a step-by-step manner. Martens and Suarez (1997) conducted sequential extraction, which was described in the following manner.

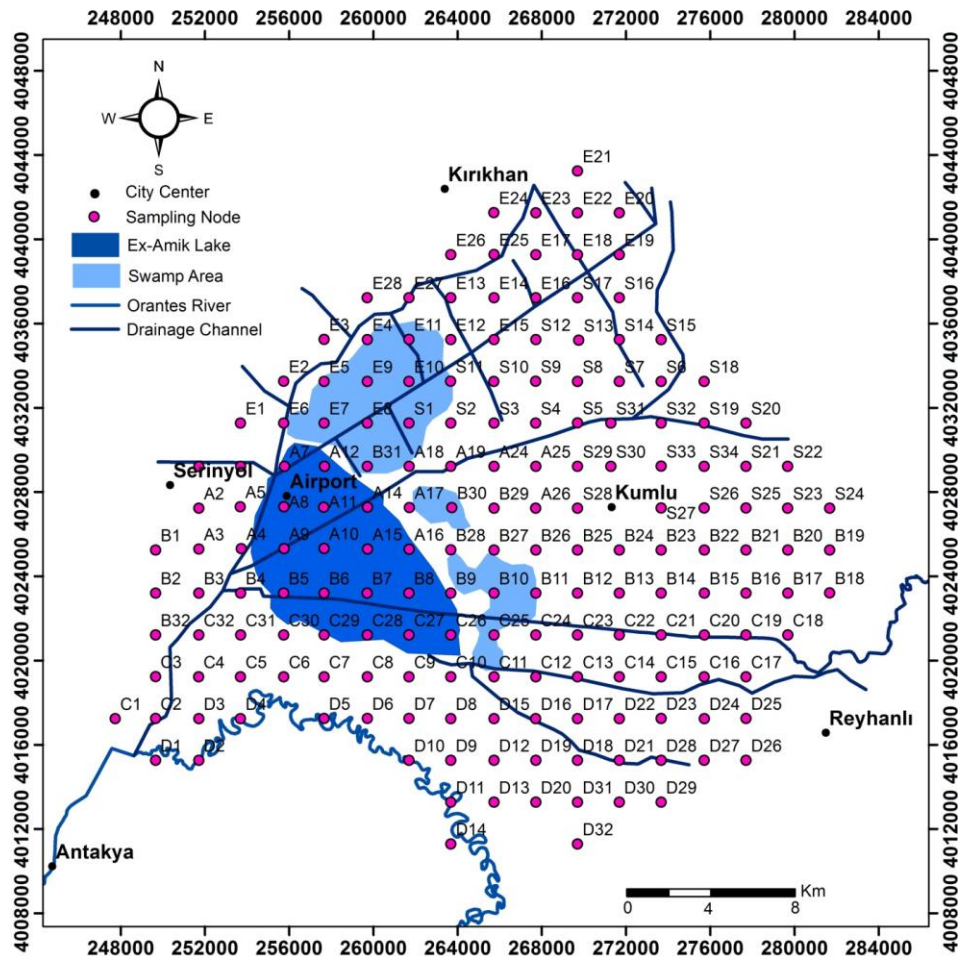


Figure 1. Sampling nodes with drainage-web, former Amik Lake and related swamp areas

Table 1. Descriptive statistics for measured soil properties and selenium fractions

Parameters	Min.	Max.	Mean	Std. err.	Std. dev.	Variance	Skewness*	Kurtosis**
OM (%)	0.10	4.035	1.509	0.062	0.824	0.679	0.610	0.138
pH (1:2.5)	7.02	8.930	7.935	0.020	0.274	0.075	0.187	0.843
EC (dS cm <sup>-1</sup> )	78.00	1189	274.0	12.23	163.6	26774	2.154	6.641
CCE (%)	0.00	43.95	19.31	0.738	9.87	97.51	-0.025	-0.649
Clay (%)	0.00	58.85	13.56	0.912	12.20	148.9	1.263	1.009
Sand (%)	4.58	70.88	26.49	0.932	12.47	155.4	1.210	1.518
Silt (%)	15.58	88.76	59.95	1.164	15.57	242.3	-0.653	-0.168
Se1 (µg kg <sup>-1</sup> )	6.43	177.3	62.68	2.370	31.70	1005	1.185	1.584
Se2 (µg kg <sup>-1</sup> )	13.10	3591	246.8	21.32	285.3	81372	9.320	107.2
Se3 (µg kg <sup>-1</sup> )	34.80	1226	189.4	12.39	165.7	27461	3.211	13.58
ΣSe (µg kg <sup>-1</sup> )	150.5	3802	498.9	27.22	364.2	132670	5.134	39.47

\*Standard error of skewness 0.182, \*\* Standard error of kurtosis 0.361 (N= 179), ΣSe is the total of Se1-3 fractions. CCE calcium carbonate equivalent.

The extraction of phosphate buffer, Se1, involves the determination of the combined presence of exchangeable and soluble ionic species of selenium, including  $\text{Se}^{4+}$ ,  $\text{Se}^{6+}$ , and  $\text{Se}^{2-}$ . A soil sample weighing 1 gram was passed through a sieve with a mesh size of 150  $\mu\text{m}$ . The resulting sieved soil sample was then subjected to an extraction process using 25 mL of a phosphate buffer solution with a concentration of 0.1 M and a pH value of 7. The extraction was carried out by shaking the mixture at a speed of 130 revolutions per minute at room temperature for a duration of 1 hour. The liquid phase was prepared for examination by centrifugation and filtering techniques. The sample was maintained at a temperature of  $+4^\circ\text{C}$  until it was subjected to analysis. In this work, the total concentration of distinct ion species of Se was used, despite the fact that they were analyzed individually.

Persulfate oxidation involving the proportion of Se2 that is either bound or adsorbed to organic matter was conducted as follows: Following a first stage whereby soil samples were subjected to the addition of 25 mL of a 0.1 M persulfate solution, the mixture was heated in a sand bath at a temperature of  $90^\circ\text{C}$  for a duration of 2 hours. Subsequently, the liquid phase was separated and subjected to analysis using the same procedure as outlined in Step 1. The introduction of persulfate resulted in the oxidation of Se (II-) present in the medium to Se (VI). Next, a volume of 0.5-2 mL of the extract was selected and supplemented with 6 M HCl until the total volume reached 25 mL. The resulting mixture was then subjected to heating in a sand bath at a temperature of  $90^\circ\text{C}$  for a duration of 30 minutes. This process aimed to facilitate the reduction of Se (VI) to Se (IV). The quantification of the  $\text{Se}^{4+}$  ion species in the solution phase was conducted.

The process of oxidation of nitric acid on Se3, refers to the residual selenium fraction associated with the mineral structure. A volume of 2.5 mL of concentrated  $\text{HNO}_3$  was introduced to the remaining samples obtained from step 2 and then subjected to heating using a sand bath.

The sample was subjected to a temperature of  $90^\circ\text{C}$  for a duration of 30 minutes. Subsequently, it was allowed to cool down before adding 20 mL of distilled water. The resulting mixture was then placed in a sand bath and maintained at a temperature of  $90^\circ\text{C}$  for a period of 1.5 hours. After routinely separating the liquid phase, the Se (0) concentration in the solution was measured by HRC-AAS. [Ozkan et al. \(2022\)](#) provided a comprehensive account of the analytical techniques involved in verifying the analysis using reference material (BCR-176 fly ash, the Sigma-Aldrich, Belgium) and optimizing High-Resolution Continuum Source Atomic Absorption Spectrometer, HRC-AAS (Analytic Jena Conrtaa 700). Table 1 displays the descriptive data according to the fractions and total Se parameters.

**Statistical analysis:** In the context of statistical analysis, the observation vector  $Y$  represents the observed characteristics. The symbol  $\mu$  denotes the overall mean, while  $a$  represents the random additive genetic influence of the individual soil. Additionally,  $e$  represents the random error connected with the observation vector  $Y$ . Fixed effects were not included in the analysis due to the inherent difficulty in discerning and quantifying the precise impact of soil forming elements at individual sample locations, even if these components do vary. The estimation of variance components and genetic factors arising from soil formation was conducted using the constrained maximum likelihood method ([Madsen and Jensen, 2010](#)). Canonical correlations were computed using genetic correlations resulting from the soil formation processes. The study used canonical correlation analysis to elucidate and quantify the association between two qualities, using the SAS 9.0 PROC CANCORR software package ([SAS Institute, 2002](#)).

Canonical correlation analysis is based on the correlation of the linear combination between two sets of variables ( $Xp$ ) ( $Yq$ ) (Johnson and Wichern, 1986). In this case, linear combinations of variables' sets can be defined as follows:

$$\begin{aligned} U_i &= a_{i1}X_1 + a_{i2}X_2 + \dots + a_{ip}X_p \\ V_i &= b_{i1}Y_1 + b_{i2}Y_2 + \dots + b_{iq}Y_q \end{aligned}$$

Where  $a_{i2}$  and  $b_{i2}$  are the canonical coefficients,  $U_i$  and  $V_i$ ,  $i$  are the  $i^{\text{th}}$  pair of canonical variables.  $U_i$  and  $V_i$  are the first pair of canonical variables associated with the first canonical correlation that is formed as follows:

$$r_i = \frac{Cov(U_1, V_1)}{\sqrt{Var(U_1)Var(V_1)}}$$

The percentage of variance explained by the first canonical variables  $U_{x_i}^2$  and  $V_{w_i}^2$  was calculated as follow:

$$U_{x_i}^2 = \frac{\sum_{j=1}^p a_{ij}^2}{p} \text{ and } V_{w_i}^2 = \frac{\sum_{j=1}^q b_{ij}^2}{q}$$

$p$  and  $q$  are the variable numbers of  $X$  and  $Y$ , respectively. The smallest value between  $p$  and  $q$  defines the number of pairs of canonical variables.

Two sets of features were established: The first included the descriptive soil properties, and the second consisted of Se fractions, including total Se.

## Results and Discussion

The examination of correlation coefficients within the soil characteristics dataset may provide insights into the collective impact of soil-forming variables and agricultural activities within the designated research region. The variables included in the analysis of soil

attributes show a strong negative association between silt content and clay and sand content, with correlation coefficients of -0.554 and -0.600, respectively. This relationship is derived from the meticulous measurement of textural fractions and the underlying processes of soil formation. The inverse correlation between EC and CCE suggests that the soil has developed from lime-based parent material or is mostly composed of secondary lime formations (Usta, 1995), with deposition playing a minor role. According to Usta (1995), the presence of clay is associated with a negative association, suggesting that the adsorption process effectively decreases the concentration of soluble salts when adequate drainage conditions are present. In contrast, it has been shown that soils containing a significant proportion of sand exhibit elevated electrical conductivity (EC) values when subjected to comparable environmental circumstances. This phenomenon might be attributed to the existence of a clay barrier located near the surface within the plain, particularly in areas of low elevation or former lake beds (Akis, 2015). The inverse correlation between lime and sand is characterized by the absence of lime buildup in areas with high sand content, often seen in the elevated regions of the research area. This lack of lime accumulation may be attributed to the limited water retention capacity of sandy soils, resulting in rapid drainage of incoming water from the soil profile. However, it is important to note that due to the bowl-shaped physiography of the plain, these particular

places also serve as the primary locations for the deposition of coarse materials that have been eroded from higher elevations. The inclusion of the former lake bottom within the research area suggests that the clay fraction, which is transported by erosion, tends to collect more prominently in this particular region. This accumulation signifies the potential development of secondary carbonate formations over time at the bottom of the lake.

The presence of selenium (Se) in the Amik plain exhibits a significant and inverse correlation with the organic-bound (Se2) and residual-Se (Se3) fractions, while demonstrating a comparatively weaker association with the available-Se fraction, as shown in Table 2. These findings suggest that the distribution of selenium in the region may be influenced by the processes of solute and solid transport. This method is substantiated by the observed positive relationship between the sum of selenium ( $\Sigma$ Se) and sand content, which represents the indigenous fraction in water with moderate erosion, as well as the electrical conductivity (EC), which is indicative of solute transport as influenced by the physiography of the plain. Nevertheless, the proportion of accessible selenium (Se1) was dependent on the concentration of S3, which is likewise closely associated with the levels of S2 components. Hence, organic matter plays a pivotal part in this process. The aforementioned findings were substantiated by the statistically significant correlations

**Table 2.** Correlations within and between sets of variables for soil properties and selenium content

Correlation coefficients for within soil properties set						
	OM	pH	EC	CCE	Clay	Sand
pH	-0.406**					
EC	-0.249**	0.174*				
CCE	-0.033	0.061	-0.275**			
Clay	-0.004	0.150*	-0.144*	0.149*		
Sand	-0.058	-0.033	0.358**	-0.387**	-0.211**	
Silt	0.068	-0.114	-0.173*	0.201**	-0.554**	-0.600**
Correlation coefficients for within selenium fractions set						
	Se1	Se2	Se3			
Se2	-0.086					
Se3	0.441**	0.397**				
$\Sigma$ Se	-0.299**	-0.818**		-0.811**		
Correlation coefficients for between the sets						
	Se1	Se2	Se3	$\Sigma$ Se		
OM	-0.116	0.299**	0.139	-0.238**		
pH	0.187*	-0.095	-0.116	0.087		
EC	-0.117	-0.225**	-0.266**	0.305**		
CCE	-0.051	0.121	0.172*	-0.189*		
Clay	0.026	0.149*	0.137	-0.164		
Sand	0.008	-0.214**	-0.218**	0.269**		
Silt	-0.059	0.083	0.082	-0.110		

OM organic matter, CCE calcium carbonate equivalent

observed between organic matter and S2 ( $r = 0.299$ ,  $p < 0.01$ ) as well as  $\Sigma\text{Se}$  ( $r = -0.238$ ,  $p < 0.01$ ). The pH (0.186\*) most likely had an impact on the regulation of the Se1 fraction in the soils. According to [Usta \(1995\)](#), there is a positive correlation between the Se2 percentage and the presence of organic matter and clay content. This relationship is attributed to the ability of organic matter and clay to protect organic matter from degradation by generating organo-mineral complexes. In contrast, it has been shown that sand (-0.214\*\*) and electrical conductivity (EC) (-0.225\*\*) exhibit a negative correlation with the ability of organic matter (OM) content to grow. The observed strong association between Se3, sand, and EC underscores the importance of solute and solid transport mechanisms. The lithological composition of the studied region mostly consists of Quaternary alluvium and colluvium, as shown by [Korkmaz et al. \(2011\)](#). However, it is possible to interpret the positive correlation of CCE as the result of primary and/or secondary carbonate minerals playing a role in enhancing this percentage via adsorption processes. Another notable observation is that carbonated rocks have a greater Se concentration (0.8 mg/kg) compared to metamorphic and

sedimentary rocks (0.6 mg/kg) as well as igneous rocks (0.5 mg/kg) ([Aubert and Pinta, 1977](#)), providing an explanation for the correlation.

Table 3 presents the results of the canonical correlations, adjusted canonical correlations, standard errors, and squared canonical correlations computed between the variable sets of soil characteristics and Se fractions. The statistical results for evaluating the hypothesis of the canonical structure are also shown in Table 3. The first canonical correlation, which represents the correlation between the first pair of canonical variables, for soil characteristics and Se fractions sets is reported to be  $0.466 \pm 0.059$  (Table 3). The first canonical correlation has statistical significance and is distinguishable from the next three canonical correlations. The observed significance of the first canonical correlation suggests a strong link between linear combinations of soil characteristics variables and linear combinations of Se fractions variables. In this particular instance, hypothesis H1 is deemed valid, hence refuting the null hypothesis H0 which posits that all canonical correlations derived from the study are equal to zero ( $\text{Pr} > F$ ; 0.001). Nevertheless, this assertion holds true just for the first

**Table 3.** Canonical correlations and their significance test

Canonical variables	Canonical correlations	Adjusted canonical correlations	St. error	Squared canonical correlations	Eigen values	Differences	Percentage (%)	Cumulative percentage (%)
1	0.466	0.418	0.059	0.217	0.279	0.180	63.0	63.0
2	0.300	0.232	0.068	0.090	0.099	0.053	22.4	85.4
3	0.211	0.159	0.072	0.045	0.047	0.029	10.5	95.9
4	0.133	0.110	0.074	0.018	0.018		4.1	100.0

Canonical variables	Probability ratio	F values	DF	Pr > F	Wilks-Lambda value
1	0.668	2.57	28	<0.001	<0.0001
2	0.854	1.53	18	0.076	-
3	0.939	1.10	10	0.365	-
4	0.982	0.77	4	0.544	-

canonical correlation that is substantiated by the Wilks-Lambda test statistic.

According to [Keskin and Ozsoy \(2004\)](#), when the variables in the data sets have different units of measurement and unequal variances, it is more appropriate to interpret standardised canonical correlation coefficients rather than canonical correlations in research. Standardized coefficients have significance due to their ability to demonstrate the magnitudes of the effects of the original variables within a given set on the occurrence of a canonical variable within the same set ([Keskin and Ozsoy, 2004](#)). Hence, Table 4 presents the standardized canonical

correlation coefficients pertaining to the first pair of canonical variables that has been shown to possess statistical significance.

The critical questions for Se in soils are related to i) soil properties can be responsible for the total Se, and ii) the availability of Se as a function of soil properties and Se fractions. Using the standardised canonical coefficients given in Table 4, the equations for the first canonical variable pair of soil properties (Soil-1) and Se fractions (Se1-1) can be calculated as follows:

$$\text{Soil-1} = 0.438 \text{ OM} - 0.352 \text{ pH} - 0.063 \text{ EC} + 0.349 \text{ CCE} +$$

0.565 clay – 0.020 sand + 0.481 silt  
 Sel-1 = -0.733 Se1 - 0.263 Se2 + 0.213 Se3 – 0.950  $\Sigma$ Se

**Table 4.** Standardized coefficients of the first canonical variable

Soil properties set		Se fractions set	
	Soil-1		Sel-1
OM	0.438	Se1	-0.733
pH	-0.352	Se2	-0.263
EC	-0.063	Se3	0.213
CCE	0.349	$\Sigma$ Se	-0.950
Clay	0.565		
Sand	-0.020		
Silt	0.481		

The loadings for the organic matter (OM), cation exchange capacity (CCE), clay, and silt variables are 0.438, 0.349, 0.565, and 0.481, respectively. These loadings exhibit a positive relationship in the formation of the Soil-1 canonical variable. The clay variable exhibits the greatest loading (0.565) among the given variables. In contrast, the variables of pH, EC, and sand exhibit negative and minor contributions, with corresponding coefficients of -0.352, -0.063, and -0.020, except for pH. The presence of selenium (Se) in a particular soil is influenced by several factors, including the parent material, adsorption, and desorption processes, precipitation, and dissolution processes, as well as solute and solid deposition processes. The susceptibility of these processes is influenced by various soil properties, including pH, redox potential (Fang and Zhao, 2018; Xing et al., 2015; Nakamaru et al., 2014), organic matter (Smazikova et al., 2019; Xing et al., 2015), soil texture and mineralogy (Antoniadis et al., 2017; Munier-Lamy et al., 2007), sesquioxide content (Araujo et al., 2020; Fan et al., 2019), and topographic parameters (Xu et al., 2018). The role of topography may be considered a significant factor influencing the occurrence of short and long-range transport phenomena. The presence of organic matter (OM), clay, silt, and sand are strongly influenced by topography under certain climatic circumstances at a localized level, as shown in this research. In a similar vein, Xu et al (2018) found that the concentration of selenium (Se) in soil is predominantly influenced by topography and the presence of iron oxide (Fe<sub>2</sub>O<sub>3</sub>). On the other hand, the availability of Se in soil is strongly correlated with factors such as pH, weathering processes, and eluviation. The drainage processes of the old Amik Lake likely had a significant effect on the transport phenomena that led to the buildup of substances before drainage and facilitated their leaching during the post-drainage era. The accumulation of organic debris, clay, silt, and secondary carbonate has been seen in the shallow ex-lake. In their study, Fan et al. (2019) documented the movement of selenium (Se) in soils as a result of humic acid and waterlogging. The waterlogging phase is associated with a fall in pH, as shown by Oren et al. (2018). In

contrast, Lindsay (2001) found that a rise in pH characterizes the post-drainage period. This is supported by the negative impact of pH (-0.352) on the Soil-1 function. In the derivation of the canonical variable for the Se fraction (Sel-1), it was seen that the variables Se1, Se2, and  $\Sigma$ Se had negative contributions (-0.733, -0.263, and -0.950, respectively), but the variable B3 had a positive contribution (+0.213). The  $\Sigma$ Se variable has the greatest effect on the establishment of the canonical correlation of Se fractions. Hence, the rationale provided for the function of Soil-1 might similarly apply in this context. Furthermore, within the research region, there is a process of organic matter decomposition after drainage as well as intermittent occurrences of waterlogging, particularly in the lake-bed. These factors together contribute to the accelerated depletion of Se. In general, it can be said that leaching processes are active inside the research region.

Another correlation that is derived from the canonical correlation analysis is the correlation between the pairs of canonical variables and the original variables in both their own set and the counter set. Table 5 presents the multiple correlation coefficients between the initial canonical variables derived from both sets, as well as the original variables inside their respective sets and the counter set. The significance of these correlation coefficients is in their ability to quantify the amount of variation accounted for by the canonical variable in both the original variables within its own set and the complementary set. This is often referred to as the coefficient of determination (R<sup>2</sup>). Among these correlations given in Table 5, the highest correlation (+0.617) between the first canonical variable (Soil-1) of the soil properties set and the original variables in its own set is between organic matter content. Conversely, the canonical variable (Sel-1) representing the initial fraction of selenium (Se) had the strongest connection (-0.795) with the aggregate sum of Se across all original variables. Furthermore, the other correlations that were identified exhibited statistical significance and were determined to be in close proximity to the aforementioned values. This observation indicates that the initial set of canonical variables, namely Soil-1 and Sel-1, effectively capture the variability within their respective variable sets. Nevertheless, the first pair of canonical variables exhibits a lesser degree of effectiveness compared to their respective variable sets in elucidating the variability inside the opposing variable set. The strongest association ( $r = +0.288$ ) is seen between the first canonical variable of selenium fractionation (Sel-1) and the original variables representing soil characteristics, namely organic matter. Simultaneously, it is seen that 28.8% of the variance in the organic matter variable can be accounted for by the first canonical variable of Se fractions. Similarly, the Soil-1 canonical variable explains 37.1% and 35.0% of the variances in the  $\Sigma$ Se

**Table 5.** Correlations between canonical variables of soil properties and selenium fractions sets and the variables in these sets

Variables	Soil-1	Sel-1	Variables	Sel-1	Soil-1
OM	0.617	0.288	Se1	-0.301	-0.141
pH	-0.489	-0.228	Se2	0.749	0.350
EC	-0.501	-0.234	Se3	0.642	0.300
CCE	0.518	0.242	ΣSe	-0.795	-0.371
Clay	0.310	0.145			
Sand	-0.599	-0.280			
Silt	0.331	0.154			

OM organic matter, CCE calcium carbonate equivalent

and Se2 variables, respectively. This suggests a strong correlation between Se's overall quantities and proportions and soil formation processes.

The redundancy analysis assesses the degree to which the standardized variances of the canonical variable pairs explain the variability in both their own set and the counter set. Table 6 presents the proportion of the overall standardized variation in the soil characteristics dataset that is accounted for by its respective canonical factors, as well as the proportion

explained by the canonical variables of the Se fractions dataset. The first canonical variable in the soil characteristics set accounts for 24.3% of the overall variation. The explanatory percentages for the second and subsequent canonical variables were 11.8%, 17.5%, and 12.5%, respectively. The first canonical variable had the greatest coefficient of determination ( $R^2$ ) at 21.8%. The cumulative proportions of the canonical variables were determined to be 66% in the end. The first canonical variable in the soil characteristics set accounted for 9.2% of the standardized variation in the

**Table 6.** Explained percentages of standardised variance in the set of soil properties and selenium fractions by canonical variables

Amounts of total variance in the set of soil properties explained by their canonical variables				Amounts of total variance in the set of selenium fractions by their canonical variables	
Canonical variable	Percentage	Cumulative percentage	Canonical $R^2$	Percentage	Cumulative percentage
1	24.3	24.3	0.218	9.2	9.2
2	11.8	36.1	0.090	2.9	12.1
3	17.5	53.4	0.045	0.6	12.7
4	12.5	66.0	0.018	0.2	12.9

Se fractions set. The aggregate proportion of four canonical variables within this dataset was found to be 12.9%.

## Conclusions

In recent decades, there has been a significant increase in the availability of research data, facilitated by the advancement and cost reduction of current technology. The use of multivariate analysis methodologies enables the extraction of novel and valuable insights while assessing the interconnections across extensive data sets. The present work effectively employs canonical correlation, a multivariate statistical method that provides a fresh viewpoint on elucidating the interrelationship between Se fractions and soil parameters. This research investigates the distribution of Se fractions in the soils of Amik Plain. The first canonical correlation demonstrates a substantial link between the soil characteristics and Se fraction data sets. The findings indicated that soil characteristics and

the transit of soluble and solid fractions of Se had a significant influence. The contributions of organic matter, clay, and silt characteristics to the fractionation of Se and its overall concentration were shown to be statistically significant among the various soil parameters. The data presented indicates that the first canonical variable has a significant capability for accurately predicting the occurrence of Se. In this regard, the soil parameters may be regarded as the independent variables used to elucidate the variability in the Se fractions, which serve as the dependent variables. The assessment of the associations between created data sets via the use of multivariate analytic techniques, such as canonical correlation analysis, yields novel and valuable insights. Nevertheless, it is important to highlight that the suitability of the canonical correlation technique for various ecosystems characterized by varying scales and plant nutrient compositions has to be empirically evaluated. Another factor to consider is the heterogeneity of the research region, which may provide limitations on the ability to



extract more information. Hence, it is essential to thoroughly evaluate the whole process in order to address the constraints associated with data screening methodologies.

### Authors contributions

**AÖ:** Contributed substantially to the conception and design of the study, soil sampling, the acquisition of data or the analysis and interpretation. Provided final approval of the version to publish. **ANÖ:** Statistical analysis, interpretation of statistical data, drafted the statistical part of the manuscript and approved the final version. **VU:** Contributed substantially to the conception and design of the study, drafted and approved the final version of the manuscript.

### Funding Information

This project was supported by the Scientific Research Foundation of Hatay Mustafa Kemal University Funding Information Grand number: MKU-BAP, Project No: 9820. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### Conflict of Interest

The authors declare that there is no conflict of interest.

### References

- Akis, R. (2015). Spatial variability of soil solute and saturated hydraulic conductivity affected by undrained water table conditions. *Precision Agriculture*, 16, 330–359. <https://doi.org/s11119-014-9379-0>
- Alfthan, G., Eurola, M., Ekholm, P., Venalainen, E. R., Root, T., Korkalainen, K., Hartikainen, H., Salminen, P., Hietaniemi, V., Aspila, P., Aro, A., & Selenium working group (2015). Effects of nationwide addition of selenium to fertilizers on foods, and animal and human health in Finland: From deficiency to optimal selenium status of the population. *Journal of Trace Elements Medicine and Biology*, 31, 142–147. <https://doi.org/10.1016/j.jtemb.2014.04.009>
- Antoniadis, V., Levizou, E., Shaheen, S. M., Ok, Y. S., Sebastian, A., Baum, C., Prasad, M. N., Wenzel, W. W., & Rinkbele, J. (2017). Trace element in the soil-plant interface: phytoavailability, translocation, and phytoremediation- A review. *Earth-Sciences Reviews*, 171, 621–645. <https://doi.org/10.1016/j.earscirev.2017.06.005>
- Araujo, A. M., Lessa, J. H. D., de Lima, F. R. D., Raymundo, J. F., Curi, N., Guilherme, L. R. G., & Lopes, G. (2019). Adsorption of selenite in tropical soils as affected by soil management, ionic strength, and soil properties. *Journal of Soil Science & Plant Nutrition*, 20, 139–148 (2020). <https://doi.org/10.1007/s42729-019-00107-x>
- Aubert, H., & Pinta, M. (1977). *Trace elements in soils*. Elsevier, Amsterdam.
- Eswayah, A. S., Smith, T. J., Scheinost, A. C., Hondow, N., & Gardiner, P. H. E. (2017). Microbial transformations of selenite by methane-oxidizing bacteria. *Applied Microbiology and Biotechnology*, 101(17), 6713–6724. <https://doi.org/10.1007/s00253-017-8380-8>
- Fan, J. X., Zeng, Y., & Sun, J. X. (2018). The transformation and migration of selenium in soil under different Eh conditions. *Journal of Soil Sediments*, 18(9), 2935–2947. <https://doi.org/10.1007/s11368-018-1980-9>
- Fan, J. X., & Zhao, G. L. (2018). Sorption characteristics and fraction distribution changes of selenite in soil. *Sustainability-Basel*, 10(7), 2491. <https://doi.org/10.3390/su10072491>
- Fan, J. X., Zhao, G. L., Sun, J. X., Hu, Y., & Wang, T. J. (2019). Effect of humic acid on Se and Fe transformations in soil during waterlogged incubation. *Science of the Total Environment*, 684,476–485. <https://doi.org/10.1016/j.scitotenv.2019.05.246>
- Hyun, S., Burns, P. E., Murarka, I., & Lee, L. S. (2006). Selenium (IV) and (VI) sorption by soils surrounding fly ash management facilities. *Vadose Zone Journal*, 5(4), 1110–1118. <http://dx.doi.org/10.2136/vzj2005.0140>
- Johnson, R. A., & Wichern, D.W. (2002). *Applied multivariate analysis*. Prentice-Hall, Inc., New Jersey.
- Kacar, B. (2009). *Soil analysis*. Nobel Academic Press, Ankara, Turkey (in Turkish).
- Keskin, S., & Ozsoy, A. N. (2004). Kanonik korelasyon analizi ve bir uygulaması. *Tarım Bilimleri Dergisi*, 10(1), 67–71.
- Korkmaz, H., B. Çetin, I. Ege, A. Karataş, A. Bom, E. Ozsahin. 2011. Environmental effects of stone pits in Hatay (Turkey). *Procedia - Social and Behavioral Sciences*, 19, 504–510. <https://doi.org/10.1016/j.sbspro.2011.05.162>
- Lee, S., Doolittle, J. J., & Woodard, H. J. (2011). Selenite adsorption and desorption in selected South Dakota soils as a function of pH and other oxyanions. *Soil Science*, 176(2), 73–79. <http://dx.doi.org/10.1097/SS.0b013e31820a0ff6>
- Li, Z., Liang, D. Peng, Q., Cui, Z., Huang, J., & Lin, Z. (2017). Interaction between selenium and soil organic matter and its impact on selenium bioavailability: a review. *Geoderma*, 295, 69–79. <https://doi.org/10.1016/j.geoderma.2017.02.019>
- Lindsay, W.L. (2001). *Chemical equilibria in soils*. The Blackburn Press, USA.
- Martens, D. A., & Soares, D.L. (1997). Selenium speciation of soil/sediment determined with sequential extraction and hydride generation atomic absorption spectrophotometry. *Science of the Total Environment*, 31, 133–139. <https://doi.org/10.1021/es960214>
- Mikkelsen, R., Page, A., & Bingham, F. (1989). Factors affecting selenium accumulation by agricultural crops. In L.W. Jacobs (Ed.) *Selenium in agriculture and the environment*, (pp. 65–94). <https://doi.org/10.2136/sssaspepub23.c4>
- Munier-Lamy, C., Deneux-Mustin, S., Mustin, C., Merlet, D., Berthelin, J., & Leyval, C. (2007). Selenium bioavailability and uptake as affected by four different plants in a loamy clay soil with particular attention to mycorrhizae inoculated ryegrass. *Journal of Environmental Radioactivity*, 97(2-3),148–158.

- <https://doi.org/10.1016/j.jenvrad.2007.04.001>  
 Navarro-Alarcon, M., & Cabrera-Vique C. (2008). Selenium in food and the human body: a review. *Science of the Total Environment*, 400, 115–41. <https://doi.org/10.1016/j.scitotenv.2008.06.024>
- Nakamaru, Y. M., & Altansuvd, J. (2014). Speciation and bioavailability of selenium and antimony in non-flooded and wetland soils: A review. *Chemosphere*, 111, 366–371. <https://doi.org/10.1016/j.chemosphere.2014.04.024>
- Ören, S., Uygur, V., & Sukuşu, E. (2018) Effect of redox potential induced changes on Fe and Mn availability in soils with differing characteristics. *Mediterranean Agricultural Sciences*, 31(3), 301-309. <https://doi.org/10.29136/mediterranean.409050>
- Ozkan, A., Uygur, V., Sungur, S., & Ozkan, V. (2022). Relationships between physico-chemical properties of the soil and selenium speciations from Amik plain, Turkey, *Fresenius Environmental Bulletin*, 31, 3805-3818.
- Ozsoy, A. N. (2019). Egg and chick quality characteristics of meat type japanese quail (coturnix coturnix japonica) line by canonical correlation analysis. *Fresenius Environmental Bulletin*, 28(4), 2582-2588.
- Rayman, M. P. (2000). The importance of selenium to human health. *Lancet*, 356, 233–41. [https://doi.org/10.1016/S0140-6736\(00\)02490-9](https://doi.org/10.1016/S0140-6736(00)02490-9)
- Reid, M. E., Duffield-Lillico, A. J., Slate, E., Natarajan, N., Turnbull, B., Jacobs, E., et al. (2008). The nutritional prevention of cancer: 400 mcg per day selenium treatment. *Nutrition and Cancer*, 60, 155–63. <https://doi.org/10.1080/01635580701684856>
- Saha, U., Fayiga, A., & Sonon, L. (2017). Selenium in the soil-plant environment: A review. *International Journal of Applied Agricultural Sciences*, 3, 1-18. <https://doi.org/10.11648/j.ijaas.20170301.11>
- SAS Institute (2002). *SAS/STAT user's guide*. Version 9.1, Cary, North Carolina, USA.
- Sharma, S. (1996). *Applied multivariate techniques*. John Willey & Sons, Inc., Canada.
- Shin, K. (1996). *SPSS guide for DOS version 5 and windows 6.1.2*. 2nd Edition, Irwin, Chicago.
- Smazikova, P., Praus, L., Szakova, J., Tremlova, J., Hanc A., & Tlustos, P. (2019). Effects of organic matter-rich amendments on selenium mobility in soils. *Pedosphere*, 29(6), 740–751. [https://doi.org/10.1016/S1002-0160\(17\)60444-2](https://doi.org/10.1016/S1002-0160(17)60444-2)
- Stevenson, F. J. (1994). *Humus chemistry: Genesis, composition, reactions*. 2nd ed. Wiley & Sons, New York.
- Tapiero, H., Townsend D. M., & Tew, K. D. (2003). The antioxidant role of selenium and selenocompounds. *Biomedicine & Pharmacotherapy*, 57, 134–44. [https://doi.org/10.1016/S0753-3322\(03\)00035-0](https://doi.org/10.1016/S0753-3322(03)00035-0)
- Tolu, J., Thiry, Y., Bueno, M., Jolivet, C., Potin-Gautier, M., & Le Hecho, I. (2014). Distribution and speciation of ambient selenium in contrasted soils, from mineral to organic rich. *Science of the Total Environment*, 479, 93-101. <https://doi.org/10.1016/j.scitotenv.2014.01.079>
- Usta, S. (1995) *Toprak Kimyası*. Ankara Üniversitesi Ziraat Fakültesi Yayınları, no: 1387, Ankara.
- Xing, K., Zhou, S. B., Wu, X. G., Zhu, Y. Y., Kong, J. J., Shao, T., & Tao, X. (2015). Concentrations and characteristics of selenium in soil samples from Dashan Region, a selenium-enriched area in China. *Soil Science & Plant Nutrition*, 61(6), 889-897. <https://doi.org/10.1080/00380768.2015.1075363>
- Xu, Y. F., Li, Y. H., Li, H. R., Wang, L., Liao, X. Y., Wang, J., & Kong, C. (2018). Effects of topography and soil properties on soil selenium distribution and bioavailability (phosphate extraction): A case study in Yongjia County, China. *Science of the Total Environment*, 633, 240-248. <https://doi.org/10.1016/j.scitotenv.2018.03.190>
- Wang, D., Xue, M. Y., Wang, Y. K., Zhou, D. Z., Tang, L., Cao, S. Y., Wei, Y. H., Yang, C., & Liang, D. L. (2019). Effects of straw amendment on selenium aging in soils: Mechanism and influential factors. *Science of the Total Environment*, 657, 871-881. <https://doi.org/10.1016/j.scitotenv.2018.12.021>
- White, R. E. (2005). *Principles and practice of soil science: The soil as a natural resource*. Wiley-Blackwell, New York.
- Winkel, I.H., Vriens, B., Jones, G.D., Schneider, L.S., Pilon-Smits, E., & Banuelos, G.S. (2015). Selenium cycling across soil-plant-atmosphere interfaces: a critical review. *Nutrients*, 7, 4199-4239. <https://doi.org/10.3390%2Fnu7064199>