RESEARCH PAPER

Multivariate linear modeling of soil properties using soil formation factors in different geological formations

Alireza Pazhuhandeh [1] , **Ali Mohammadi Torkashvand** [1],*, **Abolfazl Moeini** [2], **Ebrahim Pazira** [1]

[1] Department Soil Science, Science and Research Branch; Islamic Azad University, Tehran, Iran
[2] Department Watershed, Science and Research Branch, Islamic Azad University, Tehran, Iran

Abstract
This study aimed to investigate the relationship between these factors and soil physicochemical properties for rapid management of watersheds and Multivariate linear modeling of soil properties using soil formation factors. Four watersheds of Qazvin province were studied, namely Juyank, Madan, Niyarak, and Plangeh. First, the unit map was prepared using a geographical information system (GIS) on a scale of 1:25000 based on photogeological investigations and field visits. Then, 101 soil samples were collected and EC, pH, organic matter percentage, sand fraction, silt proportion, and clay percentages were measured. The relationship between soil formation factors and properties was investigated using linear multivariate regression in three methods of Enter, Forward, and Stepwise. Therefore, the 5 soil-forming factors were the independent variables, and the soil's physical and chemical characteristics were set as the dependent variables. The accuracy of these models was validated using two statistics of explanation coefficient and mean squared error (RMSE). The results demonstrated that the highest R^2 value of 0.78 with an RMSE of 0.56 was associated with the relationship between pH and soil-forming factors (regression by the Enter method). There was less than a 50% correlation between silt, clay, and electrical conductivity with soil-forming factors.

Keywords: Soil Formation Factors, Regression, One-Way Analysis Of Variance, Soil Characteristics

Introduction

Soil formation depends on several factors, including parent material, climate, topography, time, and living organisms. According to Jenny (1941) several soil formation factors influence the formation of soil properties (s) Where CL is the climate, O is living organisms, R is topography, P is parent material and T is time (Jenny, 1941).

The mentioned five reactions occur with different intensities and relative importance in the formation of all soil types, which may result in different profile characteristics (Thanachit et al., 2006). A wide variety of environmental factors and processes contribute to soil formation in different climates, geological conditions, low elevations, vegetation conditions, and time conditions (Buo et al., 2011). In other words, the variation of the different soil formation factors leads to different physical and chemical properties of the soil (Thanachit et al., 2006; Vingiani et al., 2010). The distribution of elements between the solid phase and the soil solution is important to determine their mobility and stabilization in the soil. The mobility and availability of elements depend on physicochemical properties (such as particle size distribution, especially

clay content, soil pH, organic matter, soil carbonate content, ionic strength of soil solution, cation exchange capacity, etc.), mineralogical properties, and biological properties of the soil (He et al., 2004). The main soil formation processes affecting the behavior of elements in the soil are 1. The release of elements from the parent material during aeration; and 2. the movement and accumulation of soil constituents, including silicate clay, oxides, and hydroxides, as well as organic material that can absorb elements (Ostwari et al., 2019; Gomes et al., 2001). The study of the origin and evolution of soils, taking into account their geologic and geomorphologic characteristics, generally provides a better understanding of soil formation processes (Phillips, 2001). The parent material factor controls soil formation in arid and semiarid regions. Furthermore, the other two aspects, living organisms, and topography, will balance this factor over time (Navidi & Abtahi, 2001). The development and evolution level of the soil determines its characteristics, which are shown in comparison with the parent material (Saketeband et al., 2017; Ortiz et al., 2002). In addition, the intensity and type of aeration strongly depend on the geology and composition of the parent material (Bluth & Kump, 1994). In the early stages of soil formation, the chemical composition of the soil is strongly controlled by the parent material, while in mature soils it reflects the effects of the aeration environment (Thanachit et al., 2006, Bouma, 1989).

In time series, soil formation is influenced by a time when it occurs on the same parent material and other factors such as climate, the activity of living organisms, and topography remain constant. The soil will exhibit the characteristics of the parent rock to a high degree of aeration occurs in a short period. In contrast, as aeration progresses, soil formation is influenced by factors other than the parent rock. Therefore, time plays a key role in soil formation. Generally, the longer soil formation takes, the thicker the soil and the less it resembles the parent material (Harden, 1982, Nael et al., 2009). Depending on the parent material, the soil may form from the underlying bedrock or a layer of discontinuous sediments. Soils that form on bedrock are known as residual soils and another type is called transported or non-residual soils (Haghian et al. 2017; Tarbok and Lutgen, 2016).

The bedrock type affects soils in two ways. First, its nature influences aeration and the soil formation rate. For example, the mineralogical composition of the parent rock affects the extent of chemical weathering. In addition, discontinuous deposits form soil more quickly than bedrock because of some aeration. Second, soil fertility is affected by the chemical composition of the materials. As a result, soil that lacks nutrients essential to plant growth is not considered crucial (Raushi et al., 2021; Araujo et al., 2017).

Plants and animals are responsible for providing soil organic matter. The organic content of some soils is completely dominated by organic matter, while desert soils have a smaller proportion of organic matter. Even though soil organic matter percentages vary, no soil is organic matter-free. Plants are the primary sources of organic matter, but animals and millions of microscopic organisms also contribute. The decomposition of organic matter provides notorious material for plants, animals, and microscopic organisms in the soil. Also, the decomposition of plants and animals produces complex organic acids, which intensify the aeration process. In addition, organic materials have excellent water retention capabilities, so they hold water in the soil. Microorganisms such as fungi, bacteria and protozoa contribute significantly to decomposing plant and animal remains to produce humus as the final product. There is no similarity between this substance, humus, and the plants and animals derived from it (Jaafari & Sarmadian, 2017; Birkeland, 1999).

Climate is the most important factor controlling soil formation determining the effect of chemical or mechanical aeration. It also significantly affects the aeration rate and severity. For example, the chemical aeration of soil in a hot and humid climate results in a thick layer, whereas the physical aeration of soil in a cold climate with low humidity produces a thin layer (Egli et al., 2008). Furthermore, precipitation will affect how much material washes from the

soil, thereby determining its fertility. Thus, in terms of the climate factor, precipitation and temperature both play an important role in soil formation (Tarbok & Lutgen, 2016).

Soil chemical reactions are greatly influenced by its topography. These include the intensity and quantity of surface wastewater (runoff) and subsurface drainage; and accordingly, the leaching rate of solutes and the formation of specialized soils, erosion intensity and the transport of degraded materials, or the transport of suspended and soluble materials by water and the degradation of materials by erosion. Furthermore, this factor affects the soil solum depth, thickness or organic matter content, moisture content of a profile, soil color and degree of clarity of soil horizons, soil reaction, and soil temperature (Jiang & Thelen, 2004).

As part of soil investigation, it is essential to know the correlations between soil properties and their quantitative expression in statistical models (Buol et al., 2011; Gray et al., 2016). This study aimed to investigate the relationship between the physicochemical properties of the soil and the soil-forming factors and to explore the possibility of creating maps of the aforementioned soil characteristics based on the national layers of soil-forming elements.

Materials and Methods

Study Area

This study was conducted in four watersheds in Qazvin province including Juyank, Madan, Niyarak, and Plangeh (Figure 1). The characteristics of the study area were presented in Table 1. The watersheds of Juyank and Madan are located in the mountains, while the climatic conditions in Niyarak and Palange are semi-arid and cold. The entire study area is located in the geological zone of central Iran.

Research Methodology

First, land use map, slope, precipitation zonation, temperature, and elevation were prepared using 1:25000 scale geologic maps. Then, these maps were combined in ArcGIS 10.8 and similar areas were combined into working units.
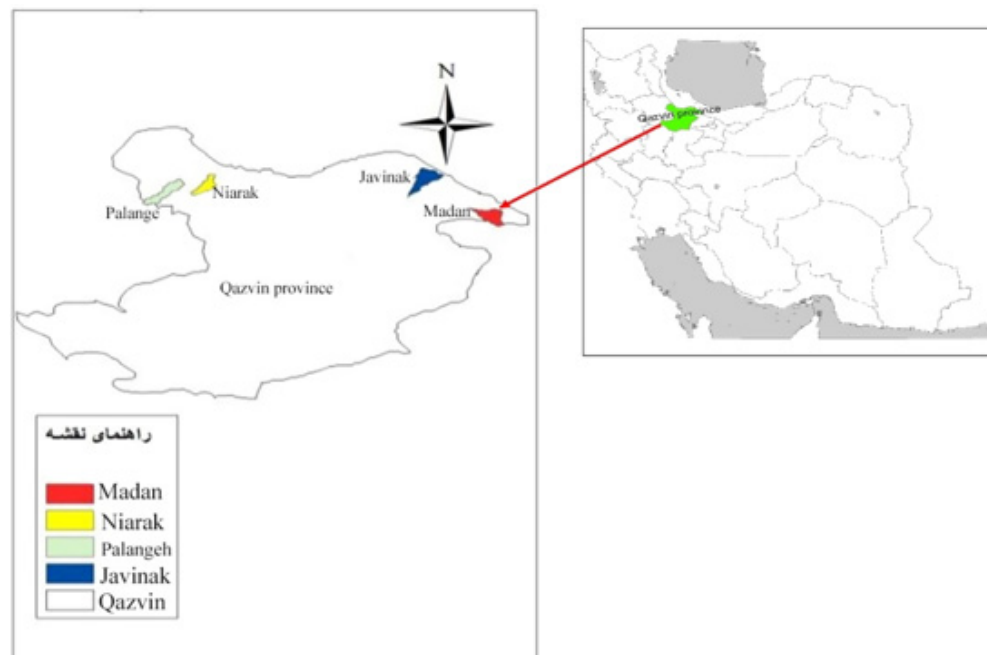

Figure 1. Four studied areas locations in the country and province

Table 1. Characteristics of the four watersheds studied

Area (hec)	Mean Annual Temperature (c)	Mean Annual Precipitation (mm)	Latitude	Longitude	Area
8536.34	9.68	566.52	26″, 27′, 36° to 29″, 36′, 36° northern latitude	41″, 16′, 50° to 08″, 26′, 50° eastern longitude	Juyank
6812.29	6.7	563.5	30″ , 18′, 36° to 30″, 24′, 36° northern latitude	30″ , 32′, 50° to 00″, 52′, 50° eastern longitude	Madan
4855.7	12.4	299.7	0.59, 27′, 36° to 29.1″, 34′, 36°northern latitude	16.2″, 20′, 49° to 8.8″, 27′, 49° eastern longitude	Niyarak
7291.06	10.4	350	57″ 24′ 36° to 02″ 33′ 36° northern latitude	43″ 49′ 07° to 40″ 18′ 49° eastern longitude	Palangeh

A final 101 units were derived based on soil-building factors in the study area. To measure the physicochemical properties of the samples, 101 samples were taken from each work unit in representative areas of that unit.

The correlation of all variables was analyzed using Pearson method and SPSS 20 software. Using SPSS software, variance analysis was conducted at a level of 5% and multivariate linear regression modeling was performed using Enter, Forward, and Stepwise methods.

Each of the used methods will be explained separately: the Enter method, all independent variables were introduced into the model simultaneously to determine the effect of all important and unimportant variables. In this method, all variables are analyzed in a single step in order of minimum tolerance. Since all variables are included in the equation regardless of their correlation coefficient with the dependent variable, the simultaneous method introduces several drawbacks, such as the fact that variables that are not significant are likely to remain in the equation, causing F and R^2 values to decrease as a result. The Forward method begins by calculating the correlation between each independent variable and the dependent variable. Then, the independent variable that explains the largest proportion of the variance in the dependent variable is included in the analysis. After the first variable is separated, the variable with the highest correlation coefficient with the dependent variable is included in the analysis. Like the Forward method, the Stepwise method introduces variables one by one. The first step is to select the variable that has the highest correlation with the dependent variable. Following the separation of the preceding variable, the second variable in the analysis is the one that causes the greatest increase in the R^2 coefficient. The main difference between this method and the Forward method is that in the Forward method, the variables entered into the analysis remain in the equation, whereas in the Stepwise method, with the entry of a new variable, the variables already entered into the equation are retested to determine whether or not they still contribute to the success of the equation. Consequently, some variables that had high explanatory power in the first step may be removed in the second step.

In all three models, soil formation factors were considered independent variables, while soil physical and chemical factors were considered dependent variables. Finally, these models were validated using explanatory coefficients and mean square errors.

Results and Discussion

In linear multivariate regression modeling, three approaches were explored, including Enter, Forward, and Stepwise. Therefore, five factors of soil formation (Figure 2) were considered independent variables, and physical and chemical factors of soil were considered dependent variables to investigate the relationship between these factors and the physical and chemical properties of the soil (Table 2).

Table 2. Multivariate linear modeling of soil parameters with three methods and validation

| Dependent variable	Method	Model	R^2	RMSE
pH	Enter	pH=0.08V-0.005Rain-0.059Temperature +0.0001Slope+0.0001Height+0.077hgeo+9.223	0.78	0.56
	Stepwise	pH=-0.005Rain-0.051Temperature +0.076hgeo+8.915	0.78	0.56
	Forward	pH=-0.005Rain-0.051Temperature +0.076hgeo+8.915	0.78	0.56
Electrical conductivity	Enter	EC=0.005Hgeo-0.023Temperature +0.0001Rain+0.001Slope+0.0001Height+0.832	0.31	0.23
	Stepwise	-	-	-
	Forward	-	-	-
Clay percentage	Enter	Clay=3.714Temperature +0.84Rain+106Slope-10.707V-1.910-86.141	0.47	12.38
	Stepwise	Clay=0.083Rain+3.890Temperature +0.024Height-1.948Hgeo-10.469V-87.544	0.467	12.37
	Forward	Clay=0.083Rain+3.890Temperature +0.024Height-1.948Hgeo-10.469V-87.544	0.467	12.37
Silt percentage	Enter	silt=0.510Temperature -3.439Hgeo-0.029Rain-0.123Slope+0.001Height+47.501	0.33	9.85
	Stepwise	silt=7.171V-3.524Hgeo+47.990	0.27	9.91
	Forward	silt=7.171V-3.524Hgeo+47.990	0.27	9.91
Sand percentage	Enter	Sand=3.743Hgeo-3.576Temperature -0.070Rain-0.119Slope-0.022Height+5.024V+142.368	0.54	15.77
	Stepewise	Sand=3.691Hgeo-0.061Rain-3.920Temperature -0.22Height+143.060	0.56	15.73
	Forward	Sand=3.691Hgeo-0.061Rain-3.920Temperature -0.22Height+143.060	0.56	15.73
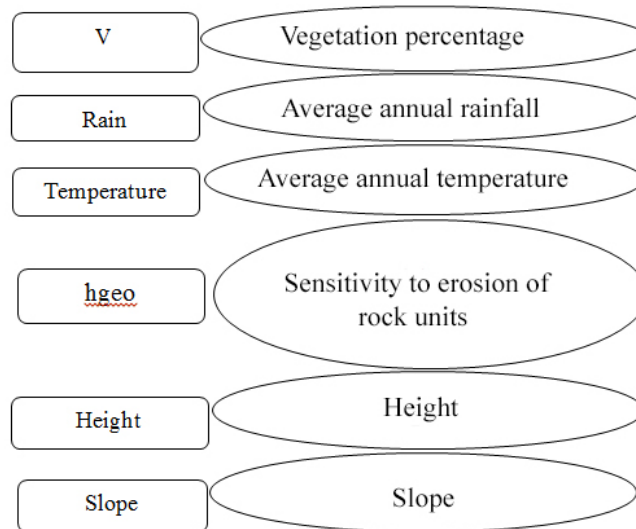
Figure 2. Symbols in regression

For acidity, all three models showed almost the same statistics, indicating that all three models can be used. Nevertheless, the Stepwise model calculates acidity faster and easier because it has fewer parameters. A three-factor model includes temperature, precipitation, and sensitivity of rock to erosion, all of which can be readily accessed. The results of this study agree with those of Thomas et al (1999), who noted that stratigraphy may explain the pH of a region. Rain removes alkaline elements such as calcium, magnesium, and potassium from the soil, while acidic elements such as hydrogen, aluminum, and manganese remain. Therefore,

rain-soaked soils are generally more acidic than dry soils in areas with a lot of rain. Although this is not always the case, other factors can also affect soil pH. As for acidity and parent material, the soil is also acidic if the main rock from which it was formed is acidic, such as granite. Furthermore, soils formed with limestone become alkaline (Zhang et al., 2019). In the long run, however, soil texture determines how much rain affects the soil. The organic matter in clay and loamy soils is more abundant, and their elements are more cohesive than those in sandy soils. Therefore, these soils are more resistant to the leaching of alkaline elements by rain (Kaiser et al., 1992). Precipitation-induced acidification of soil does not occur in a single growing season and requires hundreds of years of heavy precipitation. In sandy soils, pH changes more rapidly because the soil elements are less tightly bound together and water moves faster. The distance between soil particles allows the water to flow from the soil particles and wash the soil elements along with it. Clay soil with poor drainage enables less water to pass through soil particles, which reduces leaching (Rostaei et al., 2021). Concerning electrical conductivity, only the Enter method provided a model, and in the other two methods, the parameters were not included. A soil's electrical conductivity depends on the amount of cations. In other words, the more cations there are the higher the electrical conductivity (Salehi, 2013).

Three methods were used to model soil clay percentage. We found that the Stepwise method with a smaller RMSE was more suitable for this estimate based on two statistics. A Stepwise approach was also chosen to calculate silt percentage due to its appropriate statistics and lower number of factors. The model included organic carbon proportions and the erosion sensitivity of rocks. The above method also determined the percentage of sand faster than other approaches. All obtained models showed a positive sign indicating a direct linear relationship, and a negative sign indicating an inverse linear relationship. In terms of soil texture, clay and silt levels decreased at higher heights, whereas sand levels increased. The percentage of soil particles greatly affects vegetation. For instance, species are more likely to occur in soils that have more silt, since silt causes more moisture to be stored near the roots, while sand increases the permeability and dries the soil rapidly (Shaksteband et al., 2017).

Based on regression modeling, Su Kim et al. (2020) determined the equations for shear strength and permeability coefficients, concluding that the permeability coefficient was associated with porosity ratio, effective grain size, and uniformity coefficient. Furthermore, shear strength is associated with the ratio of fine-grained soil, uniformity coefficient and degree of saturation, dry weight density, and porosity ratio. Azadpour et al. (2014) evaluated the ability of parametric transfer functions to estimate the absorption branch of the soil water characteristic curve in Boeen Zahra city. They found that using the extracted regression equations caused the RMSE of the transfer functions of Mayer and Jarvis (1999), Westen et al. (1999), Campbell (1985), and Sexton et al. (1986) to decrease by 140, 174, 132, and 113%, respectively. Consequently, it improved the estimate of the absorption branch of the soil water characteristic curve. Bayat et al. (2016) estimated soil thermal diffusivity using regression transfer functions and concluded that the best models for estimating coefficients of thermal diffusivity were models with saturated clay, silt-to-sand ratio, and hydraulic conductivity. Ostvari et al (2020) estimated the tolerable limit of soil loss using linear and tree regression methods. They noted that as a result of the higher determination coefficient in the data, the tree regression method with an estimated average tolerable soil loss limit of 1.08 tons per hectare per year was more efficient than multiple regression with an estimated average tolerable soil loss limit of 1.13 tons per hectare per year.

Kassai & Sisák (2018) used 12,400 soil samples and a 1:100,000 scale geologic map to create a highly accurate soil map by plotting the relationship between geology and soil in Hungry. They considered three important characteristics of soil, including pH, organic matter, and alkaline cations, and noted that soil maps must be created by classifying parent materials geochemically, and climate is more important than geology. Furthermore, Javadi et al. (2018)

studied the southern part of the Urmia Plain and concluded that among the five factors of soil formation, mother material has played a significant role in soil formation. The lime present in all soil types, whether geogenic or pedogenic, is derived from parent material and was found in most formations of the studied watersheds. Moreover, Bayat et al. (2015) concluded that chalk soil-forming consequences are the main ones observed in Cretaceous formations and limestone type's consequences are mainly associated with Neogene formations in Bardsir region. Hagan et al. (2007) also studied the Sawadkoh in Mazandaran and concluded that there were no significant differences between the soil properties in terms of organic matter, organic carbon, potassium, and phosphorus in the formations, whereas other soil properties differed significantly between the various formations.

Conclusion

In this study, we found that soil formation factor maps can be used to prepare maps of soil physicochemical properties in different regions with the aid of more extensive modeling. In addition to speeding up the study, this approach will also help to propose and implement management actions and decisions more quickly and accurately, since soil-building factors and physiochemical properties are interconnected.

Competing interests

This article has no financial conflict of interest

Declaration of Funding

This research did not receive any specific funding

Data Availability

The data that support this study are available in the article and accompanying online supplementary material.

Declarations

All authors have read, understood, and have complied as applicable with the statement on "Ethical responsibilities of Authors" as found in the Instructions for Authors.

Author Contribution statement

The first author wrote the article, the second author edited it, and the third and fourth authors provided guidance.

References

Araujo, M. A., Zinn, Y. L., Lal, R., 2017. Soil parent material, texture and oxide contents have little effecton soil organic carbon retention in tropical highlands. Geoderma, 300: 1-10.

Azadpour, A., Rasulzadeh, A., Azizi M., 2014. Evaluation of the ability of parametric transfer functions to estimate the absorption branch of soil water characteristic curve (case study: Boeen Zahra city). Irrigation science and engineering (scientific-research journal), 39 (3).

Bayat, A., Farpour, M.H., Jafari. a., 2015. Physicochemical, micromorphological and mineralogical properties of clay soils of Bardsir region affected by geological formations, geomorphology and

climate. Water and Soil Journal. 30 (5): 1515-1530.

Bayat, H., Sahebi Hamrah, Sh., Islami, S., 2016. Thermal diffusion coefficient estimation using regression transfer functions. 15th Congress of Soil Sciences of Iran.

Birkeland, P. W., 1999. Soils and geomorphology. 3rd ed. Oxford Univ. Press, New York

Bouma J., 1989. Using soil survey data for quantitative land evaluation Adv. Soil Sci. Soc. Am. J, 9: 177-213.

Buol, S. W., R. J. Southard, R. C. Graham, Mc-Daniel. P. A., 2011. Soil Genesis and Classification. 6th ed. York, 560 p Wiley-Blackwell, New

Bluth, G. J. S., Kump, L. R., 1994. Lithologic and climatologic controls of river chemistry. Geochimica ET Cosmochimica Acta, 58(10): 2341-2359.

Egli M., Merkli Ch., Sartori G., Mirabella A., Plotze M., 2008. Weathering, mineralogical evolution and soil organic matter along a Holocene soil toposequence developed on carbonate-rich materials. Geomorphology, 97:675-696.

Gray, J.M., Bishop, T.F.A., Wilford, J.R., 2016. Lithology and soil relationships for soil modelling and mapping. Catena, 147: 429–440.

Gomes, P. C., Fontes, M. P., Da Silva, A. G., Mendoca, E. S., Netto, A. R., 2001. Selectivity sequence and competitive adsorption of heavy metal by Brazilian soils. Soil Science Society of America Journal, 48: 794-752.

Haghian, A., Rokh Firouz, G., Ghorbani, J., 2017. Comparison of geological formations in terms of soil characteristics and vegetation. 4th Conference on Geology and Environment. Period 4.

Harden, J.W., 1982. A quantitative index of soil development from field descriptions examples from a chronosequence in central California. Geoderma, 28: 1-28.

He, Z. L., Zhang, M. K., Calvert, D. V., Stoffella, P. J., Yong, X. E., Yu, S., 2004. Transport of heavy metals in surface runoff from vegetable and citrus fields. Soil Science Society of American Journal, Academic research library.   https://doi.org/10.2136/sssaj2004.1662

Javadi, H., Sokoti. R., Pazira, A., 2018. The relationship between geological formations and land shape with soil formation and evolution (case study: southern part of Urmia Plain). Journal of Water and Soil Resources Protection. 9 (1): 51-66.

Jenny, H., 1994. Factors of Soil Formation. Dover, New York. p. 281.

Jiang, P., Thelen, K.D., 2004. Effect of soil and topographic properties on crop yield in a northcentral corn-Soybean cropping system. J. Agron. 96: 252-258.

Kaiser, E. A., Mueller, T., Joergensen, R. G., Insam, H., Heinemeyer, O., 1992. Evaluation of methods to estimate the soil microbial biomass and the relationship with soil texture and organic matter. Soil Biology and Biochemistry, 24(7), 675-683.

Kassai, P., Sisák, I., 2018. Role of geology in the spatial prediction of soil properties in the watershed of Lake Balaton, Hungary. Journal of the Croatian Geological Survey and the Croatian Geological Society. 71/1 :29-39.

Nael, M., Khademi, H., Jalalian, A., Schulin, R., Kalbasi, M., Sotohian, F., 2009. Effect of geo-pedological conditions on the distribution and chemical speciation of selected trace elements in forest soils of western Alborz, Iran. Geoderma, 152:157-170.

Navidi, N., Abtahi, A., 2001. Effects of climate and topography in forest soils genesis Khirrod Kenar of Nowshahr in Mazandaran province. Journal of Soil and Water Science, 15: 299-316.

Ortiz, M., Simon, C., Dorronsoro, F., Marti, N., Garcia, I., 2002. Soil evolution over the quaternary period in a mediterranean climate (SE Spain). Catena, 48: 131-148.

Ostwari, Y., Mousavi, A., Mozafari, H., 2019. Calculation of the tolerable limit of soil loss using linear and tree regression methods. Water and Soil Journal (Agricultural Sciences and Industries). 34 (1): 173-179.

Phillips, J.D., 2001. Divergent evolution and the spatial structure of soil landscape variability. Catena. 43: 101-113.

Raushi, S., Safadoost, A., Mossadeghi, M.R., 2021. Effect of soil texture and improvement materials on some soil properties. The first national conference of community-oriented researches in agriculture. Natural resources and environment. Hamedan.

Saketeband, F., Shahab Arkhazlou, H., Ghorbani, A. Mohammadi Moghadam, S., 2017. The effect of altitude on physical and chemical properties of soil. The first international conference on organic agriculture. Moghadas Ardabili University.

Salehi, A., 2013. An investigation of the changing physical and chemical properties of soil in Khash Namkhaneh of Khairudkanar forest based on the composition of tree cover and topographical factors. PhD Thesis, Faculty of Natural Resources, University of Tehran. p 187.

Su Kim, K., Kim, M., Se L., Hwang, E., 2020. Regression Equations for Estimating Landslide-Triggering Factors Using Soil Characteristics. Appl. Sci. 10: 3560.   https://doi.org/10.3390/app10103560

Thanachit, S., Suddhiprakarn, A., Kheoruenromne, I., Gilkes, R. J., 2006. The geochemistry of soils on a catena on basalt at Khon Buri, northeast Thailand. Geoderma, 135: 81-96.

Thomas, A.L., Dambrinc, E., King, D., party, J.P., Probst, A., 1999. A spatial study of the relationships between streamwater acidity and geology, soils and relief. Journal of HYDROLOGY 217: 35-45.

Vingiani, S., Terribile, F., Meunier, A., Petit, S., 2010. Weathering of basaltic pebbles in a red soil from Sardinia: A microsite approach for the identification of secondary mineral phases. Catena, 83: 96-106.

Zhang, Y.-Y. Wu, W., Liu, H., 2019. Factors affecting variations of soil pH in different horizons in hilly regions.