



## Investigating soil grouping using conventional and modern clustering models in some parts of Qazvin plain

Zahra Rasaei<sup>1</sup> | Fereydoon Sarmadian<sup>2✉</sup> | Azam Jafari<sup>3</sup>

1. Soil Science Department, Faculty of Agricultural, College of Agriculture and Natural Resources, University of Tehran, Karaj, Iran. E-mail: [zahra.rasaei@ut.ac.ir](mailto:zahra.rasaei@ut.ac.ir)
2. Corresponding Author, Soil Science Department, Faculty of Agricultural, College of Agriculture and Natural Resources, University of Tehran, Karaj, Iran. E-mail: [fsarmad@ut.ac.ir](mailto:fsarmad@ut.ac.ir)
3. Soil Science Department, Faculty of Agriculture, Shahid Bahonar University of Kerman, Kerman, Iran. E-mail: [a.jafari@uk.ac.ir](mailto:a.jafari@uk.ac.ir)

Article Info	ABSTRACT
<p><b>Article type:</b> Research Article</p> <p><b>Article history:</b></p> <p><b>Received:</b> May. 12, 2024</p> <p><b>Revised:</b> June. 15, 2024</p> <p><b>Accepted:</b> July. 21, 2024</p> <p><b>Published online:</b> Oct. 2024</p> <p><b>Keywords:</b> Co-Clustering, Digital Soil Mapping, Hierarchical Clustering.</p>	<p>Soil is a crucial component in achieving sustainable development goals due to its significant role in addressing environmental challenges. It is essential to differentiate soils that have similar management requirements. This necessity has prompted soil scientists to employ numerical classification models to categorize soils based on their similarities. In this study, we utilized two types of clustering models, traditional and modern, to classify soils from certain areas of the Qazvin Plain. Using one-way and two-way clustering models, we grouped 297 soils from the region based on a comprehensive set of their morphological, physicochemical, and environmental attributes. The classifications derived from these two models were assessed using internal and external evaluation indicators, with the distribution map of soil subgroups serving as a ground truth reference map. The results indicated that the hierarchical clustering model, with a lower Davis-Bouldin index (DB: 1.38) and a higher adjusted Rand index (ARI: 0.49), outperformed the biclustering model. However, the classifications from the bidirectional clustering model corresponded reasonably well with the topographical and soil changes in the region, as evidenced by the higher Shannon's difference index in the bidirectional clustering model (1.82) compared to the hierarchical clustering model (1.62). Overall, the study's findings underscore the utility of the co-clustering model as a contemporary data mining technique for soil classification and identification of soil management similarity patterns.</p>

Cite this article: Rasaei, Z., Sarmadian, F., & Jafari, A. (2024) Investigating soil grouping using conventional and modern clustering models in some parts of Qazvin plain, *Iranian Journal of Soil and Water Research*, 55 (8), 1273-1295. <https://doi.org/10.22059/ijswr.2024.376397.669711>

© The Author(s).

Publisher: The University of Tehran Press.



DOI: <https://doi.org/10.22059/ijswr.2024.376397.669711>



## EXTENDED ABSTRACT

### Introduction

The classification of soils is crucial for their proper management and identification. Soil scientists have recognized the importance of numerical classification models for soil grouping. Creating continuous maps using digital mapping models allows for a better understanding of soil class distribution, aiding in improved soil management. Traditional clustering methods group soils based on their distinct properties, while two-way clustering methods group soils within subsets of similar characteristics. This study aims to compare the effectiveness of one-way and two-way clustering models in soil grouping and in identifying relationships between different soils.

### Materials and Methods

This study utilized a dataset of 297 soil samples from some parts of the Qazvin plain. A broad spectrum of morphological, physicochemical, and environmental variables was used for soil grouping. The hierarchical clustering method was employed for one-way soil grouping, and the two-way clustering method was used for co-clustering of them. The Davis-Bouldin (DB) index was used to evaluate the groupings obtained from these models based on the degree of soil separation and intra-group variance. The Adjusted Rand Index (ARI) was used for external evaluation of groupings, with the distribution map of soil subgroups serving as a ground truth reference. Shannon's entropy index was used to assess the efficiency of these models in representing soil variability in the study area.

### Results and Discussion

The study found that both models were successful in differentiating the region's soils based on topographical and physiographic unit changes. However, the two-way clustering model demonstrated a slightly different pattern in soil separation, particularly in the central and southern parts of the study area. A numerical comparison of the results showed that the one-way clustering model provided better soil separation and less variance (DB: 1.38), and was more congruent with the distribution map of soil subgroups in the region (ARI: 0.49). The two-way clustering model effectively represented the pattern of soil changes in the study area, as evidenced by a higher Shannon index (1.82) compared to the hierarchical clustering model (1.62).

### Conclusion

Although numerical comparative evaluations of the groupings reveal the superior efficiency of the hierarchical clustering model in separating soil groups, the two-way clustering model successfully grouped the region's soils according to their changes and the region's physiographic changes. This model also effectively represented soil changes in the region, as indicated by a higher Shannon entropy index. The study's findings affirm the efficacy of the two-way clustering model as a modern data mining technique in identifying similar soils and, consequently, in their grouping and modeling in digital soil mapping studies. The use of this model is recommended for examining soils in different parts of the country.

### Author Contributions

Conceptualization, Z.R., F.S. and A.J.; methodology, Z.R., and A.J.; software, Z.R.; validation, Z.R., and A.J.; formal analysis, Z.R.; investigation, Z.R., F.S. and A.J.; resources, Z.R. and F.S.; data curation, F.S.; writing—original draft preparation, Z.R.; writing—review and editing, Z.R., F.S. and A.J.; visualization, Z.R., F.S. and A.J.; supervision, F.S.; project administration, F.S.; funding acquisition, Z.R. All authors have read and agreed to the published version of the manuscript.

### Data Availability Statement

Not applicable.

### Acknowledgements

This work is based on research funded by the Iran National Science Foundation (INSF) under project No. 4024343. The first author would like to express her appreciation to the INSF for its financial support of this project, and to the University of Tehran for providing the conditions necessary for the completion of this project.

### Ethical considerations

The study was approved by the Ethics Committee of the University of ABCD (Ethical code: IR.UT.RES.2024.500). The authors avoided data fabrication, falsification, plagiarism, and misconduct.

## بررسی گروه‌بندی خاک با استفاده از مدل‌های خوشه‌بندی مرسوم و مدرن در بخش‌هایی از دشت قزوین

زهرا رسائی<sup>۱</sup> | فریدون سرمدیان<sup>۲</sup> | اعظم جعفری<sup>۳</sup>۱. گروه علوم و مهندسی خاک، دانشکده کشاورزی، دانشکدگان کشاورزی و منابع طبیعی، دانشگاه تهران، کرج، ایران. رایانامه: [zahra.rasaei@ut.ac.ir](mailto:zahra.rasaei@ut.ac.ir)۲. نویسنده مسئول، گروه علوم و مهندسی خاک، دانشکده کشاورزی، دانشکدگان کشاورزی و منابع طبیعی، دانشگاه تهران، کرج، ایران. رایانامه: [fsamad@ut.ac.ir](mailto:fsamad@ut.ac.ir)۳. گروه علوم خاک، دانشکده کشاورزی، دانشگاه شهید باهنر کرمان، کرمان، ایران. رایانامه: [a.jafari@uk.ac.ir](mailto:a.jafari@uk.ac.ir)

## اطلاعات مقاله

## چکیده

نوع مقاله: مقاله پژوهشی

تاریخ دریافت: ۱۴۰۳/۲/۲۳

تاریخ بازنگری: ۱۴۰۳/۳/۲۹

تاریخ پذیرش: ۱۴۰۳/۴/۳۱

تاریخ انتشار: آبان ۱۴۰۳

## واژه‌های کلیدی:

خوشه‌بندی دوطرفه،

خوشه‌بندی سلسله مراتبی،

نقشه‌برداری رقومی خاک.

خاک به عنوان یکی از اجزای اصلی دستیابی به اهداف توسعه پایدار، نقش مهمی در مدیریت مسائل محیط زیستی دارد. بنابراین، تفکیک خاک‌های با نیازهای مدیریتی مشابه ضروری می‌باشد. این امر باعث شده است که دانشمندان خاک‌شناسی از مدل‌های طبقه‌بندی عددی برای گروه‌بندی خاک‌ها بر اساس میزان شباهت آن‌ها استفاده کنند. از میان مدل‌های کمی ارائه شده در این زمینه، مطالعه حاضر دو مدل خوشه‌بندی مرسوم و مدرن را برای گروه‌بندی خاک‌های اراضی قسمت‌هایی از دشت قزوین بکار برده است. بدین منظور، ۲۹۷ خاک‌رخ مطالعه شده در منطقه بر اساس طیف گسترده‌ای از ویژگی‌های ریخت‌شناسی، فیزیکی-شیمیایی و محیطی آن‌ها با استفاده از مدل‌های خوشه‌بندی یک‌طرفه و دوطرفه مورد گروه‌بندی قرار گرفتند. گروه‌بندی‌های بدست آمده از این دو مدل بر اساس شاخص‌های ارزیابی درونی و بیرونی (با در نظر گرفتن نقشه توزیع زیرگروه‌های خاک به عنوان نقشه مرجع واقعیت زمینی) مورد بررسی قرار گرفت. نتایج نشان داد، مدل خوشه‌بندی سلسله مراتبی با میزان کم‌تر شاخص دوپس-بولدین (DB: ۱/۳۸) و افزایش میزان شاخص رند تعدیل شده (ARI: ۰/۴۹) نسبت به مدل خوشه‌بندی دوطرفه کارایی بهتری دارد. با این حال، گروه‌بندی‌های بدست آمده از مدل خوشه‌بندی دوطرفه به میزان قابل قبولی با تغییرات پستی و بلندی و تغییرات خاک‌ها در منطقه تطابق دارند. این امر با میزان شاخص تفرق شانن بیشتر در مدل خوشه‌بندی دوطرفه (۱/۸۲) نسبت به مدل خوشه‌بندی سلسله مراتبی (۱/۶۲) تایید می‌شود. بطور کلی یافته‌های این پژوهش، بر استفاده از مدل خوشه‌بندی دوطرفه به عنوان یک مدل داده کاوی مدرن در گروه‌بندی خاک‌ها و یافتن الگوی تشابه مدیریتی آن‌ها تاکید دارند.

استناد: رسائی؛ زهرا، سرمدیان؛ فریدون، جعفری؛ اعظم، (۱۴۰۳) بررسی گروه‌بندی خاک با استفاده از مدل‌های خوشه‌بندی مرسوم و مدرن در بخش‌هایی از دشت قزوین

، مجله تحقیقات آب و خاک ایران، ۵۵ (۸)، ۱۲۹۵-۱۲۷۳. <https://doi.org/10.22059/ijswr.2024.376397.669711>

© نویسندگان.

ناشر: مؤسسه انتشارات دانشگاه تهران.

DOI: <https://doi.org/10.22059/ijswr.2024.376397.669711>

## مقدمه

خاک به عنوان یکی از اجزای اصلی دستیابی به اهداف توسعه پایدار نقش مهمی در مدیریت مسائل زیست‌محیطی دارد. بنابراین تهیه الگوی دقیق پراکنش ویژگی‌ها و کلاس‌های خاک که به‌خوبی جنبه‌های مختلف مدیریتی خاک را در بر بگیرد؛ می‌تواند در این زمینه راهگشا باشد (Bouma et al., 2019). تمایل برای یافتن ارتباط بین خاک‌ها و گروه‌بندی آن‌ها همواره موضوع جالب توجه مسئولین و کاربران برای شناسایی و مدیریت آن‌ها بوده است. در این راستا، آگاهی از تنوع و تغییر خاک‌ها و الگوی کلاس‌بندی آن‌ها می‌تواند مدیریت بهتر کلاس‌های خاک با ویژگی‌های مشترک و مشابه را راحت‌تر و مؤثرتر سازد (Edokpayi et al., 2017). این امر خاک‌شناسان را بر آن داشته است تا در پی یافتن راه‌های مناسب برای تفکیک و گروه‌بندی خاک‌ها بر اساس میزان شباهت آن‌ها به یکدیگر بر اساس ویژگی‌های مختلف (ریخت‌شناسی، ژنتیکی، آزمایشگاهی و موقعیت قرارگیری آن‌ها در زمین‌نما) برآیند تا بتوانند راحت‌تر مورد بررسی و مدیریت قرار بگیرند (Lee et al., 2023).

تعریف سامانه‌های طبقه‌بندی تاکسونومیکی از جمله راهبردهای پیشنهادی از سوی جامعه علوم خاک آمریکا در این زمینه بوده است که در آن خاک‌ها بر اساس ویژگی‌های مختلف ریخت‌شناسی و فیزیکی-شیمیایی آن‌ها (بسته به دقت مطالعه) در سطوح مختلف تاکسونومیکی کلاس‌بندی می‌شدند (Arnold, 2006; Sonn et al., 2019). بطورکلی، این طبقه‌بندی خاک‌ها با استفاده از توصیف مشخصات خاک‌ها از طریق مشاهده چشمی و تعیین ویژگی‌های فیزیکوشیمیایی هر لایه خاک با هدف درک واضح ویژگی‌های خاک و کمک به مدیریت خاک از طریق تفسیر خاک انجام می‌شود (Lee et al., 2023).

با پیشرفت علم، روش‌های زمین‌آماري<sup>۱</sup> با بکارگیری مدل‌های ریاضی برای مطالعه توزیع و گروه‌بندی عددی خاک‌ها مورد توجه قرار گرفته‌اند که خروجی آن‌ها قابل ارزیابی بوده و صحت و کارایی گروه‌بندی‌ها قابل بررسی می‌باشند. در این روش، خاک‌ها عمدتاً بر اساس ویژگی‌های مختلف آن‌ها از جمله ویژگی‌های فیزیکی شیمیایی مورد گروه‌بندی قرار می‌گیرند. این در حالی است که تمامی نمونه‌ها به تمامی ویژگی‌ها مرتبط و وابسته نیستند. به منظور در نظر گرفتن این اصل، روش خوشه‌بندی دوسویه تعریف و ارائه شده است (Hartigan, 1972). در این روش برخلاف روش‌های خوشه‌بندی معمول که فقط نمونه‌ها (در اینجا خاک‌ها) را بر اساس ویژگی‌های آن‌ها مورد آنالیز قرار می‌دهند، در خوشه‌بندی دوبعدی<sup>۲</sup> هم نمونه‌ها (متغیرهای کلاسی خاک) و هم ویژگی‌های مختلف مربوط به آن‌ها (متغیرهای مستقل) گروه‌بندی می‌شوند. سپس نمونه‌ها (خاک‌ها) بر اساس شباهت زیر گروه‌های با ویژگی‌های یکسان گروه‌بندی می‌شوند (Pontes et al., 2015).

با توجه به ضرورت دستیابی به کلاس‌های خاک مشابه به منظور مدیریت بهتر آن‌ها، مطالعه حاضر با استفاده از دو مدل طبقه‌بندی عددی مرسوم و مدرن سعی در یافتن خاک‌های مشابه و کلاس‌بندی آن‌ها دارد. بدین منظور، مدل‌های خوشه‌بندی عددی یک جهته (مرسوم) و در دو جهت (نوبین) برای گروه‌بندی خاک‌ها بر اساس ویژگی‌های مختلف آن‌ها مورد استفاده قرار می‌گیرند. بدین ترتیب، علاوه بر گروه‌بندی خاک‌ها، کارایی مدل خوشه‌بندی دوطرفه برای حل و مدیریت طبقه‌بندی خاک‌ها برای اولین بار مورد بررسی و ارزیابی قرار می‌گیرد. این تحقیق در قسمتی از دشت قزوین به عنوان یکی از دشت‌های مهم برای تولید انواع محصولات کشاورزی کشور انجام شده است. اراضی این منطقه مطابق با پتانسیل و استعداد خاک‌ها مورد استفاده قرار نمی‌گیرند. بنابراین آگاهی از نحوه پراکنش خاک‌های منطقه و تعیین گروه‌بندی‌های خاک‌های با نیازهای مدیریتی یکسان در قالب نقشه‌های خاک می‌تواند به عنوان یکی از راهکارهای ضروری در راستای مدیریت بهتر این منابع و استفاده بهینه از خاک‌ها برای انواع استفاده‌ها تلقی گردد.

## پیشینه پژوهش

### پیشینه نظری

مدل‌های کمی کلاس‌بندی خاک‌ها بر اساس اندازه‌گیری فاصله عددی بین خاک‌ها پایه‌گذاری شده و اقدام به تفکیک آن‌ها در گروه‌های با ویژگی‌های مشابه‌تر می‌کنند. به‌طوری که طبقه‌بندی عددی خاک‌ها شامل مجموع فرآیندهایی است که سعی در گروه‌بندی خاک‌ها به منظور دستیابی به خالص‌ترین کلاس‌ها از نظر ویژگی‌های مختلف خاکی و مدیریتی آن‌ها دارد؛ بطوریکه خاک‌های موجود در هر گروه بیشترین شباهت را با یکدیگر داشته باشند (Carré and Jacobson, 2009; Láng et al., 2013). از این رو در بسیاری از مطالعات کمی

و پردازش‌های عددی انجام شده برای کلاس‌بندی خاک‌ها، از شاخص‌های تشابه مبتنی بر مفهوم فاصله استفاده می‌شود (Carré and Jacobson; Láng et al., 2013; Hughes et al., 2014). ساده‌ترین و معمول‌ترین شاخص فاصله، فاصله اقلیدسی است که اساس و پایه بسیاری از تکنیک‌های گروه‌بندی سنتی خاک‌ها است (Liberti et al., 2012). بر اساس اندازه‌گیری فاصله و شباهت، الگوریتم‌های خوشه‌بندی نظارت نشده و نظارت شده مختلفی برای گروه‌بندی مشاهدات از جمله خاک ارائه شده است.

بررسی‌ها بیانگر کاربرد گسترده این مدل‌های کمی داده‌کاوی در یافتن ارتباط بین خاک‌ها و گروه‌بندی آن‌ها می‌باشد. از این میان می‌توان به مدل خوشه‌بندی سلسله‌مراتبی (Seaton et al., 2021; Zolfaghari et al., 2019; Egbueri, 2023; Shelia and Hoogenboom, 2020)، میانگین‌های کاممونی (Sunori et al., 2022; Xu et al., 2021; Zhao et al., 2023; Dunkl and Ließ, 2022; Heil et al., 2019) و فازی (Ließ, 2022; Heil et al., 2019) و فازی (Cebeci and Yildiz, 2015; Heil et al., 2019; Zhao et al., 2023) درخت‌های تصمیم (Taghizadeh-Mehrjardi et al., 2015; Adhikari et al., 2014) و غیره برای یافتن ارتباط بین خاک‌ها و دستیابی به گروه‌های خاک‌های مشابه با اهداف مدیریتی مختلف در گروه‌بندی خاک‌ها بر اساس فاصله اقلیدسی بین خاک‌ها اشاره کرد. که نتایج بیانگر کارایی خوب این روش‌ها در گروه‌بندی خاک‌ها می‌باشند.

تجزیه و تحلیل خوشه‌بندی دوطرفه یک روش مفید برای کشف الگوهای منسجم محلی پنهان در یک ماتریس داده است. بر خلاف روش‌های خوشه‌بندی مرسوم، که گروه‌هایی از الگوهای منسجم را با استفاده از کل مجموعه ویژگی‌ها جستجو می‌کند، روش خوشه‌بندی دوسویه الگوی همزمان را در هر دو جهت ردیف و ستون در یک ماتریس داده انجام می‌دهد. این روش به طور خاص برای مدیریت مجموعه‌های داده با تعداد زیادی متغیر و داده‌های با ابعاد بالا طراحی شده است، که امکان خوشه‌بندی همزمان سطرها (نمونه‌ها) و ستون‌ها (ویژگی‌ها) را فراهم می‌کند (Noronha et al., 2022; Chakraborty and Vardeman, 2022) که به خوبی قادر به مدیریت پایگاه داده با داده‌های گم‌شده می‌باشد (Li et al., 2020). با توجه به بررسی‌های منابع در دسترس، این روش بطور گسترده در بیوانفورماتیک و گروه‌بندی نمونه‌های ژن و میکروآرایه مورد استفاده قرار گرفته است (Xie et al., 2019; Pontes et al., 2015; Aouabed et al., 2021; de França et al., 2021). این در حالی است که محققین به تازگی از این روش در حوزه‌های دیگر از جمله سید خرید، متن‌کاوی، تجزیه و تحلیل زیست‌پزشکی، تجزیه و تحلیل نمودار، سامانه‌های توصیه‌بازاریابی، اقتصاد، تجزیه و تحلیل داده‌های ورزشی، الگوهای مصرف آب و غیره بهره‌جسته‌اند (Orzechowski et al. 2019; Xie et al., 2019; Chakraborty and Vardeman, 2022). تا آنجا که به کاربرد آن در علوم محیطی از جمله کشاورزی مربوط می‌شود، این روش به تازگی در مطالعات بوم‌شناختی و کشاورزی برای گروه‌بندی محصولات کشاورزی و تخمین عملکرد گونه‌های مختلف اجرا شده است (Pham et al., Chakraborty and Vardeman, 2022; Aubert et al., 2021; 2022). با این حال، بر اساس دانش ما، استفاده از روش‌های خوشه‌بندی دوطرفه در مطالعات خاک و زمین مورد بررسی قرار نگرفته‌اند یا به طور گسترده در مقالات موجود مستند نشده‌اند.

### پیشینه تجربی

مطالعات زیادی در سرتاسر جهان از روش خوشه‌بندی سلسله‌مراتبی در گروه‌بندی خاک‌ها با اهداف متفاوت در مطالعات مختلف بهره‌برده‌اند. به عنوان مثال، Zolfaghari et al. 2019 از روش خوشه‌بندی سلسله‌مراتبی برای گروه‌بندی خاک‌ها به منظور دستیابی به کلاس‌های خاک با مدیریت مشابه در قالب نقشه شدت بیابان‌زایی در منطقه سیستان استفاده کردند. با توجه به نتایج حاصله، ایشان به اهمیت و ضرورت آنالیز خوشه‌بندی در یافتن خاک‌های مشابه با نیازهای مدیریتی مشترک تأکید کردند. (Edokpayi et al. 2017) از یک الگوریتم خوشه‌بندی سلسله‌مراتبی برای گروه‌بندی خاک‌ها بر اساس مواد اولیه و ویژگی‌های شیمیایی خاص استفاده کرد. مقایسه گروه‌بندی‌های کمی حاصل با نقشه وراثت خاک، سطح تطابق پایینی را بین این دو نقشه نشان داد. (Tchorbagiieff et al., 2019) در مقابل، توافق خوبی بین گروه‌های حاصل از خوشه‌بندی سلسله‌مراتبی با واحدهای ریخت‌شناسی زمین منطقه مورد مطالعه خود پیدا کردند. (Seaton et al. 2021) به طور مشابه بر قابلیت و کارایی مدل خوشه‌بندی برای دستیابی به گروه‌های خاک با نیازهای مدیریتی مشابه بر اساس ویژگی‌های مختلف خاک سطحی در ترکیب با شاخص‌های کیفیت خاک تأکید کردند. ایشان همچنین بر کیفیت بالاتر این گروه‌بندی‌ها نسبت به طبقه‌بندی‌های ژنتیکی و نقشه‌های مرسوم خاک اشاره کردند. (Egbueri 2023) از روش خوشه‌بندی سلسله‌مراتبی در ترکیب با مدل‌های تعلیم ماشین از جمله شبکه عصبی مصنوعی در بررسی فرسایش خاک و حساسیت زمین‌لغزش در جنوب شرقی نیجریه استفاده کردند. رسائی و همکاران (۱۳۹۹) برای ارزیابی دقت واحدهای نقشه خاک مرسوم از روش خوشه‌بندی سلسله‌مراتبی به

همراه چند روش طبقه‌بندی عددی دیگر بهره جستند. نتایج این مطالعه بیان‌گر کارایی خوب مدل‌های کمی مورد استفاده از جمله مدل خوشه‌بندی سلسله مراتبی در شناسایی و تفکیک گروه‌های مناسب با خاک‌های مشابه می‌باشد. سیستانی و همکاران (۱۳۹۶) از روش خوشه‌بندی سلسله مراتبی برای تعیین منشأ فلزات سنگین خاک‌های مجاور صنایع فولاد کرمان استفاده کردند. اگرچه گزارش مستندی مبنی بر کاربرد روش خوشه‌بندی دوطرفه در مطالعات خاک در دسترس نمی‌باشد، بررسی‌ها بیان‌گر کاربرد این مدل در سایر حوزه‌های مشابه از جمله کشاورزی، منابع طبیعی و اکولوژی می‌باشد. به عنوان مثال، (Jhariya and Singh, 2020) از روش خوشه‌بندی دوطرفه در ارزیابی تأثیر رژیم‌های مختلف بر روی جامعه گیاهی همراه با تغییرات در وضعیت ریزمغذی‌های خاک استفاده کردند. Liu et al. (2021) از روش‌های خوشه‌بندی یک‌طرفه و دوطرفه برای آشکارسازی الگوهای مکانی-زمانی و عوامل کنترل‌کننده ژئوشیمی آب‌های زیرزمینی استفاده کردند. در روشی مشابه، (Iqbal et al., 2018) از روش‌های خوشه‌بندی یک‌طرفه و دوطرفه برای طبقه‌بندی جامعه‌شناسی گیاهی علف‌های هرز یک الگوی کشت زراعی-اکولوژیکی استفاده کردند.

## روش‌شناسی پژوهش

### منطقه مورد مطالعه

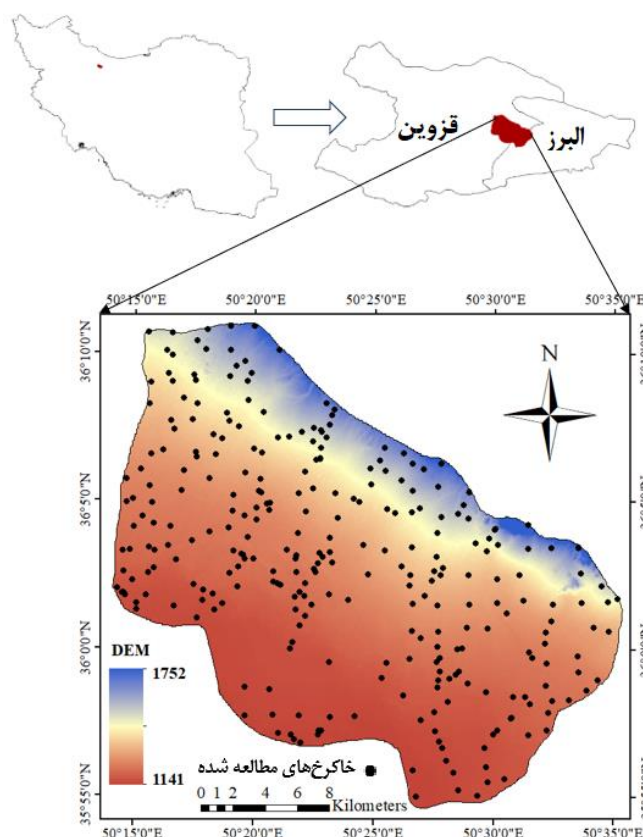
این مطالعه در بخشی از دشت قزوین در مرز استان‌های قزوین و البرز به مساحت تقریبی ۶۰۰۰۰ هکتار انجام شد (شکل ۱). منطقه مورد نظر بین عرض‌های جغرافیایی ۳۵ درجه و ۵۵ دقیقه تا ۳۶ درجه و ۱۰ دقیقه شمالی و طول‌های جغرافیایی ۵۰ درجه و ۱۵ دقیقه تا ۵۰ درجه و ۳۵ دقیقه شرقی قرار دارد. ارتفاع منطقه از ۱۱۴۱ متر تا ۱۷۵۲ متر بالای سطح دریا متغیر است. با حرکت از قسمت‌های شمالی به جنوبی منطقه میانگین دمای سالانه از ۱۳/۳ تا ۱۵/۵ درجه سانتی‌گراد افزایش و بارندگی از ۳۱۴/۲ به ۲۵۳/۵ میلی‌متر کاهش می‌یابد. بر اساس محاسبات انجام شده در نرم‌افزار نیوهال<sup>۱</sup> (Van Wambeke, 2000) خاک‌های منطقه در سه رژیم رطوبتی زیریک خشک، اریدیک ضعیف و آکوئیک و رژیم حرارتی ترمیک قرار دارند (Soil Survey Staff, 2022). مواد مادری تشکیل دهنده خاک‌های بخش‌های شمالی و مرکزی منطقه رسوبات آبرفتی و کوهرفتی شامل سنگ‌های آذرین بازالتی آهکی مربوط به دوره کواترنری می‌باشند که در بخش‌های جنوبی منطقه به کفه‌های گلی و نمکی تغییر می‌یابند. منطقه مورد مطالعه نسبتاً مسطح است و از نظر پستی و بلندی از شمال به جنوب واحدهای شکل زمین مختلف شامل تپه، فلات، دشت دامنه‌ای و اراضی پست در منطقه دیده می‌شوند. شیب در حدود ۷۰ درصد از اراضی منطقه (شامل دشت و اراضی پست) کمتر از ۵ درصد است و در سایر قسمت‌ها (تپه و فلات) تا ۸ درصد می‌رسد. در اثر تغییرات همزمان عامل‌های مهم خاک‌سازی منطقه (پستی و بلندی و اقلیمی) و همچنین رژیم رطوبتی از زیریک به آکوئیک از قسمت‌های بالای منطقه به سمت پایین، طیف وسیعی از خاک‌ها از بدون تکامل خاک‌های شور، گچی و قلیائی در منطقه تشکیل شده‌اند. دشت قزوین یکی از دشت‌های مهم کشور برای تولیدات کشاورزی است. گندم، جو، یونجه و ذرت آبی در کشت زمستانه و تابستانه در دشت‌های میانی منطقه، مرتع غیرشور و کشت دیم در قسمت‌های شمالی و اراضی شور در قسمت‌های جنوبی از مهم‌ترین کاربری‌های منطقه می‌باشند (Mousavi et al., 2022; Mousavi et al., 2023; رحمانی و همکاران، ۱۴۰۱؛ رضائی و همکاران، ۱۴۰۲).

### داده‌های خاک

مطالعات میدانی در مقیاس حد نیمه‌تفصیلی طی یک دوره ۶ ساله (۲۰۱۶ تا ۲۰۲۲) انجام شد. در نهایت، بر اساس ترکیبی از روش‌های معمول نمونه‌برداری از جمله تصادفی، شبکه‌ای و فرا مکعب لاتین مشروط<sup>۲</sup> (cLHS) ۲۹۷ خاک‌رخ مشاهداتی و مته با میانگین فاصله ۴۰۰ تا ۶۰۰ متر در نظر گرفته شد (شکل ۱). نقاط مشخص شده تا عمق ۲۰۰ سانتی‌متری یا بالاتر از لایه محدود کننده حفر و بر اساس راهنمای تشریح صحرائی خاک‌ها تشریح و از کلیه افق‌های ژنتیکی قابل‌شناسایی، نمونه‌برداری شد (Schoeneberger et al., 2012). سپس تمامی ویژگی‌های خاک‌ها در نمونه‌های خاک دست‌نخورده و عبور داده شده از الک ۲ میلی‌متری طبق روش‌های استاندارد اندازه‌گیری شدند (Sparks et al., 2020; Dane and Topp, 2020). از این میان، ۱۰ ویژگی فیزیکی و شیمیایی خاک‌ها شامل درصد سنگریزه و اجزای بافت خاک (شن، سیلت، رس)، کربن آلی، کربنات کلسیم معادل، گچ، اسیدیتته در گل اشباع، هدایت الکتریکی در عصاره گل اشباع و نسبت جذب سدیم در این مطالعه مورد استفاده قرار گرفتند. این داده‌ها از گستره وسیعی از مطالعات پیشین انجام شده در



منطقه (Neyestani et al., 2021; Mousavi et al., 2022; Mousavi et al., 2023; موسوی و همکاران، ۱۴۰۰ (۱،۲،۳)؛ رحمانی و همکاران، ۱۴۰۱؛ خاموشی و همکاران، ۱۴۰۱؛ رضائی و همکاران، ۱۴۰۲؛ خاموشی و همکاران، ۱۳۹۷؛ موسوی و همکاران، ۱۳۹۸؛ ممتازی بروجنی و سرمدیان، ۱۴۰۱) در دسترس می‌باشند.



شکل ۱. محل منطقه مطالعاتی در ایران و در مرز استان‌های قزوین و البرز. محل حفر خاک‌رخ‌های مطالعاتی (دایره‌های سیاه رنگ) به همراه تغییرات ارتفاع (DEM) در منطقه مورد مطالعه نیز نمایش داده شده‌اند.

از آنجایی که خاک‌رخ‌ها در لایه‌های با عمق‌های مختلف بررسی و مطالعه شده‌اند، به منظور هماهنگ کردن لایه‌های تمامی خاک‌رخ‌ها طبق یک استاندارد مشخص، شش عمق استاندارد ۲۰-۰، ۴۰-۲۰، ۶۰-۴۰، ۸۰-۶۰، ۱۰۰-۸۰، ۱۰۰-۱۰۰ و ۱۵۰-۱۰۰ سانتی‌متری در نظر گرفته شد (Nachtergaele, 2023) و مقادیر تمامی ویژگی‌های خاک بر اساس درون‌یابی وزنی توابع خطی (Equal-area spline function) (Malone et al., 2009) بسته aqp (Beaudette et al., 2023) نرم‌افزار R (R Core Team, 2022) بصورت عمودی در عمق کل خاک‌رخ متوسط‌گیری وزنی شدند. بطوری که برای هر ویژگی خاک شش عدد متناظر با شش لایه عمق استاندارد هر خاک‌رخ بدست آمد. از آنجایی که عمق خاک‌رخ‌های مطالعاتی متغیر می‌باشد، برای خاک‌رخ‌های با عمق کم‌تر از ۱۵۰ سانتی‌متری (آخرین عمق استاندارد در نظر گرفته شده)، مقادیر ویژگی‌های خاک تا عمق انتهایی خاک‌رخ میانگین‌گیری گردیدند. بنابراین، برای خاک‌رخ‌های با عمق ۱۰۰ سانتی‌متر یا کم‌تر (مرز پایین آخرین محدوده عمق استاندارد، ۱۵۰-۱۰۰ سانتی‌متری) برای ویژگی‌های خاکی قرار گرفته در این عمق‌ها (که فاقد داده بودند) عدد صفر در نظر گرفته شد.

#### متغیرهای محیطی

با توجه به عامل‌های تشکیل خاک ارائه شده در مدل اسکورپن<sup>۱</sup> (McBratney et al., 2003)، نمایندگان متغیرهای محیطی مختلف شامل پستی و بلندی، پوشش گیاهی، اقلیم و مواد مادری با استفاده از منابع مختلف اطلاعاتی در دسترس تهیه شدند. مشتقات درجه اول و دوم مدل رقومی ارتفاع به عنوان عامل‌های پستی و بلندی در نظر گرفته شدند. بدین منظور، مدل رقومی ارتفاعی آلواس پالسا (<https://asf.alaska.edu>) با دقت تفکیک مکانی ۱۲/۵ متر مورد استفاده قرار گرفت. بر این اساس ۴۵ عدد متغیر

پستی و بلندی-زمین‌ریخت‌شناسی<sup>۱</sup> طبق روش‌های استاندارد (Conrad et al., 2015) در نرم‌افزار SAGA استخراج شدند. همچنین، میانگین ۶ ساله تصاویر ماهواره‌ای لندست ۸ مربوط به فصل‌های بهار و تابستان سال‌های ۲۰۱۳ تا ۲۰۲۳ که با بازه تاریخی نمونه‌برداری از خاک‌ها منطبق می‌باشند، برای استخراج شاخص‌های تصاویر ماهواره‌ای استفاده شد. بدین ترتیب که ابتدا در سامانه Google Earth Engine (Gorelick et al., 2017) کلیه پیش‌پردازش‌های لازم از جمله حذف ابر، تصحیحات رادیومتر و بهبود کنتراست انجام شد. سپس ۲۷ شاخص مختلف با استفاده از روش‌های استاندارد (Index DataBase, 2021) تهیه شدند. نقشه مواد مادری با رومی‌سازی نقشه‌های زمین‌شناسی ۱:۱۰۰۰۰۰ موجود در منطقه (سازمان زمین‌شناسی و اکتشاف معادن ایران) و نقشه ژئوform از روی هم انداختن نقشه‌های زمین‌شناسی و شکل زمین در نرم‌افزار ArcGIS تهیه شدند. همچنین، نقشه‌های میانگین سالانه دما و بارندگی با استفاده از داده‌های اقلیمی ماهانه ۲۴ ایستگاه هواشناسی در منطقه مورد مطالعه با درون‌یابی بر اساس روش درون‌یابی تجربی کریجینگ بی‌زی<sup>۲</sup> با قدرت تفکیک مکانی ۲۵۰ متر در نرم‌افزار ArcGIS تهیه شدند (Mousavi et al., 2022؛ Mousavi et al., 2023؛ موسوی و همکاران، ۱۴۰۰؛ رحمانی و همکاران، ۱۴۰۱).

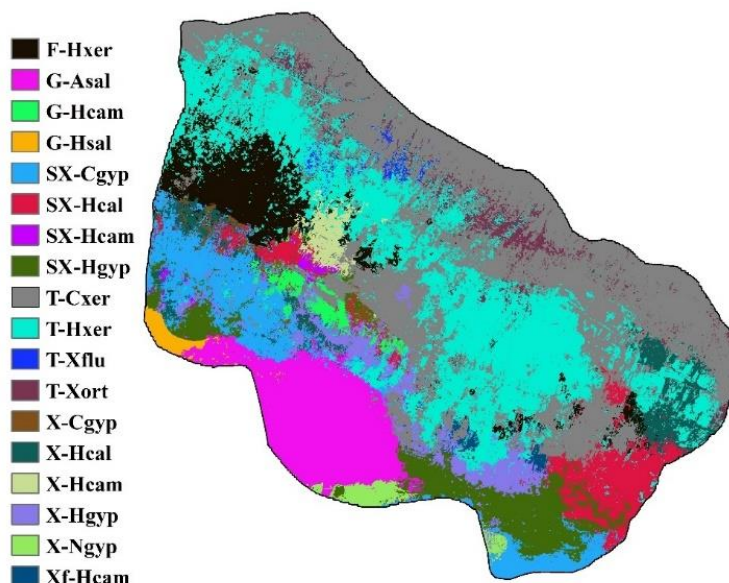
### رده‌بندی تاکسونومیکی خاک‌ها

همان‌طور که در مطالعات قبلی انجام‌شده در منطقه مورد مطالعه نیز آورده شده است (خاموشی و همکاران، ۱۳۹۷؛ موسوی و همکاران، ۱۳۹۸؛ ممتازی بروجنی و سرمیدیان، ۱۴۰۱؛ Neyestani et al., 2021)، خاک‌ها طبق اطلاعات ریخت‌شناسی (افق‌های مشخصه سطحی و زیرسطحی، رژیم‌های رطوبتی و حرارتی) و آزمایشگاهی آن‌ها و با در نظر گرفتن معیارهای کلید طبقه‌بندی آمریکایی خاک (Soil Survey Staff, 2022) تا سطح فامیل رده‌بندی شدند. بدین ترتیب، خاک‌ها در ۳ رده، ۷ زیر رده، ۱۱ گروه بزرگ، ۱۸ زیرگروه و ۴۲ فامیل قرار دارند. در این مطالعه، سطوح مختلف تاکسونومیکی خاک‌ها به عنوان نقشه پایه و مرجع (نماینده واقعیت‌های زمینی) برای ارزیابی گروه‌بندی‌های بدست آمده از مدل‌های خوشه‌بندی مدنظر قرار گرفتند. به منظور رعایت اختصار، در اینجا فقط زیرگروه‌های خاک‌ها به تفصیل در جدول ۱ آورده شده‌اند، همچنین، نقشه پراکنش هریک از کلاس‌های خاک در منطقه در شکل ۲ نمایش داده شده است.

جدول ۱. زیرگروه‌های مختلف تاکسونومیکی خاک‌های مطالعه شده به همراه درصد فراوانی هریک در منطقه مورد مطالعه.

شماره	زیرگروه	علامت اختصاری	درصد فراوانی	ویژگی مشخصه
۱	Fluentic Haploxerepts	F-Hxer	۷/۴۱	افق کمبیک، فلوتتیک، رژیم رطوبتی زیریک
۲	Gypsic Aquisalids	G-Aqsal	۳/۳۷	افق‌های جیبسیک و سالیک، رژیم رطوبتی آکوئیک
۳	Gypsic Haplocambids	G-Hcam	۲/۰۲	افق‌های جیبسیک و کمبیک، رژیم رطوبتی اریدیک
۴	Gypsic Haplosalids	G-Hsal	۲/۰۲	افق‌های جیبسیک و سالیک، رژیم رطوبتی اریدیک
۵				افق‌های جیبسیک و کلسیک، رژیم‌های رطوبتی اریدیک و زیریک، سودیک
	Sodic Xeric Calcigypsis	SX-Cgyp	۵/۷۲	
۶	Sodic Xeric Haplocalcids	SX-Hcal	۵/۳۹	افق کلسیک، رژیم‌های رطوبتی اریدیک و زیریک، سودیک
۷	Sodic Xeric Haplocambids	SX-Hcam	۱/۰۱	افق کمبیک، رژیم‌های رطوبتی اریدیک و زیریک، سودیک
۸	Sodic Xeric Haplogypsis	SX-Hgyp	۴/۷۱	افق جیبسیک، رژیم‌های رطوبتی اریدیک و زیریک، سودیک
۹	Typic Calcixerepts	T-Cxer	۲۴/۹۲	افق کلسیک، رژیم رطوبتی زیریک
۱۰	Typic Haploxerepts	T-Hxer	۱۵/۸۲	افق کمبیک، رژیم رطوبتی زیریک
۱۱	Typic Xerofluvents	T-Xflu	۲/۰۲	فلوتتیک، رژیم رطوبتی زیریک
۱۲	Typic Xerorthents	T-Xor	۶/۷۳	رژیم رطوبتی زیریک، اورتننت
۱۳	Xeric Calcigypsis	X-Cgyp	۲/۳۶	افق‌های جیبسیک و کلسیک، رژیم‌های رطوبتی اریدیک و زیریک
۱۴	Xeric Haplocalcids	X-Hcal	۶/۷۳	افق کلسیک، رژیم‌های رطوبتی اریدیک و زیریک
۱۵	Xeric Haplocambids	X-Hcam	۲/۰۲	افق کمبیک، رژیم‌های رطوبتی اریدیک و زیریک
۱۶	Xeric Haplogypsis	X-Hgyp	۴/۳۸	افق جیبسیک، رژیم‌های رطوبتی اریدیک و زیریک
۱۷	Xeric Natrigypsis	X-Ngyp	۲/۰۲	افق‌های جیبسیک و ناتریک، رژیم‌های رطوبتی اریدیک و زیریک
۱۸	Xerofluentic Haplocambids	Xf-Hcam	۱/۳۵	افق کمبیک، فلوتتیک، رژیم‌های رطوبتی اریدیک و زیریک





شکل ۲. نقشه پراکنش زیرگروه‌های مختلف تاکسونومیکی خاک‌های مطالعه شده در منطقه مورد مطالعه. برای اطلاع از نام کامل هر یک از کلاس‌های خاک به جدول ۱ مراجعه بفرمائید.

همان‌طور که مشاهده می‌شود، حدود ۳۲٪ از منطقه را خاک‌های آهکی با افق مشخصه غالب کلسیک تشکیل می‌دهند (T-Cxer و X-Hcal) که بیشتر در قسمت‌های شمالی و مرکزی منطقه پراکنده شده‌اند. با کاهش شیب منطقه با حرکت از قسمت شمالی منطقه به سمت جنوب و با تغییر رژیم رطوبتی خاک (از زریک به اریدیک و سپس اکوئیک در جنوبی‌ترین قسمت منطقه، در اراضی پست) لایه‌های با تجمع آهک که عمده‌ترین مشخصه خاک‌ها در مناطق بالاتر بودند، با لایه‌های تجمع گچ، نمک، و با درصد بالای سدیم ترکیب و جایگزین می‌شوند که منجر به پیدایش خاک‌های شور، گچی و سدیمی در این قسمت شده است. عمده تجمع این خاک‌ها در اراضی پست (در جنوبی‌ترین قسمت منطقه)، جنوب غرب و جنوب شرق منطقه دیده می‌شوند. این خاک‌ها روی هم رفته بالغ بر ۳۳ درصد خاک‌های منطقه را به خود اختصاص می‌دهند (جدول ۱). حدود ۳۵ درصد مابقی قسمت‌ها را خاک‌های کم عمق و یا با تکامل کم خاکرخ (F-Hxer، T-Hxer، T-Xflu، T-Xort، Xf-Hcam و X-Hcam) تشکیل می‌دهند که عمدتاً در قسمت‌های شمالی و مرکزی منطقه پراکنده هستند (شکل ۲). تفسیر بیشتر علت و نحوه تشکیل خاک‌های منطقه در سایر مطالعات از جمله رحمانی و همکاران (۱۴۰۱) و ممتازی و همکاران (۱۴۰۱) در دسترس می‌باشد.

#### تهیه پایگاه داده

در نهایت تمامی این نقشه‌ها با استفاده از روش نمونه‌برداری مجدد<sup>۱</sup> به قدرت تفکیک مکانی ۱۵ متر در ۱۵ متر تغییر داده شدند و برای تمامی نقشه‌ها سامانه مختصات مرجع متریک (WGS 1984, UTM 39N) تعریف شد. با روی هم انداختن موقعیت جغرافیایی خاکرخ‌ها و این نقشه‌ها، مقادیر این متغیرهای محیطی در موقعیت جغرافیایی خاکرخ‌ها استخراج شدند.

همچنین، علاوه بر این ویژگی‌های خاکی و متغیرهای محیطی، ریخت‌شناسی و تاکسونومی متمایز کننده خاک‌ها بر اساس زیرگروه‌های مختلف خاک در منطقه مورد مطالعه، دوازده مشخصه شامل رژیم‌های رطوبتی (زریک، اریدیک و اکوئیک)، افق‌های مشخصه (کلسیک، چپسیک، کمبیک، سالیک و ناتریک)، ویژگی‌های مشخصه (سودیک و فلوتیک)، زیر رده (اورتنت) و انقطاع بافتی که نماینده بیشترین تفاوت در خاک‌ها می‌باشند انتخاب شدند (جدول ۱). با فرض تساوی وزن تمامی ویژگی‌ها، خاک‌های مختلف بر اساس حضور یا عدم حضور این موارد بصورت صفر (عدم حضور) و یک (حضور) کدگذاری شدند.

در نهایت، به منظور انتخاب بهترین زیر مجموعه از متغیرهای خاکی و محیطی برای مدل‌سازی و گروه‌بندی خاک‌ها، از همبستگی بین این متغیرها و خاک‌ها در محل ۲۹۷ خاکرخ مطالعاتی با استفاده از بسته FSelector نرم‌افزار R (Romanski et al., 2023) استفاده شد.

### گروه‌بندی خاک‌ها

پارامتر مهمی که باید در گروه‌بندی خاک‌ها در نظر گرفته شود، تعداد بهینه کلاس است (Weatherill and Burton, 2009; Kassambara., 2017). برای این منظور می‌توان از شاخص‌های مختلفی استفاده کرد (Kassambara., 2017). در این پژوهش از روش آرنج<sup>۱</sup> (Ketchen and Shook, 1996) برای انتخاب تعداد خوشه مناسب استفاده شد که با هدف اصلی روش‌های خوشه‌بندی که تفکیک خوشه‌هایی با حداقل تغییرات درون خوشه‌ای است، هم‌راستا می‌باشد. این روش به مجموع تغییرات درون خوشه‌ای به عنوان تابعی از تعداد خوشه‌ها نگاه می‌کند؛ بدین معنی که تعدادی از خوشه‌ها را انتخاب می‌کند تا جایی که با افزودن یک خوشه دیگر، میزان تغییرات درون خوشه‌ای کاهش نیابد (Kassambara., 2017).

در نهایت، از آنجایی که این ویژگی‌های خاکی و محیطی بر اساس واحدهای مختلف اندازه‌گیری شده‌اند، به منظور از بین بردن اثر مقیاس، ویژگی‌های مورد بررسی با استفاده از رابطه ۱ استاندارد شدند (Everitt et al., 2011).

$$\frac{x_i - \text{center}(x)}{\text{scale}(x)} \quad \text{رابطه ۱}$$

که  $x_i$ : متغیرهای مورد مطالعه،  $\text{center}(x)$ : میانگین مقادیر متغیرها و  $\text{scale}(x)$ : واریانس متغیرها می‌باشند.

سپس، ماتریس فاصله بین خاک‌ها با توجه به ویژگی‌های خاکی-محیطی انتخاب و استاندارد شده و همچنین ویژگی‌های تاکسونومیکی متناظر با آن‌ها با استفاده از مفهوم فاصله تهیه گردید. پس از به دست آمدن گروه‌های خاک، مدل جنگل تصادفی به روش اعتبارسنجی متقاطع<sup>۲</sup> برای تهیه نقشه توزیع مکانی کلاس‌های خاک در منطقه استفاده شد. مدل‌سازی این کلاس‌ها بر اساس رابطه بین آن‌ها و متغیرهای محیطی انتخاب شده انجام شد.

### مدل خوشه‌بندی سلسله مراتبی<sup>۳</sup>

این الگوریتم ساده‌ترین و پایه‌ای‌ترین الگوریتم در فرآیند داده‌کاوی نمونه‌ها از جمله خاک می‌باشد. این مدل یک روش تجزیه و تحلیل آماری است برای توسعه مجموعه‌ای از خوشه‌ها که با جفت شدن متوالی متغیرها ایجاد می‌شوند، که در نهایت کلاس‌های همگن خاک را بر اساس ویژگی‌های آن‌ها شناسایی می‌کند. بدین ترتیب که خاک‌ها را طی مراحل مختلف با یکدیگر ترکیب کرده و کلاس‌های بزرگ‌تری ایجاد می‌کند. این مدل سعی در یافتن خوشه‌های همگن با حداکثر شباهت درون کلاسی و حداکثر تفاوت با سایر کلاس‌ها دارد. بدین صورت که یک خاک با سایر خاک‌های درون آن کلاسی که به آن تعلق گرفته است کم‌ترین فاصله و با خاک‌های بیرون از کلاس حداکثر تفاوت را از نظر ویژگی‌های مختلف داشته باشد. برای تعیین شباهت بین خاک‌ها از ماتریس فاصله استفاده می‌شود (Everitt et al., 2011; Zirnea et al., 2013; Edokpayi et al., 2017). روش‌های مختلفی برای تعیین فاصله وجود دارد که در این مطالعه از فاصله اقلیدسی استفاده شده است (Liberti et al., 2012).

$$d_{\text{euc}} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad \text{رابطه ۲}$$

که در آن،  $d_{\text{euc}}$ : فاصله اقلیدسی بین دو مشاهده  $X = X_1, X_2, \dots, X_m$  و  $Y = Y_1, Y_2, \dots, Y_m$  می‌باشد. که در آن مجموع تفاضل بین مقادیر یک متغیر در دو مشاهده  $X$  و  $Y$  محاسبه می‌گردد. در این مطالعه مدل خوشه‌بندی سلسله مراتبی تجمعی با استفاده از تابع  $\text{hclust}$  در نرم‌افزار  $R$  مورد استفاده قرار گرفت.

### مدل خوشه‌بندی دوطرفه<sup>۴</sup>

خوشه‌بندی دوطرفه یا مستقیم در سال ۱۹۷۲ برای حل مسئله خوشه‌بندی پایگاه‌های داده بزرگ و همچنین خوشه‌بندی همزمان ردیف و ستون‌های داده توسط هارتیگان ارائه شد که امروزه به عنوان زیر مجموعه‌ای از روش‌های داده‌کاوی در رویکردهای خوشه‌بندی مدرن در نظر گرفته می‌شود (Hartigan, 1972). هدف خوشه‌بندی دوطرفه شناسایی زیرگروه‌هایی از مشاهدات است که در زیرمجموعه‌ای از متغیرها شباهت دارند در حالی که در متغیرهای باقی‌مانده متمایز هستند (Pontes et al., 2015; Kaiser, 2011; Xie et al., 2019). روش‌های خوشه‌بندی دوطرفه در زمینه‌های مختلف مورد مطالعه و استفاده قرار گرفته‌اند. به طوری که محققان این روش‌ها را برای کاوش مجموعه داده‌های پیچیده و پرداختن به پرس و جوهای علمی خاص، ارائه توانایی کشف همزمان خوشه‌های ردیف و خوشه‌های ستون،

1. Elbow method  
2. K-fold cross-validation Random Forest model  
3. Hierarchical clustering  
4. Co-clustering or Biclustering

ارائه بینش و کشف روابط ارزشمند در داده‌ها مفید یافته‌اند (Noronha et al., 2022; Chakraborty and Vardeman, 2022). اگر فرض شود که  $B$  یک دو خوشه است که در آن  $I$  مجموعه‌ای از  $|I|$  نمونه را تشکیل می‌دهد و  $J$  یک مجموعه از  $|J|$  عوارض باشد بطوریکه  $b_{ij}$  نشان‌دهنده رابطه بین ردیف  $i$  و ستون  $j$  است،  $B$  را می‌توان به صورت شکل ۳ بیان کرد. اگر یک ماتریس داده  $n \times m$  داشته باشیم که در آن  $A$  متشکل از مشاهدات  $X$  و متغیرهای  $Y$ ، با  $a_{ij}$  نشان‌دهنده عناصر موجود در سلول‌ها باشد، تجزیه و تحلیل خوشه‌بندی دوطرفه به دنبال شناسایی زیر گروه‌های مشاهدات بصورت  $A_{IJ}$  است (شکل ۳).

	$y_1$	...	$y_i$	...	$y_m$
$x_1$	$a_{11}$	...	$a_{i1}$	...	$a_{m1}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
$x_j$	$a_{1j}$	...	$a_{ij}$	...	$a_{mj}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
$x_n$	$a_{1n}$	...	$a_{in}$	...	$a_{mn}$

شکل ۳. نمایش شماتیک یک ماتریس خوشه‌بندی دوطرفه

طیف وسیعی از روش‌ها برای این روش خاص ارائه شده‌اند که عمدتاً در زمینه پایگاه‌های داده در مقیاس بزرگ اعمال می‌شوند (Pontes et al., 2015). در اینجا، ما گروه‌بندی خاک‌ها را با استفاده از الگوریتم پایه CC (Cheng and Church, 2000) از بسته biclust در نرم‌افزار R (Kaiser et al., 2023) تهیه می‌کنیم. گروه‌بندی‌ها با در نظر گرفتن آستانه  $0/0.1$  به عنوان حداکثر امتیاز پذیرفته شده (دلتا:  $0/0.1$ ) با در نظر گرفتن تعداد کلاس‌های بهینه انجام می‌شود.

### ارزیابی مدل‌های گروه‌بندی

کارایی مدل‌های گروه‌بندی بر اساس معیارهای مختلف بصورت داخلی و خارجی مورد ارزیابی قرار می‌گیرد. در روش‌های ارزیابی درونی اغلب گروه‌بندی‌ها بر اساس ماتریس فاصله بین آن‌ها مورد ارزیابی قرار می‌گیرند. این روش اعتبارسنجی فقط مجموعه داده و کلاس‌ها را به عنوان ورودی می‌گیرد و از اطلاعات ذاتی داده‌ها برای ارزیابی کیفیت خوشه‌بندی استفاده می‌کند. به‌طور کلی اغلب شاخص‌های ارزیابی داخلی مدل‌های خوشه‌بندی بر مبنای منسجم بودن و تفکیک شدن خوب گروه‌ها از هم و کاهش واریانس درون گروهی استوار می‌باشند. این در حالی است که شاخص‌های ارزیابی بیرونی بر اساس مجموعه داده واقعی انجام می‌شوند. بدین صورت که با در نظر گرفتن گروه‌بندی‌های مرجع، میزان شباهت و همبستگی بین نتایج خوشه‌بندی‌ها با واقعیت‌های زمینی مرجع را محاسبه می‌کنند (Handl et al., 2005).

### شاخص دیویس-بولدین<sup>۱</sup>

این شاخص امکان ارزیابی داخلی کیفیت خوشه‌بندی‌ها با مقایسه میانگین شباهت بین خوشه‌های دوتایی مشابه را می‌دهد و به عنوان میانگین شباهت هر خوشه با خوشه‌ای که بیشترین شباهت به آن را دارد، محاسبه می‌شود. در این زمینه، شباهت به عنوان نسبت بین فاصله بین خوشه‌ای و درون خوشه‌ای تعریف می‌شود. به این ترتیب، این شاخص، خوشه‌های به خوبی تفکیک شده با پراکندگی کمتر را به عنوان دارای امتیاز بهتر رتبه‌بندی می‌کند. بطور کلی، این شاخص بیانگر میانگین شباهت بین هر کلاس و نزدیک‌ترین کلاس همسایه اش می‌باشد (Davies and Bouldin, 1979; Handl et al., 2005).

$$DB = \left(\frac{1}{n}\right) \sum \max (R_{ij}) \quad \text{رابطه ۳}$$

که  $n$  بیانگر تعداد کلاس‌ها و  $R_{ij}$  میزان فاصله (عدم شباهت) بین کلاس  $i$  و کلاس  $j$  با بیشترین شباهت با این کلاس است. این شاخص بر اساس فاصله اقلیدسی بین مراکز گروه‌های خاک‌ها و با کمک تابع  $index.DB$  در بسته  $clusterSim$  نرم‌افزار R (Walesiak, 2023) محاسبه گردید.

1. Davies-Bouldin index (DB)

### شاخص تفرق شانن<sup>۱</sup>

این شاخص یکی از پرکاربردترین شاخص‌ها برای اندازه‌گیری تفرق مبتنی بر فراوانی نسبی کلاس‌های خاک مورد مطالعه می‌باشد که به عنوان شاخص شانن غیریکنواختی یا انتروپی شناخته می‌شود. رابطه ۵ معادله ریاضی اندازه‌گیری این شاخص را نشان می‌دهد (Ibanez et al., 1995; Saldana and Ibanez, 2004).

$$H' = - \sum_{i=1}^n p_i \cdot \ln p_i \quad \text{رابطه ۴}$$

که  $p_i$  نسبت اجزای موجود در واحد  $i$  ام است که از نسبت  $n_i$  (تعداد اجزای متعلق به واحد  $i$  ام) به  $N$  (تعداد کل اجزای مورد مطالعه) بدست می‌آید. این شاخص بر اساس کلاس‌های خاک در سطوح مختلف تاکسونومیک (رده، زیر رده، گروه بزرگ و زیرگروه) در گروه‌های خاک بدست آمده از مدل‌های خوشه‌بندی مورد بررسی بدست آمد. همچنین، نقشه توزیع شاخص شانن برای گروه‌بندی‌های حاصل از این دو مدل تهیه گردید.

### شاخص رند تعدیل‌شده<sup>۲</sup>

امتیاز رند یک معیار ارزیابی بیرونی مدل‌های خوشه‌بندی است که با در نظر گرفتن همه جفت‌های نمونه و با استفاده از تعداد جفت‌های نمونه که در کلاس‌های مشابه یا متفاوت در کلاس اصلی و پیش‌بینی‌شده تخصیص داده شده‌اند، معیار تشابه بین دو کلاس است. مقدار ۱ را برای کلاس‌های یکسان و صفر را برای کلاس‌بندی تصادفی برمی‌گرداند. شاخص رند تعدیل‌شده (ARI) گونه‌ای از شاخص رند است که با استفاده از مدل جایگشت<sup>۳</sup> برای کلاس‌بندی‌ها از نظر شانن تصحیح می‌شود.

$$ARI = \frac{RI - E(RI)}{1 - E(RI)} \quad \text{رابطه ۴}$$

که در آن  $E(RI)$  مقدار مورد انتظار  $RI$  تحت مدل جایگشت است. مقدار این شاخص بین -۱ تا ۱ متغیر می‌باشد. مقدار ۱ نشان می‌دهد که کلاس‌ها یکسان هستند، در حالی که مقدار صفر نشان می‌دهد که کلاس‌ها به طور تصادفی و مستقل از یکدیگر ترسیم شده‌اند. این شاخص با استفاده از تابع  $ARI$  در بسته  $aricode$  نرم‌افزار  $R$  (Chiquet et al., 2023) محاسبه شد. برای این منظور، نقشه پراکنش زیرگروه‌های خاک به‌عنوان واقعیت‌های زمینی مورد استفاده قرار گرفت. علت این امر این است که زیرگروه‌های تاکسونومیک با توجه به مشاهدات عینی و واقعی شخص کارشناس خبره از منطقه مورد مطالعه، تغییرات خاک‌ها، ویژگی‌ها و مشخصات مورفولوژیکی خاک‌ها و با در نظر گرفتن اندازه‌گیری‌های فیزیکی و شیمیایی آن‌ها انجام شده است. بنابراین بازتاب خوب و قابل قبولی از واقعیت‌های موجود در منطقه می‌باشد.

## نتایج و بحث

نتایج بررسی‌های آماری ویژگی‌های فیزیکی و شیمیایی خاک‌ها در ۲۹۷ خاکرخ مطالعاتی در جدول ۲ آورده شده است. از آنجایی که اندازه‌گیری‌های عمق استاندارد در عمق ۰-۱۵۰ سانتی‌متری خاکرخ‌ها صورت گرفته است، این آنالیز نیز برای میانگین‌های عمقی در ۱۵۰ سانتی‌متری بالای خاک‌ها محاسبه شده است.

تجزیه و تحلیل آماری الگوی قابل توجهی از تغییرات بالا را در تمام متغیرهای خاک در منطقه مورد مطالعه نشان می‌دهد. تمامی ویژگی‌های خاکی بررسی شده به استثنای اسیدیته، مقادیر بسیار بالای ضریب تغییرات ( $CV$  بیش‌تر از ۳۵٪) دارند (Wilding, 1985). هدایت الکتریکی بیشترین میزان تغییرات را نشان می‌دهد و گچ، نسبت جذب سدیم و درصد سنگریزه به ترتیب در رتبه‌های بعدی قرار دارند. این در حالی است که اسیدیته کم‌ترین میزان تغییرپذیری در خاک را نشان می‌دهد ( $CV$  کم‌تر از ۱۵٪). سنگریزه، هدایت الکتریکی، گچ و نسبت جذب سدیم بالاترین میزان چولگی و کشیدگی را نشان می‌دهند که بیانگر توزیع غیرنرمال آن‌ها با یک پیک تیز است (Hair et al., 2022). تغییرپذیری بالای این ویژگی‌های خاک تأکیدی بر تعامل عوامل پیچیده و متعدد، از جمله فرآیندهای تشکیل خاک و شیوه‌های مدیریتی مختلف خاک‌ها می‌باشد (Soropa et al., 2021; Selmy et al., 2022). از طرفی، تغییرپذیری کمتر در اسیدیته خاک ممکن است مربوط به ظرفیت بافری این ویژگی خاک باشد (Selmy et al., 2022).

جدول ۲. نتایج آنالیز آماری ویژگی‌های فیزیکی و شیمیایی عمق ۰-۱۵۰ سانتی‌متری ۲۹۷ خاکرخ مطالعاتی

ویژگی خاک	واحد	حداقل	حداکثر	میانگین	انحراف معیار	ضریب تغییرات (%)	چولگی	کشیدگی
سنگریزه	درصد	۰/۰	۸۰/۰	۱۰/۴	۱۸/۳	۱۷۶/۳	۱/۸	۲/۱
شن	درصد	۶/۵	۸۸/۲	۴۲/۶	۱۹/۴	۴۵/۵	۰/۱	-/۸
سیلت	درصد	۵/۳	۷۰/۷	۲۹/۴	۱۰/۹	۳۷/۳	۰/۲	-/۱
رس	درصد	۳/۳	۶۴/۸	۲۷/۹	۱۳/۶	۴۸/۸	۰/۶	-/۴
اسیدیته	-	۷/۳	۹/۲	۷/۹	۰/۳	۳/۹	۱/۰	۱/۶
هدایت الکتریکی	دسی‌زیمنس بر متر	۰/۲	۱۱۶/۴	۶/۵	۱۶/۱	۲۴۶/۹	۴/۵	۲۲/۳
کربن آلی	درصد	۰/۰	۱/۷	۰/۵	۰/۳	۵۱/۴	۱/۳	۲/۸
کربنات کلسیم	درصد	۱/۲	۵۱/۸	۱۵/۳	۸/۷	۵۷/۰	۱/۱	۱/۵
گچ	درصد	۰/۰	۲۸/۹	۲/۱	۴/۶	۲۲۰/۰	۳/۰	۱۰/۶
نسبت جذب سدیم	-	۰/۸	۱۱۶/۸	۱۰/۰	۱۸/۴	۱۸۳/۷	۳/۰	۹/۷

فرآیند میانگین‌گیری ویژگی‌های خاکی در ۶ عمق مختلف و افزودن متغیرهای محیطی (پستی و بلندی، سنجش از دور و اقلیمی) در محل هر یک از خاکرخ‌ها به همراه معیارهای تاکسونومیکی در نهایت به یک پایگاه داده با اندازه ۲۹۷ در ۱۴۸ (۲۹۷ خاکرخ در ردیف‌ها و ۱۴۸ متغیر مستقل شامل ۱۲ مشخصه ریخت‌شناسی، ۶۰ متغیر خاکی، ۴۵ متغیر پستی و بلندی، ۲۷ متغیر سنجش از دور، ۲ متغیر اقلیمی، ۲ متغیر مواد مادری و شکل زمین در ستون‌ها) منجر گردید.

#### انتخاب ویژگی‌های خاکی و متغیرهای محیطی مهم برای مدل‌سازی

جدول ۳ متغیرهای مهم انتخاب شده بر اساس آنالیز همبستگی انجام شده بین خاکرخ‌ها با مقادیر ویژگی‌های خاکی در عمق‌های مختلف و متغیرهای محیطی را نشان می‌دهد. همانطور که ملاحظه می‌گردد، روی هم رفته ۲۱ متغیر انتخاب شده است که از این میان ۱۲ متغیر مربوط به ویژگی‌های خاکی در عمق‌های مختلف می‌باشند.

جدول ۳. میزان همبستگی بین متغیرهای مهم خاکی و محیطی انتخاب شده مربوط به ۲۹۷ خاکرخ در منطقه مورد مطالعه

عامل خاک‌سازی اسکورین	متغیرهای انتخاب شده	همبستگی	توضیحات	منبع
CaSO4 80-100	۰/۹۴	درصد گچ در عمق ۸۰-۱۰۰ سانتی‌متری	آزمایشگاه	
CaSO4 60-80	۰/۷۴	درصد گچ در عمق ۶۰-۸۰ سانتی‌متری	آزمایشگاه	
EC 40-60	۰/۷۳	میزان هدایت الکتریکی در عمق ۴۰-۶۰ سانتی‌متری	آزمایشگاه	
SAR 60-80	۰/۷۰	میزان نسبت جذب سدیم در عمق ۶۰-۸۰ سانتی‌متری	آزمایشگاه	
EC 100-150	۰/۶۹	میزان هدایت الکتریکی در عمق ۱۰۰-۱۵۰ سانتی‌متری	آزمایشگاه	
SAR 40-60	۰/۶۸	میزان نسبت جذب سدیم در عمق ۴۰-۶۰ سانتی‌متری	آزمایشگاه	
SAR 80-100	۰/۶۸	میزان نسبت جذب سدیم در عمق ۸۰-۱۰۰ سانتی‌متری	آزمایشگاه	
Clay 20-40	۰/۶۵	درصد رس در عمق ۲۰-۴۰ سانتی‌متری	آزمایشگاه	
CaCO3 60-80	۰/۶۰	درصد کربنات کلسیم در عمق ۶۰-۸۰ سانتی‌متری	آزمایشگاه	
CaCO3 80-100	۰/۶۰	درصد کربنات کلسیم در عمق ۸۰-۱۰۰ سانتی‌متری	آزمایشگاه	
Clay 40-60	۰/۵۸	درصد رس در عمق ۴۰-۶۰ سانتی‌متری	آزمایشگاه	
Gravel 60-80	۰/۴۹	درصد سنگریزه در عمق ۶۰-۸۰ سانتی‌متری	آزمایشگاه	
PreSum	۰/۶۸	میانگین بارندگی سالانه	داده‌های اقلیمی	
CTVI	۰/۴۲	شاخص پوشش گیاهی تبدیل و اصلاح‌شده	سنجش از دور	
MrVBF	۰/۶۹	شاخص همواری کف دره با درجه تفکیک بالا	مدل رقومی ارتفاع	
StH	۰/۶۵	ارتفاع استاندارد	مدل رقومی ارتفاع	
DB	۰/۶۲	حوزه زهکشی	مدل رقومی ارتفاع	
MCA	۰/۶۱	مساحت حوزه تعدیل‌شده	مدل رقومی ارتفاع	
ValD	۰/۵۶	عمق دره	مدل رقومی ارتفاع	
Geoform	۰/۶۸	نقشه واحدهای ژئوform	نقشه زمین‌شناسی	
SI	۰/۶۳	شاخص شوری	سنجش از دور	

خاک (s)

اقلیم (c)

پوشش گیاهی (o)

پستی و بلندی (r)

مواد مادری و پستی و بلندی (p,r)

مواد مادری و خاک (p,s)



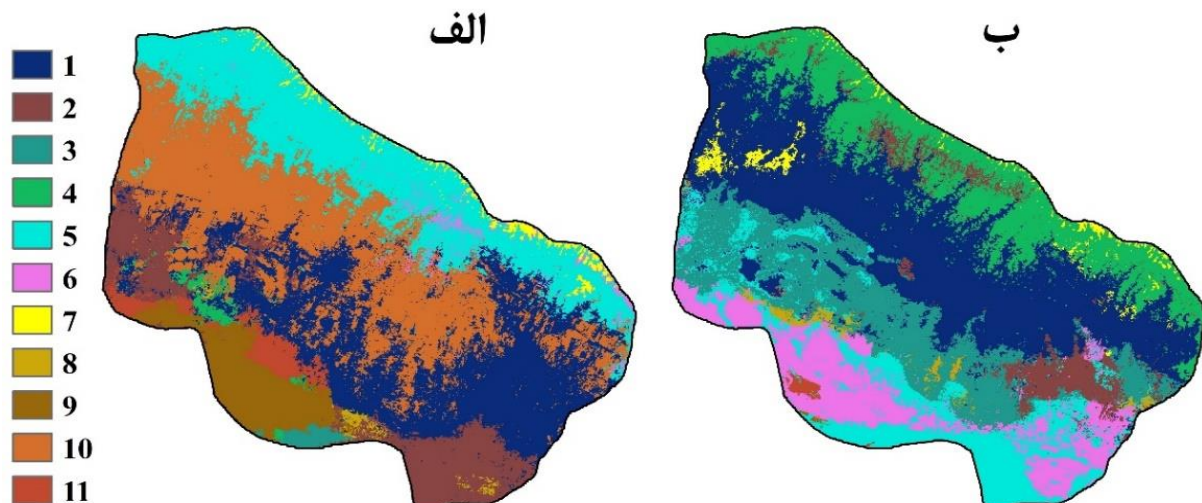
از میان ویژگی‌های خاکی (به تفکیک عمق‌های استاندارد) وارد شده به مدل درصد گچ در عمق‌های ۶۰-۸۰ و ۸۰-۱۰۰ سانتی‌متری، میزان هدایت الکتریکی در عمق ۴۰-۶۰ سانتی‌متری و نسبت جذب سدیم در عمق ۶۰-۸۰ سانتی‌متری در بالاترین درجه اهمیت قرار دارند. ویژگی‌های خاکی انتخاب شده به خوبی تغییرات بین خاک‌ها و تفاوت آن‌ها را نشان می‌دهند و به خوبی بیانگر حضور خاک‌های گچی، آهکی، شور و سدیمی در منطقه مورد مطالعه می‌باشند (شکل ۲). این ویژگی‌ها با خاک‌های منطقه در ارتباط می‌باشند (جدول ۱) که نشان می‌دهند مهم‌ترین تغییرات آن‌ها که باعث تمایز در این خاک‌ها شده است در عمق‌های میانی و پایینی خاک‌ها رخ داده است که متأثر از اقلیم و پستی و بلندی به عنوان دو عامل خاک‌سازی مهم منطقه می‌باشند. بنابراین، این‌ها به عنوان ویژگی‌های خاکی متمایزکننده خاک‌های منطقه در گروه‌بندی آن‌ها مؤثر می‌باشند. همچنین، درصد رس، کربنات کلسیم و سنگریزه از ویژگی‌های مهم متمایزکننده خاک‌های منطقه می‌باشند (جدول‌های ۱ و ۲) که مهم‌ترین تغییرات آن‌ها در عمق‌های میانی و پایینی خاک‌ها می‌باشند. علت این امر را می‌توان به تغییرات کم پستی و بلندی در منطقه دانست که عمده قسمت‌های منطقه بصورت دشت و با درصد شیب کم می‌باشند که باعث تجمع بیشتر این ویژگی‌ها در عمق‌های پایین‌تر خاک‌ها شده است.

سایر متغیرهای انتخاب شده شامل شش متغیر پستی و بلندی، دو متغیر سنجش از دور و یک متغیر اقلیمی می‌باشند (جدول ۳). از میان متغیرهای پستی و بلندی (استخراج شده از مدل رقومی ارتفاع)، شاخص همواری کف دره با درجه تفکیک بالا (MrVBF) بیشترین میزان اهمیت را دارد (همبستگی: ۰/۶۹) که نتیجه‌ای منطقی است زیرا این ویژگی پستی و بلندی، نقشی اساسی در شکل‌دهی خاک و تعیین الگوهای مکانی ویژگی‌های خاک از طریق کنترل آن‌ها بر حرکت آب و رسوب‌گذاری را دارد (Hu et al., 2019; Li et al., 2020; Jafari et al., 2012; Deressa et al., 2018; Rasaei et al., 2020). این عامل در تشکیل خاک‌های دارای تجمع رسوب با افق‌های کلسیک، جیسیک، سالیک در هر دو رژیم رطوبتی خاک Xeric و Aridic اشاره کرده‌اند. همچنین، بسیاری از محققان بر شاخص‌های RS به‌ویژه شاخص شوری به‌عنوان عامل مهم در نقشه‌برداری خاک‌های متأثر از نمک و گچ تأکید کرده‌اند (Taghizadeh-Mehrjardi et al., 2014; Zhang and Huang, 2019; Gad et al., 2021; Naimi et al., 2021; Wang et al., 2024). همان‌طور که ملاحظه می‌شود، متغیرهای انتخاب شده به خوبی عامل‌های مهم خاک‌سازی ارائه شده در مدل اسکورپین را پوشش می‌دهند.

### مدل‌های گروه‌بندی

#### خوشه‌بندی سلسله مراتبی

بر اساس آنالیز آرنج انجام شده (WSS)، تعداد ۱۱ کلاس برای خوشه‌بندی خاک‌ها توسط هر دو مدل در نظر گرفته شد. شکل ۴ الف پراکنش گروه‌های بدست آمده از مدل خوشه‌بندی مرسوم یکطرفه (سلسله مراتبی) و جدول ۴ نتایج ارزیابی‌های کمی این مدل را نشان می‌دهد. همان‌طور که ملاحظه می‌شود، ارزیابی داخلی گروه‌بندی‌های بدست آمده از این مدل به DB نسبتاً کم منجر گردید که نشان می‌دهد که این مدل به خوبی توانسته است خاک‌ها را از هم تفکیک و سپس گروه‌بندی کند (جدول ۴).



شکل ۴. نقشه پراکنش گروه‌های خاک بدست آمده از دو مدل خوشه‌بندی سلسله مراتبی (الف) و دوطرفه (ب). به منظور اطلاع از خاک‌های موجود در هر خوشه به شکل ۵ مراجعه شود.



جدول ۴. نتایج مقایسه کمی نتایج گروه بندی های خاک در دو مدل خوشه بندی سلسله مراتبی (یک طرفه) و دوطرفه.

شاخص	شاخص تفرق شانون				شاخص DB	مدل
	زیر گروه	گروه بزرگ	زیر رده	رده		
ARI	۳/۴۷	۳/۴۰	۳/۲۶	۲/۸۹	۱/۳۸	خوشه بندی یک طرفه
۰/۴۹	۳/۸۴	۳/۷۱	۳/۶۹	۲/۶۲	۳/۶۳	خوشه بندی دوطرفه

نتایج بیانگر هم خوانی گروه بندی های بدست آمده از این مدل با تغییرات پستی و بلندی (شکل ۱) و در نتیجه تغییرات خاکها در منطقه مورد مطالعه است. بطوری که در مناطق بالادست منطقه که تغییرات شیب و پستی و بلندی بارزتر از قسمت های پایین است خاکها را در سه گروه (۵، ۶ و ۷) قرار داده است. این قسمت عمدتاً شامل خاکهای فاقد تکامل خاکرخی یا خاکرخی های با تکامل کم می باشد. این مدل همچنین به خوبی قسمت های میانی منطقه مورد مطالعه را از قسمت های شمالی و جنوبی با پستی و بلندی متفاوت تر، تفکیک کرده است. این در حالی است که در جنوب منطقه با تنوع بیشتر زیر گروه های خاکها (شکل ۲)، این مدل نیز گروه های بیشتری را تفکیک کرده است. همان طور که در شکل ۲ نیز دیده می شود، این قسمت ها حاوی خاک های متأثر از تجمع نمک (گچی، شوری و قلیائی) می باشند که به دلیل تأثیر همزمان پستی و بلندی و همچنین شرایط اقلیمی (بارندگی کمتر و دمای بیشتر) تشکیل یافته اند. به طور کلی، از مقایسه کلی دو نقشه (شکل ۲ و ۴ الف)، به نظر می رسد که مدل خوشه بندی سلسله مراتبی در تفکیک خاکها به تغییرات پستی و بلندی و متغیرهای محیطی متأثر از آن اهمیت داده است.

همچنین، مقایسه کیفی این مدل با توزیع زیر گروه های خاک (شکل ۲) بیانگر یک الگوی مشابه در تفکیک خاکها در این دو مدل می باشد بطوریکه گروه های ۵، ۱۰ و ۹ عمدتاً شامل خاک های T-Cxer، T-Hxer و G-Asal می باشند. همچنین، توزیع زیر گروه های خاک در گروه بندی های بدست آمده از این مدل در شکل ۵ الف ارائه شده است. همان طور که دیده می شود، گروه های ۳، ۶ و ۷ این مدل از نظر تاکسونومیک خالص می باشند. این نتایج بیانگر دقت خوب و بالای مدل خوشه بندی سلسله مراتبی در شناسایی و تفکیک خاک های منطقه در گروه های مختلف می باشد بطوریکه این خاکها با توجه به ویژگی های مختلف خاکی و محیطی آنها شباهت بیشتری را نشان می دهند. با این حال، گروه های ۱، ۲ و ۱۰ این مدل (شکل ۵ الف) بیشترین میزان ناخالصی تاکسونومیک را نشان می دهند. این امر ممکن است به دلیل تنوع بالای خاک های منطقه در قسمت های مرکزی و متمایل به سمت جنوب منطقه باشد که در فاصله کم خاکها تنوع بالاتری را نسبت به سایر قسمت های منطقه نشان می دهند (شکل ۲). بنابراین، فاصله اقلیدسی به دلیل تغییرات زیاد بین خاکها در فاصله کم به خوبی نتوانسته است تفاوت بین آنها را نمایان سازد.

با این حال، بر اساس شاخص ارزیابی بیرونی، شاخص رند تعدیل شده (ARI) این مدل (۰/۴۹) دقت متوسط را نشان می دهد که به تبع آن حاکی از مشابهت و هم خوانی متوسط نتایج این مدل با نقشه واقعیت زمینی زیر گروه ها (شکل ۲) می باشد (جدول ۴). در مطالعه ای ادکویی و همکاران (۲۰۱۷) از الگوریتم خوشه بندی سلسله مراتبی برای گروه بندی خاکها بر اساس مواد مادری و برخی ویژگی های شیمیایی استفاده کردند. مقایسه گروه بندی های کمی بدست آمده با نقشه موروثی خاک نشان دهنده توافق کم بین این دو نقشه بود.

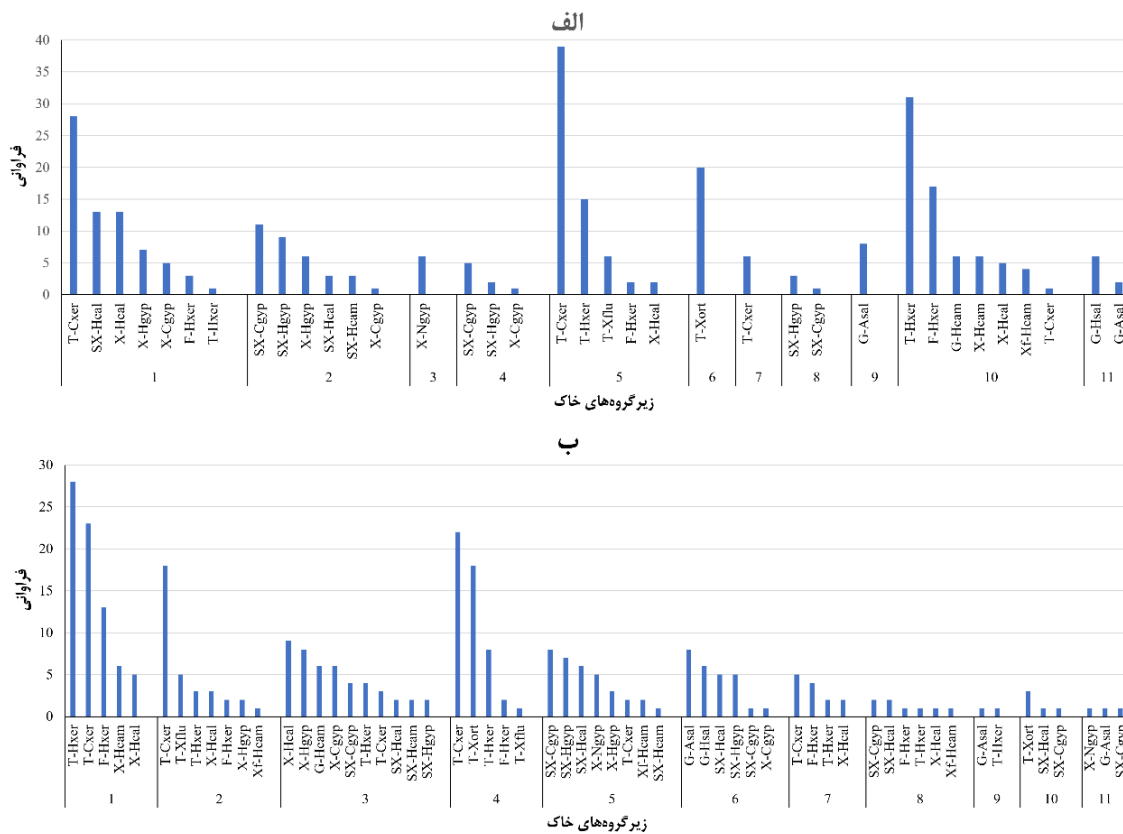
#### خوشه بندی دوطرفه

مشابه با مدل خوشه بندی سلسله مراتبی، خاکها در این مدل نیز در یازده گروه تفکیک شدند؛ که شکل ۴ ب توزیع مکانی آنها در منطقه مورد مطالعه و جدول ۴ نیز نتایج ارزیابی های کمی این مدل را نشان می دهند. همان طور که ملاحظه می شود، ارزیابی داخلی گروه بندی های بدست آمده از این مدل به DB نسبتاً بالا منجر گردید که بیانگر درجه تفکیک پایین گروه های خاک با واریانس درون گروهی بالا می باشد. بنابراین با در نظر گرفتن این شاخص می توان بیان کرد که این مدل به خوبی نتوانسته است خاکها با دقت قابل قبولی گروه بندی کند (جدول ۴).

با این حال، با در نظر گرفتن تغییرات ارتفاع در شکل ۱، این مدل مشابه با مدل خوشه بندی سلسله مراتبی خاک های مورد مطالعه را با توجه به تغییرات پستی و بلندی که مهم ترین عامل مؤثر در تشکیل خاک های منطقه می باشد تفکیک کرده است. به عبارت دیگر، این مدل نیز به این موضوع اهمیت ویژه ای داده و خاک های قسمت های شمالی و مرکزی منطقه را به خوبی با توجه به تغییرات ارتفاعی، پستی و بلندی و اقلیم تفکیک کرده است. اما در نیمه جنوبی منطقه و اراضی پست این موضوع دیده نمی شود. این امر ممکن است نشان دهد که این مدل بر خلاف مدل خوشه بندی سلسله مراتبی در قسمت هایی که تغییرات بیشتری (از نظر متغیرهای خاکی و محیطی) وجود دارد بهتر نتوانسته است عمل کند. که این را می توان به ماهیت ذاتی مدل نسبت داد که در کار با داده های پیچیده تر بهتر عمل می کند (Noronha

(et al., 2022; Chakraborty and Vardeman, 2022).

از طرف دیگر مقایسه نتایج گروه‌بندی‌های این مدل با توزیع زیرگروه‌های خاک در منطقه بیان‌گر توافق کم بین این دو مدل می‌باشد. به عنوان مثال، گروه‌های ۳، ۵ و ۶ ناخالصی‌های تاکسونومیکی زیادی دارند که تأکیدی بر عدم کارایی خوب مدل در شناسایی و تفکیک این خاک‌ها می‌باشد (شکل‌های ۲ و ۵ ب). این هم‌خوانی کم بین این دو مدل با مقادیر کم‌تر شاخص ARI نیز تأیید شده است (جدول ۴). بنابراین می‌توان بیان کرد که استفاده از روش خوشه‌بندی در دو طرف (ویژگی‌ها و خاکرخ‌ها) نیز نتوانسته است به خوبی تفاوت بین خاک‌ها در قسمت‌های پایین دست منطقه جایی که خاک‌های با تنوع بیشتر تاکسونومیکی در فاصله کم رخ داده است را بیان کند. این امر را می‌توان به تعداد کم و پراکنش محدود این خاک‌ها در قسمت‌های پایین دست دانست. با این حال، با در نظر گرفتن ویژگی‌های خاکی و محیطی استفاده شده در تفکیک این خاک‌ها می‌توان عنوان کرد که این مدل دقت خوبی در تفکیک خاک‌ها اعمال کرده است. به عنوان مثال، اگر گروه‌های ۳، ۵ و ۶ با بیشترین درصد ناخالصی تاکسونومیکی را در نظر بگیریم، می‌توان دریافت که همه این خاک‌ها ویژگی‌های خاکی انتخاب شده (گچ، هدایت الکتریکی، کربنات کلسیم، نسبت جذب سدیم) را دارا می‌باشند.



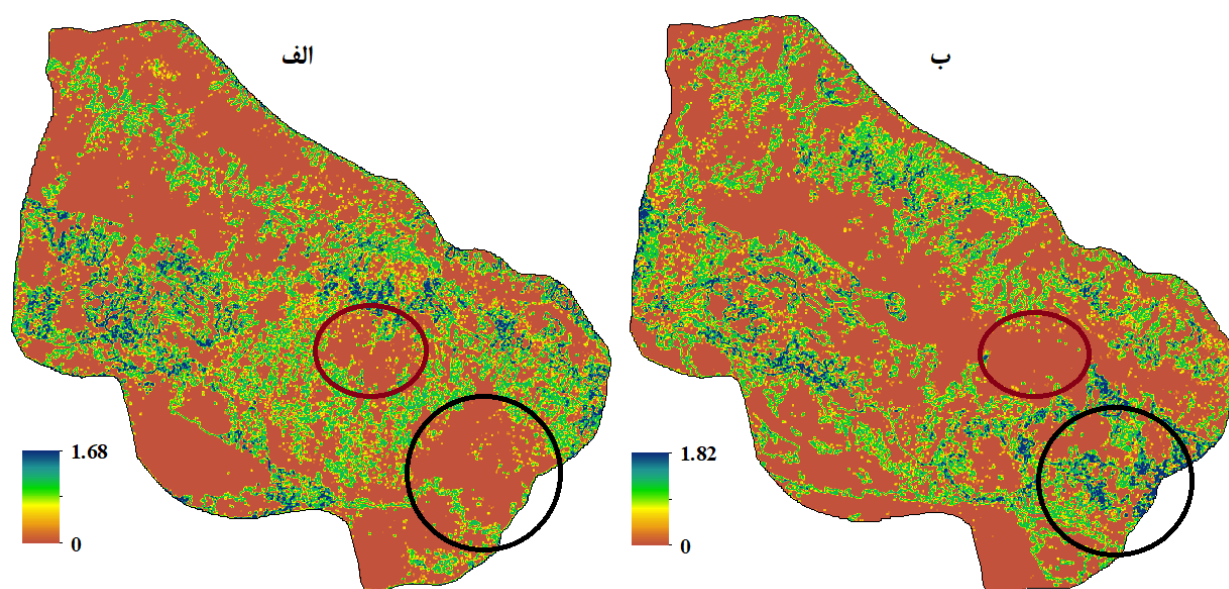
شکل ۵. توزیع زیرگروه‌های خاک در گروه‌های خاک بدست آمده از دو مدل خوشه‌بندی سلسله‌مراتبی (الف) و دوطرفه (ب). برای اطلاع از علامت‌های اختصاری مربوط به زیرگروه‌های خاک به جدول ۱ مراجعه کنید.

### مقایسه مدل‌های خوشه‌بندی

گروه‌بندی‌های به دست آمده از مدل خوشه‌بندی سلسله‌مراتبی نسبت به مدل خوشه‌بندی دوطرفه با درجه بالاتر و با واریانس کم‌تر از یکدیگر تفکیک و متمایز شده‌اند (شاخص DB کم‌تر، جدول ۴). با در نظر گرفتن نقشه توزیع زیرگروه‌های خاک (شکل ۲) به عنوان مبنای ارزیابی گروه‌بندی‌های بدست آمده از این دو مدل خوشه‌بندی مرسوم و مدرن، در قالب شاخص کمی ARI (جدول ۴)، مدل خوشه‌بندی سلسله‌مراتبی به مراتب بهتر از مدل خوشه‌بندی دوطرفه عمل کرده است. این امر را می‌توان به مشابهت ماهیت سلسله‌مراتبی مدل خوشه‌بندی مرسوم با سامانه رده‌بندی تاکسونومیکی منطبق با فرآیندهای تشکیل خاک نسبت داد (Bekele et al., 2024).

مطابق با جدول ۴، در هر دو مدل با رفتن به سطوح پایین‌تر تاکسونومی مقدار شاخص شانس افزایش پیدا می‌کند. این امر اگرچه نشان دهنده افزایش میزان تغییرات خاک‌ها در سطوح پایین‌تر و کارایی مدل‌ها در بیان این تغییرات می‌باشند (Jafari et al., 2013)، اسفندیارپور

و باقری، ۱۳۹۴)، در اینجا این موضوع بیانگر کاهش کارایی مدل‌های خوشه‌بندی مورد استفاده در تفکیک خاک‌ها و تخصیص آن‌ها به گروه‌های همگن‌تر با افزایش میزان جزئیات و تغییرات خاک‌ها در سطوح پایین‌تر تاکسونومی می‌باشد. با این حال، مقایسه نتایج دو مدل خوشه‌بندی مورد بررسی بیانگر کارایی بهتر مدل خوشه‌بندی مرسوم نسبت به مدل مدرن خوشه‌بندی دوطرفه می‌باشد (میزان شاخص شانن کم‌تر در سطوح مختلف تاکسونومیک بجز رده) که در شکل ۵ (سطح زیرگروه) نیز به وضوح قابل مشاهده می‌باشد. بنابراین با توجه به افزایش تعداد زیرگروه‌های خاک در کلاس‌های خاک به دست آمده از مدل خوشه‌بندی دوطرفه نسبت به مدل خوشه‌بندی سلسله مراتبی (شکل ۵) افزایش میزان شاخص تفرق شانن در این مدل منطقی می‌باشد (Minasny et al., 2010). که این با تغییر در میزان مقادیر عددی این شاخص در سطوح مختلف نیز قابل استنباط است، بطوریکه میزان شاخص در سطح رده در هر دو مدل بیانگر میزان تغییرات متوسط می‌باشد (۲/۹۹-۲/۵۰) اما در سطوح پایین‌تر مدل خوشه‌بندی سلسله مراتبی تغییرات زیاد (۳/۴۹-۳/۰۰) و مدل خوشه‌بندی دوطرفه تغییرات خیلی زیاد (بیشتر از ۳/۵۰) را نشان می‌دهند (Fernando et al., 1998). از طرف دیگر، می‌توان چنین استنباط کرد که مدل خوشه‌بندی مدرن دوطرفه به خوبی توانسته است تغییرات خاک‌های منطقه مورد مطالعه را در سطوح مختلف تاکسونومیک نشان دهد. شکل ۶ تغییرات میزان این شاخص در منطقه مورد مطالعه برای گروه‌بندی‌های بدست آمده از دو مدل خوشه‌بندی را نشان می‌دهد.



شکل ۶. نقشه توزیع شاخص شانن گروه‌های خاک بدست آمده از دو مدل خوشه‌بندی سلسله‌مراتبی (الف) و دوطرفه (ب).

مطابق با شکل ۶، هر دو مدل به خوبی تغییرات خاک‌ها در منطقه را نشان می‌دهند (مقایسه با نقشه زیرگروه‌های خاک، شکل ۲). میزان شاخص شانن گروه‌های بدست آمده از مدل خوشه‌بندی دوطرفه بیشتر از مدل خوشه‌بندی سلسله مراتبی می‌باشد. مقایسه دو نقشه بیانگر کارایی بهتر مدل خوشه‌بندی دوطرفه در به تصویر کشیدن تغییرات خاک‌ها در قسمت‌هایی از منطقه می‌باشد. به عنوان مثال، در قسمت جنوب غربی منطقه (دایره سیاه رنگ) مدل خوشه‌بندی دوطرفه بهتر توانسته است تغییرات خاک‌ها را نشان دهد، و در قسمت نشان داده شده با دایره قهوه‌ای روی شکل، مدل خوشه‌بندی دوطرفه یکنواختی بیشتری (کاهش در میزان شاخص شانن) در خاک‌های این قسمت را نشان می‌دهد که منطقی‌تر از نتایج بدست آمده از مدل خوشه‌بندی سلسله مراتبی می‌باشد (مقایسه با نقشه زیرگروه‌های خاک، شکل ۲).

بنابراین با توجه به نتایج به دست آمده می‌توان بیان کرد که مدل خوشه‌بندی دوطرفه نیز مانند سایر مدل‌های عددی گروه‌بندی خاک‌ها در مدل‌سازی خاک مورد استفاده قرار بگیرد. اگرچه این مطالعه نمونه‌ای کوچک از کارایی این مدل در مطالعات خاک می‌باشد، انتظار می‌رود که با افزایش تعداد مشاهدات خاک‌ها کارایی مدل در تفکیک و تمایز خاک‌ها افزایش یابد (Orzechowski et al., 2019; Noronha et al., 2022; Chakraborty and Vardeman, 2022). به گفته (Adams et al., 1992) در مطالعات خاک، تجزیه خوشه ای برای سازمان‌دهی درجه تشابه بسیار مناسب است، بنابراین برای دستیابی به اهداف طبقه‌بندی استفاده می‌شود.

## نتیجه‌گیری

در این مطالعه کارایی مدل‌های خوشه‌بندی مرسوم (خوشه‌بندی سلسله مراتبی) و مدرن (خوشه‌بندی دوطرفه) در شناسایی، تفکیک و گروه‌بندی خاک‌ها بر اساس ویژگی‌های مختلف آن‌ها (خاکی و محیطی) مورد مطالعه قرار گرفت. بررسی‌ها نشان داد که از میان متغیرهای فیزیکی و شیمیایی خاکی، درصد گچ، درصد کربنات کلسیم، هدایت الکتریکی و نسبت جذب سدیم در عمق‌های زیرین خاک‌ها در تمایز آن‌ها مؤثر بودند که به خوبی با تغییرات خاک‌های مهم منطقه از جمله خاک‌های شور، سدیمی و گچی مطابقت دارند. همچنین، از میان متغیرهای محیطی پستی و بلندی، اقلیم و مواد مادری مهم‌ترین عامل‌های مؤثر در تفکیک و گروه‌بندی خاک‌ها بودند. این متغیرها به خوبی با عامل‌های خاک‌سازی مهم در منطقه هم‌خوانی دارند.

اگرچه با توجه به ارزیابی‌های کیفی و کمی انجام شده برای مقایسه گروه‌بندی‌های به دست آمده از دو مدل خوشه‌بندی به کار رفته بیان‌گر کارایی بالاتر مدل خوشه‌بندی سلسله مراتبی در ارائه گروه‌های خاک با میزان واریانس نسبی کمتر (شاخص DB کم‌تر) و همبستگی و توافق بیشتر با نقشه زیرگروه‌های خاک (به عنوان نقشه مرجع واقعیت زمینی) بود، مدل خوشه‌بندی دوطرفه نیز به خوبی توانسته است خاک‌های منطقه را با توجه به تغییرات آن‌ها هم‌راستا با تغییرات فاکتورهای محیطی گروه‌بندی کند. بنابراین با توجه به یافته‌های این پژوهش، استفاده از این مدل به عنوان یک مدل داده‌کاوی مدرن در یافتن الگوی تشابه بین خاک‌ها و در نتیجه گروه‌بندی و مدل‌سازی آن‌ها در مطالعات نقشه‌برداری رقومی خاک در قسمت‌های مختلف کشور پیشنهاد می‌شود.

## قدردانی

این اثر تحت حمایت مادی صندوق حمایت از پژوهشگران و فناوران کشور (INSF) برگرفته شده از طرح شماره ۴۰۲۴۳۴۳، انجام شده است. بدین وسیله نویسنده اول مقاله مراتب قدردانی خود را از صندوق INSF بابت حمایت‌های مالی طرح و همچنین از دانشگاه تهران بابت فراهم آوردن امکانات اجرای آن اعلام می‌دارد.

## "هیچ‌گونه تعارض منافع بین نویسندگان وجود ندارد"

## منابع

- اسفندیارپور بروجنی، عیسی و باقری، محسن. (۱۳۹۴). تجزیه و تحلیل عدم قطعیت یک نقشه خاک با استفاده از شاخص‌های مجاورت تاکسونومیکی و تفرق خاک. نشریه مدیریت خاک و تولید پایدار، ۵ (۴): ۱۴۷-۱۶۰.
- خاموشی، سیدعرفان؛ سرمیدیان، فریدون و کشاورزی، علی. (۱۳۹۷). تهیه نقشه رقومی خاک با مدل جنگل‌های تصادفی و ارزیابی تناسب اراضی برای بخشی از منطقه آبیک، استان قزوین. مرتع و آبخیزداری، مجله منابع طبیعی ایران، ۷۱ (۴): ۸۸۵-۸۹۹.
- خاموشی، سیدعرفان؛ سرمیدیان، فریدون و امید، محمود. (۱۴۰۱). تهیه نقشه رقومی کربن آلی ذخیره شده در خاک با استفاده از روش‌های یادگیری ماشینی. مجله تحقیقات آب و خاک ایران، ۵۳ (۱۱): ۲۶۷۱-۲۶۸۱.
- رحمانی، اصغر؛ سرمیدیان، فریدون و عارفی، حسین، (۱۴۰۱). مدل‌سازی و پیش‌بینی رقومی کلاس‌های زیرگروه خاک با استفاده از رویکرد یادگیری عمیق در بخشی از اراضی خشک و نیمه‌خشک دشت قزوین، مجله تحقیقات آب و خاک ایران، ۵۳ (۱۰): ۲۴۷۷-۲۴۹۹.
- رسائی، زهرا؛ محمدی، جهانگرد و جعفری، اعظم. (۱۳۹۹). مقایسه گروه‌بندی و کیفیت مرزبندی‌های موروثی خاک با مدل‌های داده‌کاوی کمی: مطالعه موردی بخش‌هایی از استان چهارمحال و بختیاری. تحقیقات کاربردی خاک، ۸ (۴): ۲۸-۴۳.
- رضایی، گردآفرین؛ سرمیدیان، فریدون؛ محمدی ترکاشوند، علی؛ سید محمدی، جواد و مرعشی‌علی‌آبادی، مریم. (۱۴۰۲). نقشه‌برداری رقومی تغییرات سطحی و زیرسطحی کربن آلی و شوری خاک در بخشی از اراضی دشت قزوین (مطالعه موردی: مناطق آبیک و نظر آباد). نشریه آب و خاک، ۳۳۱-۳۱۵، (۲): ۳۳۱-۳۱۵.
- سیستانی، ندا؛ معین‌الدینی، مظاهر؛ علی‌طالشی، محمدصالح؛ خراسانی، نعمت‌اله؛ حمیدیان، امیرحسین و عظیمی یانچشمه، رخساره. (۱۳۹۶). منشأیابی آلودگی فلزات سنگین خاک‌های مجاور صنایع فولاد کرمان. محیط‌زیست طبیعی، مجله منابع طبیعی ایران، ۷۰ (۳): ۶۲۷-۶۴۱.
- ممتازی بروجنی، میلاد و سرمیدیان، فریدون. (۲۰۱۴). پیش‌بینی مکانی کلاس خاک به کمک مدل درخت تصمیم‌گیری C5.0 تقویت‌شده با بوستینگ در بخشی از اراضی شهرستان آبیک. مرتع و آبخیزداری، مجله منابع طبیعی ایران، ۷۵ (۴): ۵۵۳-۵۷۲.
- موسوی، سیدروح‌اله؛ سرمیدیان، فریدون و رحمانی، اصغر. (۱۳۹۸). مدل‌سازی و پیش‌بینی مکانی کلاس خاک با استفاده از الگوریتم یادگیری رگرسیون

درختی توسعه یافته و جنگل‌های تصادفی در بخشی از اراضی دشت قزوین. تحقیقات آب و خاک ایران، ۵۰(۱۰)، ۲۵۲۵-۲۵۳۸.

موسوی، سیدروح اله؛ سرمیدیان، فریدون؛ امید، محمود و بوگارت، پاتریک. (۱۴۰۰-۱). کاربرد مدل‌های یادگیری ماشین در برآورد مکانی فسفر و پتاسیم خاک در بخشی از اراضی دشت آبیک. نشریه علمی پژوهش‌های خاک، ۳۵(۴)، ۳۹۷-۴۱۲.

موسوی، سیدروح اله؛ سرمیدیان، فریدون؛ امید، محمود و بوگارت، پاتریک. (۱۴۰۰-۲). مدل‌سازی تغییرات عمقی کربنات کلسیم معادل خاک با استفاده از الگوریتم‌های یادگیری ماشین در دشت قزوین. نشریه آب و خاک (علوم و صنایع کشاورزی)، ۲۵(۵)، ۷۱۹-۷۳۴.

موسوی، سیدروح اله؛ سرمیدیان، فریدون؛ امید، محمود و بوگارت، پاتریک. (۱۴۰۰-۳). مدل‌سازی رقومی تغییرات سه بعدی شوری خاک با استفاده از الگوریتم‌های یادگیری ماشین در اراضی خشک و نیمه‌خشک دشت قزوین. تحقیقات آب و خاک ایران، ۵۲(۷)، ۱۹۱۵-۱۹۲۹.

## REFERENCES

- Adams, M. B., Turner, R. S., & Schmoyer, D. D. (1992). Evaluation of direct delayed response project soil sampling classes: Northeastern United States. *Soil Science Society of America Journal*, 56, 177-187.
- Adhikari, K., Minasny, B., Greve, M. B., & Greve, M. H. (2014). Constructing a soil class map of Denmark based on the FAO legend using digital techniques. *Geoderma*, 214-215, 101-113.
- Aouabed, H., Elloumi, M., & Santamaría, R. (2021). An evaluation study of biclusters visualization techniques of gene expression data. *Journal of Integrative Bioinformatics*, 18(4), 20210019.
- Arnold, R. (2006). Soil Survey and Soil Classification. In S. Grunwald (Ed.), *Environmental Soil-Landscape Modeling: Geographic Information Technologies and Pedometrics* (pp. 37-60). CRC Press.
- Aubert, J., Schbath, S., & Robin, S. (2021). Model-based biclustering for overdispersed count data with application in microbial ecology. *Methods in Ecology and Evolution*, 12(6), 1050-1061.
- Beaudette, D. E., Roudier, P., & O'Geen, A. T. (2023). Algorithms for quantitative pedology: A toolkit for soil scientists. *Computers and Geosciences*, 52, 258-268.
- Bekele, A., Beyene, S., Yimer, F., & Kiflu, A. (2024). Numerical classification of termite-mediated soils along toposequences and rangeland use influenced soil properties in southeast Ethiopia. *Heliyon*, 10(1), e23726.
- Bouma, J., Montanarella, L., & Evanylo, G. (2019). The challenge for the soil science community to contribute to the implementation of the UN Sustainable Development Goals. *Soil Use and Management*, 35(4), 538-546.
- Carré, F., & Jacobson, M. (2009). Numerical classification of soil profile data using distance metrics. *Geoderma*, 148(3), 336-345.
- Cebeci, Z., & Yildiz, F. (2015). Comparison of K-means and Fuzzy C-means algorithms on different cluster structures. *Journal of Agricultural Informatics*, 6(3), 13-23.
- Chakraborty, A., & Vardeman, S. B. (2022). Some Bayesian biclustering methods: Modeling and inference. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 15(4), 409-538.
- Cheng, Y., & Church, G.M. (2000). Biclustering of Expression Data. Proceedings of the 8th International Conference on Intelligent Systems for Molecular Biology, 93-103.
- Chiquet, J., Rigail, G., Sundqvist, M., Dervieux, V., & Bersani, F. (2023). Efficient Computations of Standard Clustering Comparison Measures, R Package 'aricode'. CRAN.
- Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., & Böhrer, J. (2015). System for automated geoscientific analyses (SAGA) v. 2.1.4. *Geoscientific Model Development*, 8(7), 1991-2007.
- Dane, J. H., & Topp, C. G. (Eds.). (2020). Methods of soil analysis, Part 4: Physical methods (Vol. 20). John Wiley & Sons.
- Davies, D. L., & Bouldin, D. W. (1979). A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2), 224-227.
- de França, F. O., & Coelho, A. L. V. (2015). A biclustering approach for classification with mislabeled data. *Expert Systems with Applications*, 42(12), 5065-5075.
- Deressa, A., Yli-halla, M., Mohamed, M., & Wogi, L. (2018). Soil classification of humid Western Ethiopia: a transect study along a toposequence in Didessa watershed. *Catena*, 163, 184-195.
- Dunkl, I., & Ließ, M. (2022). On the benefits of clustering approaches in digital soil mapping: an application example concerning soil texture regionalization. *SOIL*, 8, 541-558.
- Edokpayi, A. A., Agho, C. A., Adeg, S. A., & Okpamen, S. U. (2017). Comparison of the Different Hierarchical Clustering Techniques for the Classification of Soils under Oil Palm in Nigeria. *International Journal of Basic Science and Technology*, 3(1), 37-46.
- Egbueri, J. C. (2023). Soil erosion and landslide susceptibility insights based on hierarchical clustering and multilayer perceptron networks: a Nigerian case study. *Volume 20*, pages 10763-10786.





- Esfandiarpour Borujeni, I., & Bagheri Bodaghabadi, M. (2016). Analysis of uncertainty of a soil map using taxonomic adjacency and pedodiversity indices. *Journal of Soil Management and Sustainable Production*, 5(4), 147-160. (In Persian with English abstract).
- Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). Cluster analysis (5th ed.). Wiley.
- Fernando, E. S., Balatibat, J. B., Peras, J. R., & Jumawid, R. J. J. (1998). Resource inventory and assessment of biodiversity in Subic Bay Metropolitan Authority. Terminal Report. Research Project funded by DOST-PCARRD, SBMA and UPLB.
- Gad, M.M.S., Mohamed M.H.A., Mohamed, M.R. (2021). Soil salinity mapping using remote sensing and GIS. *Geomatica*, 75(4), 295-309.
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote sensing of Environment*, 202, 18-27.
- Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2022). A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM) (3rd ed.). Thousand Oaks, CA: Sage.
- Handl, J., Knowles, J., & Kell, D. B. (2005). Computational cluster validation in post-genomic data analysis. *Bioinformatics*, 21(15), 3201-3212.
- Hartigan, J. A. (1972). Direct Clustering of a Data Matrix. *Journal of the American Statistical Association*, 6(337), 123-129.
- Heil, J., Häring, V., Marschner, B., & Stumpe, B. (2019). Advantages of fuzzy k-means over k-means clustering in the classification of diffuse reflectance soil spectra: A case study with West African soils. *Geoderma*, 337, 11-21.
- Hu, C., Wright, A. L., & Lian, G. (2019). Estimating the Spatial Distribution of Soil Properties Using Environmental Variables at a Catchment Scale in the Loess Hilly Area, China. *International Journal of Environmental Research and Public Health*, 16(3), 491.
- Hughes, P. A., McBratney, A. B., Minasny, B., & Campbell, S. (2014). End members, end points and extragrades in numerical soil classification. *Geoderma*, 226-227, 365-375.
- Ibanez, J. J., De Alba, S., Bermudes, F. F., & Garcia-Alvarez, A. (1995). Pedodiversity: concepts and measurements. *Catena*, 24, 215-232.
- Index DataBase. (2021). A Database for Remote Sensing Indices. Retrieved from <https://www.indexdatabase.de/>.
- Jafari, A., Ayoubi, S., Khademi, H., Finke, P. A., & Toomanian, N. (2013). Selection of a taxonomic level for soil mapping using diversity and map purity indices: a case study from an Iranian arid region. *Geomorphology*, 201, 86-97.
- Jafari, A., Finke, P. A., Wauw, J. V., Ayoubi, S., & Khademi, H. (2012). Spatial prediction of USDA great soil groups in the arid Zarand region, Iran: comparing logistic regression approaches to predict diagnostic horizons and soil types. *European Journal of Soil Science*, 63, 284-298.
- Jhariya, M. K., & Singh, L. (2021). Herbaceous diversity and biomass under different fire regimes in a seasonally dry forest ecosystem. *Environment, Development and Sustainability*, 23(5), 6800-6818.
- Kaiser, S. (2011). Biclustering: Methods, Software and Application.
- Kaiser, S., Santamaria, R., Khamiakova, T., Sill, M., Theron, R., Quintales, L., Leisch, F., De Troyer, E., & Leon, S. (2023). biclust: BiCluster Algorithms.
- Kassambara, A. (2017). Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning. STHDA.
- Ketchen Jr., D. J., & Shook, C. L. (1996). The application of cluster analysis in strategic management research: an analysis and critique. *Strategic Management Journal*, 17(6), 441-458.
- Khamoshi, S. E., Sarmadian, F., & Keshavarzi, A. (2019). Digital soil mapping Using Random Forests and Land Suitability Evaluation for Abyek Region, Qazvin Province. *Journal of Range and Watershed Management*, 71(4), 885-899. (In Persian with English abstract).
- Khamoshi, S. E., Sarmadian, F., & Omid, M. (2023). Predicting and Mapping of Soil Organic Carbon Stock Using Machine Learning Algorithm. *Iranian Journal of Soil and Water Research*, 53(11), 2671-2681. (In Persian with English abstract).
- Láng, V., Fuchs, M., Waltner, I., & Michéli, E. (2013). Soil taxonomic distance, a tool for correlation: As exemplified by the Hungarian Brown Forest Soils and related WRB Reference Soil Groups. *Geoderma*, 192, 269-276.
- Lee, D. B., Kim, Y. N., Sonn, Y. K., & Kim, K. H. (2023). Comparison of Soil Taxonomy (2022) and WRB (2022) Systems for classifying Paddy Soils with different drainage grades in South Korea. *Land*, 12(6), 1204.
- Li et al. (2020). Biclustering with missing data. *Information Sciences*, 510, 304-316.



- Liberti, L., Lavor, C., Maculan, N., & Mucherino, A. (2014). Euclidean distance geometry and applications. *SIAM review*, 56(1), 3-69.
- Liu, H., Yang, J., Ye, M., Tang, Z., Dong, J., & Xing, T. (2021). Using one-way clustering and co-clustering methods to reveal spatio-temporal patterns and controlling factors of groundwater geochemistry. *Journal of Hydrology*, 603, 127085.
- Malone, B. P., McBratney, A. B., Minasny, B., & Laslett, G. M. (2009). Mapping continuous depth functions of soil carbon storage and available water capacity. *Geoderma*, 154(1), 138-152.
- McBratney, A. B., Santos, M. M., & Minasny, B. (2003). On digital soil mapping. *Geoderma*, 117(1-2), 3-52.
- Minasny, B., McBratney, A. B., & Hartemink, A. E. (2010). Global pedodiversity, taxonomic distance, and the World Reference Base. *Geoderma*, 155, 132-139.
- Momtazi Burojeni, M., & Sarmadian, F. (2023). Spatial prediction of soil classes using C5.0 boosted decision tree model in Abyek Area. *Iranian Journal of Soil and Water Research*, 75(4), 553-572. (In Persian with English abstract).
- Mousavi, S. R. A., Sarmadian, F., & Rahmani, A. (In Persian with English abstract). Modelling and Prediction of Soil Classes Using Boosting Regression Tree and Random Forests Machine Learning Algorithms in Some Part of Qazvin Plain. *Journal of Water and Soil*, 50(10), 2525-2538. (In Persian with English abstract).
- Mousavi, S. R. A., Sarmadian, F., Omid, M., & Bogaert, P. (2021). Modeling the Vertical Soil Calcium Carbonate Equivalent Variation by Machine Learning Algorithms in Qazvin Plain. *Journal of Water and Soil*, 35(5), 719-734. (In Persian with English abstract).
- Mousavi, S. R. A., Sarmadian, F., Omid, M., & Bogaert, P. (2021). Digital Modeling of Three-Dimensional Soil Salinity Variation Using Machine Learning Algorithms in Arid and Semi-Arid lands of Qazvin Plain. *Iranian Journal of Soil and Water Research*, 52(7), 1915-1929. (In Persian with English abstract).
- Mousavi, S. R. A., Sarmadian, F., Omid, M., & Bogaert, P. (2022). Application of Machine Learning Models in Spatial Estimation of Soil Phosphorus and Potassium in Some Parts of Abyek Plain. *Iranian Journal of Soil Science*, 35(4), 397-411. (In Persian with English abstract).
- Mousavi, S. R., Sarmadian, F., Angelini, M. E., Bogaert, P., & Omid, M. (2023). Cause-effect relationships using structural equation modeling for soil properties in arid and semi-arid regions. *Catena*, 232, 107392.
- Mousavi, S. R., Sarmadian, F., Omid, M., & Bogaert, P. (2022). Three-dimensional mapping of soil organic carbon using soil and environmental covariates in an arid and semi-arid region of Iran. *Measurement*, 201, 111706.
- Nachtergaele, F., van Velthuisen, H., Verelst, L., Wiberg, D., Henry, M., Chiozza, F., ... & Tramberend, S. (2023). Harmonized world soil database version 2.0. Food and Agriculture Organization of the United Nations.
- Naimi, S., Ayoubi, S., Zeraatpisheh, M., & Dematte, J. A. M. (2021). Ground observations and environmental covariates integration for mapping of soil salinity: a machine learning-based approach. *Remote Sensing*, 13(23), 4825.
- Neyestani, M., Sarmadian, F., Jafari, A., Keshavarzi, A., & Sharififar, A. (2021). Digital mapping of soil classes using spatial extrapolation with imbalanced data. *Geoderma Regional*, 26, e00422.
- Noronha, M. D., Henriques, R., Madeira, S. C., & Zárata, L. E. (2022). Impact of metrics on biclustering solution and quality: A review. *Pattern Recognition*, 127, 108612.
- Orzechowski, P., Boryczko, K., & Moore, J. H. (2019). Scalable biclustering—the future of big data exploration? *GigaScience*, 8(7), giz078.
- Pham, H., Reisner, J., Swift, A., Olafsson, S., & Vardeman, S. (2022). Crop phenotype prediction using biclustering to explain genotype-by-environment interactions. *Frontiers in Plant Science*, 13, 975976.
- Pontes, B., Giráldez, R., & Aguilar-Ruiz, J. S. (2015). Biclustering on expression data: A review. *Journal of Biomedical Informatics*, 57, 163-180.
- R Development Core Team. (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna Austria.
- Rahmani, A., Sarmadian, F., & Arefi, H. (2023). Digital modeling and prediction of soil subgroup classes using deep learning approach in a part of arid and semi-arid lands of Qazvin Plain. *Iranian Journal of Soil and Water Research*, 53(11), 2477-2499. (In Persian with English abstract).
- Rasaei, Z., Mohammadi, J., & Jafari, A. (2021). Comparison of grouping and the quality of legacy soil map boundaries with numerical data mining models: a case study of some regions of Chaharmahal-va-Bakhtiari province. *Applied Soil Research*, 8(4), 28-43. (In Persian with English abstract).
- Rasaei, Z., Rossiter, D.G., & Farshad, A. (2020). Rescue and renewal of legacy soil resource inventories in



- Iran as an input to digital soil mapping. *Geoderma regional*, 21, e00262.
- Rezaie, G., Sarmadian, F., Mohammadi Torkashvand, A., Seyedmohammadi, J., & Marashi Aliabadi, M. (2023). Digital mapping of surface and subsurface soil organic carbon and soil salinity variation in a part of Qazvin plain (Case study: Abyek and Nazarabad regions). *Journal of Water and Soil*, 37(2), 315-331. (In Persian with English abstract).
- Romanski, P., Kotthoff, L., & Schratz, P. (2023). FSelector R Package. CRAN.
- Saldana, A., & Ibanez, J.J. (2004). Pedodiversity analysis at large scales: an example of three fluvial terrains of the Henares River (central Spain). *Geoderma*, 62, 123-1384.
- Schoeneberger, P.J., Wysocki, D.A., Benham, E.C., & Soil Survey Staff. (2021). *Field book for describing and sampling soils, Version 3.0*. Natural Resources Conservation Service, National Soil Survey Center, Lincoln, NE.
- Seaton, F. M., Barrett, G., Burden, A., Creer, S., Fitos, E., Garbutt, A., Griffiths, R. I., Henrys, P., Jones, D. L., Keenan, P., Keith, A., Lebron, I., Maskell, L., Pereira, M. G., Reinsch, S., Smart, S. M., Williams, B., Emmett, B. A., & Robinson, D. A. (2021). Soil health cluster analysis based on national monitoring of soil indicators. *European Journal of Soil Science*, 72(6), 2414-2429.
- Selmy, S., Abd El-Aziz, S., El-Desoky, A., & El-Sayed, M. (2022). Characterizing, predicting, and mapping of soil spatial variability in Gharb El-Mawhoub area of Dakhla Oasis using geostatistics and GIS approaches. *Journal of the Saudi Society of Agricultural Sciences*, 21(6), 383-396.
- Shelia, V., & Hoogenboom, G. (2020). A new approach to clustering soil profile data using the modified distance matrix.
- Sistani, N., Moein-Aldini, M., Ali Taleshi, M. S., Khorasani, N. A., Hamidian, A. H., & Azimi Yancheshmeh, R. (2018). Source identification of heavy metal pollution nearby Kerman steel industries. *Journal of Natural Environment*, 70(3), 627-641. (In Persian with English abstract).
- Soil Survey Staff. (2022). *Keys to Soil Taxonomy*, 13th ed. USDA-Natural Resources Conservation Service.
- Sonn, Y., Seo, B., Go, W., Jeon, S., Hyun, B., & Yun, S. (2019). Consideration of suffix symbol on soil taxonomy and world reference base for soil resources classification. *Korean Journal of Soil Science and Fertilizer*, 52(4), 345-351.
- Soropa, G. A., Mbisva, O. M., Nyamangara, J., Nyakatawa, E. Z., Nyapwere, N., & Lark, R. M. (2021). Spatial variability and mapping of soil fertility status in a high-potential smallholder farming area under sub-humid conditions in Zimbabwe. *SN Applied Sciences*, 3(396).
- Sparks, D. L., Page, A. L., Helmke, P. A., & Loeppert, R. H. (Eds.) (2020). *Methods of soil analysis, part 3: Chemical methods* (Vol. 14). John Wiley & Sons.
- Sunori, S. K., Pant, J., Pant, H., Manu, M., & Juneja, P. (2022, October). Design of K-Means Clustering & SVM based Models for Soil Fertility Classification. In *2022 3rd International Conference on Smart Electronics and Communication (ICOSEC)* (pp. 862-866). IEEE.
- Taghizadeh-Mehrjardi, R., Minasny, B., Sarmadian, F., & Malone, B. P. (2014). Digital mapping of soil salinity in Ardakan region, central Iran. *Geoderma*, 213, 15-28.
- Taghizadeh-Mehrjardi, R., Nabiollahi, K., Minasny, B., & Triantafyllidis, J. (2015). Comparing data mining classifiers to predict spatial distribution of USDA-family soil groups in Baneh region, Iran. *Geoderma*, 253-254, 67-77.
- Tchorbadjieff, A., Kostev, T., Stoyanova, V., & Tcherkezova, E. (2019). K-means clustering of a soil sampling scheme with data on the morphography of the Ogosta Valley Northwestern Bulgaria. *European Journal of Geography*, 10(2).
- Van Wambeke, A. R. (2000). The Newhall Simulation Model for estimating soil moisture & temperature regimes. Conservation Service: Department of Crop and Soil Sciences Cornell University, Ithaca, NY USA.
- Walesiak, M. (2023). *clusterSim: Searching for Optimal Clustering Procedure for a Data Set*. R package version 0.50-23.
- Wang, N., Chen, S., Huang, J., Frappart, F., Taghizadeh, R., Zhang, X., ... & Shi, Z. (2024). Global Soil Salinity Estimation at 10 m Using Multi-Source Remote Sensing. *Journal of Remote Sensing*, 4, 0130.
- Weatherill, G., & Burton, P. W. (2009). Delineation of shallow seismic source zones using K-means cluster analysis, with application to the Aegean region. *Geophysical Journal International*, 176(2), 565-588.
- Wilding, L. P. (1985). Spatial variability: its documentation, accommodation and implication to soil survey. In D. R. Nielsen & J. Bouma (Eds.), *Soil Spatial Variability* (pp. 166-189). Pudoc.
- Xie, J., Guo, A.-Y., Fennell, A., Ma, Q., & Zhao, J. (2019). It is time to apply biclustering: A comprehensive review of biclustering applications in biological and biomedical data. *Briefings in Bioinformatics*, 20(4),

1450–1465.

- Xu, H., Croot, P., & Zhang, C. (2021). Discovering hidden spatial patterns and their associations with controlling factors for potentially toxic elements in topsoil using hot spot analysis and K-means clustering analysis. *Environment International*, 151, 106456.
- Zhang, X., & Huang, B. (2019). Prediction of soil salinity with soil-reflected spectra: a comparison of two regression methods. *Scientific Reports*, 9, 5067.
- Zhao, W., Ma, J., Liu, Q., Song, J., Tysklind, M., Liu, C., Wang, D., Qu, Y., Wu, Y., & Wu, F. (2023). Comparison and application of SOFM, fuzzy c-means and k-means clustering algorithms for natural soil environment regionalization in China. *Environmental Research*, 216(Pt 2), 114519.
- Zirnea, S., Lazar, I., Foudjo, B.U.S., Vasilache, T., & Lazar, G. (2013). Cluster Analysis Based of Geochemical Properties of Phosphogypsum Dump Located Near Bacau City in Romania. *APCBEE Procedia*, 5, 317–322.
- Zolfaghari, F., Khosravi, H., Shahriyari, A., Jabbari, M., & Abolhasani, A. (2019). Hierarchical cluster analysis to identify the homogeneous desertification management units. *PLoS One*, 14(12), e0226355.