

Spatiotemporal Mapping of Soil Organic Carbon Using an Inverse Modeling Approach: A Case Study of the Zayandeh Rud Watershed in Isfahan Province

S. Masoumi¹, H.R. Matinfar^{1*}, S.R. Mousavi²

1- Department of Soil Science, Faculty of Agriculture, Lorestan University, Iran

(* - Corresponding Author Email: matinfar.h@lu.ac.ir)

2- Department of Soil Science, Faculty of Agriculture, University of Tehran, Karaj, Iran

Received: 21 June 2025

Revised: 21 July 2025

Accepted: 07 September 2025

Available Online: 07 September 2025

How to cite this article:

Masoumi, S., Matinfar, H.R., & Mousavi, S.R. (2025). Spatiotemporal mapping of soil organic carbon using an inverse modeling approach: A case study of the Zayandeh Rud watershed in Isfahan province. *Journal of Water and Soil*, 39(3), 291-312. (In Persian with English abstract).

<https://doi.org/10.22067/jsw.2025.94129.1488>

Introduction

Soil organic carbon (SOC), as one of the most important components of the global carbon cycle, plays a vital role in maintaining soil quality, enhancing fertility, and moderating climate change. The spatio-temporal variations of SOC are influenced by various factors, including land use, climatic conditions, topography, and human activities. Additionally, SOC contributes to diverse functions in natural and agricultural ecosystems, such as increasing soil fertility, controlling erosion, enhancing water permeability in soil, and reducing the effects of greenhouse gases.

Materials and Methods

Given the critical role of SOC in enhancing soil quality, this study aims to investigate the spatio-temporal variability of SOC using a reverse modeling approach based on a spatial model developed in 2024. The model is extended to analyze data from the years prior to 2015, 2010, and 2000, incorporating environmental variables derived from remote sensing (RS), topographic attributes, climatic data, land use, and geological information within the Zayandeh Rud watershed. For the environmental covariates, RS data, land use, and climatic information were obtained from Google Earth Engine's open-source spatial database for the relevant years from 2000 to 2024. In total, 76 auxiliary variables were prepared, including vegetation indices derived from band ratios of RS data, as well as the digital elevation model (DEM), geology, land use, and climatic factors, which were used as representatives of soil-forming factors. A relative importance feature selection method was employed to finalize the dataset of environmental covariates. Furthermore, three machine learning models, Random Forest (RF), Support Vector Regression (SVR), and Extreme Gradient Boosting (XGBoost) tree, were utilized to explore the relationships between environmental factors and SOC. Four common statistical indices, including the coefficient of determination (R^2), concordance correlation coefficient (CCC), root mean square error (RMSE), and percentage of normalized root mean square error (nRMSE), were used to evaluate the performance of the machine learning models. Two uncertainty quantification approaches for prediction, namely bootstrap and k-fold cross-validation, were also applied.



©2025 The author(s). This is an open access article distributed under [Creative Commons Attribution 4.0 International License \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/).

<https://doi.org/10.22067/jsw.2025.94129.1488>

Results and Discussion

The evaluation of machine learning models for predicting SOC revealed a notable decline in the performance of all models from the year 2024 back to 2000, as indicated by the R^2 statistic. Among the models assessed, the RF model exhibited superior performance, achieving the highest R^2 values for the years 2024 and 2015, thereby indicating its effectiveness in capturing the complexities of SOC dynamics. The SVR model demonstrated intermediate performance, while the XGBoost model showed relatively weaker results compared to the other two models. Despite these variances in performance, all three machine learning models effectively established a robust connection between SOC and the predictive environmental variables, affirming their suitability for this analysis. Furthermore, the uncertainty quantification of SOC predictions highlighted that the bootstrap method outperformed the k-fold method, yielding lower values for both standard deviation (SD) and mean uncertainty, which suggests that the bootstrap approach provides a more reliable prediction of SOC variability. In terms of the relative importance of environmental variables in predicting SOC, the analysis across all time periods indicated that climatic factors played the most significant role, closely followed by topographical attributes. Other environmental variables, including land use, geology, and RS data, had a lesser impact on explaining spatial variations in SOC. The spatial analysis indicated alarming increases in areas with very low SOC content, suggesting soil degradation risks. Furthermore, higher rainfall and lower temperatures were associated with the highest SOC levels, emphasizing the need for effective soil management strategies.

Conclusion

This study emphasizes the necessity of ongoing monitoring and management of soil organic carbon (SOC) to combat soil degradation and promote sustainable agriculture. The findings also provide a framework for creating soil property maps in data-scarce regions, enhancing decision-making for effective soil management strategies. Overall, the findings from this study highlight the need for continuous monitoring and management of SOC to mitigate risks related to soil degradation and to promote sustainable agricultural practices.

Keywords: Digital mapping, Environmental variables, Machine learning algorithms, Temporal and spatial variability

مقاله پژوهشی

جلد ۳۹، شماره ۳، مرداد-شهریور ۱۴۰۴، ص. ۲۹۱-۳۱۲

نقشه‌برداری زمانی مکانی کربن آلی خاک با استفاده از رویکرد مدلسازی معکوس: مطالعه موردی حوضه آبخیز زاینده‌رود استان اصفهان

شیما معصومی^۱ - حمیدرضا متین فر^{۱*} - سید روح اله موسوی^۲

تاریخ دریافت: ۱۴۰۴/۰۳/۳۱

تاریخ پذیرش: ۱۴۰۴/۰۶/۱۶

چکیده

تغییرات زمانی و مکانی کربن آلی خاک تحت تأثیر عوامل متعددی از جمله کاربری اراضی، شرایط اقلیمی، پستی و بلندی و فعالیت‌های انسانی قرار دارد. با توجه به جایگاه کربن آلی در بهبود کیفیت خاک، پژوهش حاضر با هدف بررسی تغییرات زمانی و مکانی کربن آلی خاک (SOC) با استفاده از رویکرد مدلسازی معکوس بر مبنای مدل مکانی تهیه شده در سال ۲۰۲۴ و تعمیم آن به سال‌های ماقبل ۲۰۱۵، ۲۰۱۰ و ۲۰۰۰ با استفاده از متغیرهای محیطی در حوضه آبخیز زاینده‌رود انجام گردید. همچنین سه مدل یادگیری ماشین جنگل تصادفی (RF)، رگرسیون بردار پشتیبان (SVR) و درخت تقویت شده با گرادیان افراطی (XGBoost) برای بررسی ارتباط بین متغیرهای محیطی و SOC به همراه دو رویکرد کمی‌سازی عدم قطعیت پیش‌بینی شامل بوتسراپ و k-فولد استفاده گردید. نتایج اعتبارسنجی کارایی مدل‌های یادگیری ماشین حاکی از دقت بالاتر مدل RF نسبت به دو مدل دیگر بود. اگرچه به‌طور مشابه در هر سه مدل روند دقت پیش‌بینی SOC از سال ۲۰۲۴ به ۲۰۰۰ کاهش یافت. همچنین رویکرد عدم قطعیت بوتسراپ بر اساس آماره‌های انحراف معیار و میانگین عدم قطعیت از اطمینان بالاتری در پیش‌بینی SOC برخوردار بود. اهمیت نسبی متغیرهای محیطی نیز نشان‌دهنده این بود که عوامل اقلیمی و پستی و بلندی از درجه اهمیت بالاتری در پیش‌بینی SOC بودند. همچنین، تغییرات مکانی SOC با روند الگوی پراکنش متغیرهای اقلیمی و پستی و بلندی همخوانی داشت، به‌طوری‌که مناطق با بارش و ارتفاع بیشتر به‌همراه دمای پایین‌تر، محتوای SOC بیشتری را نشان دادند. نقشه‌های پیش‌بینی مکانی و زمانی نیز بیانگر افزایش سطح اراضی با محتوای SOC بسیار پایین (کمتر از ۰/۵٪) در طی دوره زمانی بلند مورد بررسی از ۲۰۰۰ تا ۲۰۲۴ بود. به‌طور کلی رویکرد مورد استفاده در این پژوهش می‌تواند به‌عنوان یک چارچوب مناسب برای دست‌یابی به روند تغییرات ویژگی‌های خاک در مناطقی باشد، که به‌طور عمده فاقد پایگاه داده‌های اطلاعات مکانی خاک در گذشته می‌باشند.

واژه‌های کلیدی: الگوریتم‌های یادگیری ماشین، تغییرپذیری زمانی و مکانی، متغیرهای محیطی، نقشه‌برداری رقومی

مقدمه

اراضی، شرایط اقلیمی، پستی و بلندی و فعالیت‌های انسانی قرار دارد (Parras-Alcántara, Díaz-Jaimes, & Lozano-García, 2015). همچنین SOC در زیست‌بوم‌های طبیعی و زراعی از کارکردهای متنوعی همچون افزایش حاصلخیزی خاک، کنترل فرسایش، افزایش نفوذپذیری آب در خاک و کاهش اثرات گازهای گلخانه‌ای برخوردار

کربن آلی خاک (SOC) به‌عنوان یکی از مهمترین اجزای چرخه جهانی کربن، نقش حیاتی در حفظ کیفیت خاک، افزایش حاصلخیزی و تعدیل تغییرات اقلیمی ایفا می‌کند (Yang et al., 2020). تغییرات زمانی و مکانی SOC تحت تأثیر عوامل متعددی از جمله کاربری

۱- گروه علوم و مهندسی خاک، دانشکده کشاورزی، دانشگاه لرستان، ایران

(*)- نویسنده مسئول: (Email: matinfar.h@lu.ac.ir)

۲- گروه علوم و مهندسی خاک، دانشکده کشاورزی، دانشگاه تهران، کرج، ایران

است (Laban, Metternicht, & Davies, 2018). از این رو SOC یک ویژگی کلیدی در تعیین سلامت و کیفیت خاک محسوب می‌شود که تمامی ویژگی‌های فیزیکی، شیمیایی و زیستی خاک را تحت تأثیر خود قرار می‌دهد (Mousavi, Sarmadian, Angelini, Bogaert, & Omid, 2023; Matinfar, Maghsodi, Mousavi, & Rahmani, 2021).

نقشه‌های خاک یکی از ابزارهای بسیار مؤثر در مباحث تصمیم‌گیری‌های توسعه کشاورزی (تناسب اراضی، الگوی کشت)، حفاظت خاک و بهبود بخشیدن به روند تصمیم‌گیری در مورد برنامه‌ریزی‌های مدیریت سرزمین می‌باشد (Adedeji, 2019)، اما باید توجه داشت که به‌روز رسانی نقشه‌ها و اطلاعات خاک با روش‌های معمول^۱ برای کشور وسیعی مانند ایران فرآیندی زمانبر و هزینه‌بردار است. از این رو تهیه نقشه با روش‌های جدید و با قابلیت تعمیم در مقیاس‌های وسیع منطقه-ای و ملی، از اهمیت ویژه‌ای برخوردار است (Zeraatpisheh et al., 2020). امروزه استفاده از روش‌های پدومتری جهت برآورد توزیع مکانی داده‌های اندازه‌گیری شده‌ی خاک، به دلیل در نظر گرفتن پیوستگی مکانی داده‌ها، نسبت به روش‌های معمول و زمین‌آماري^۲ روبه افزایش است و نقشه‌برداری رقومی خاک^۳ (DSM) به منزله‌ی زیر مجموعه‌ای از علم پدومتری، به دلیل افزایش منابع داده‌های کمکی یا محیطی با دسترسی آسان، به‌طور کاملاً عملی توانسته است به انتقادات وارده بر رویکرد خاک‌شناسی مرسوم فایق آید (Vasiliniuc, Patriche, & Mousavi et al., 2023; Pîrnău, & Roșca, 2013). این درحالی است که در رویکرد مرسوم بدلیل وابستگی به دانش کارشناس، عدم ارائه نتایج کمی از نتایج نقشه‌برداری، ناتوانی در ارائه تغییرات پیوسته پراکنش مکانی و زمانی ویژگی‌های خاک عملاً یک روش پرهزینه و وقت‌گیر بوده و در نتیجه برای اهداف کاربردی دارای محدودیت است (Kulikova, Chechenkov, Osipova, Shopina, & Semenov, 2023; Taghizadeh-Mehrjardi, Neupane, Sood, & Kumar, 2017).

در مورد پراکنش مکانی SOC مطالعات متعددی توسط محققین پیشین در مناطق خشک و نیمه‌خشک ایران و جهان از قبیل مطالعه‌ای که در مناطق مرتعی استان ایلام با هدف بررسی میزان ذخیره کربن آلی خاک^۴ سطحی با استفاده از دو مدل یادگیری ماشین جنگل تصادفی (RF) و کوئیست (CB) انجام گردید و محققین گزارش نمودند که ترکیبی از داده‌های سنجش از دور و پستی و بلندی کارایی خوبی در فرآیند پیش‌بینی داشتند (Rostaminia, Rahmani, Mousavi, & Taghizadeh-Mehrjardi, 2021). در مطالعه‌ی دیگری در اراضی کشاورزی ایتالیا (Adeniyi, Brenning, &

Maerker, 2024) محققان گزارش کردند که مدل RF نسبت به دو مدل شبکه عصبی مصنوعی (ANN) و ماشین یادگیری افراطی^۵ (ELM) کارایی بالاتری داشت و متغیرهای پستی و بلندی مهمترین اهمیت را در پیش‌بینی SOC در اراضی کشاورزی مورد مطالعه داشتند، در همین راستا مطالعات مشابهی توسط (Zeratpisheh et al., 2023) در ایالت ورموت آمریکا با در نظر گرفتن تغییرات SOC در زمین‌نماها و اراضی با کاربری‌های متنوع، (Matinfar et al., 2021) در اراضی زراعی واقع در مجاورت شهر خرم آباد، (Su et al., 2024) در مناطق جنگلی با اقلیم نیمه‌خشک در ایران و (Garosi, Ayoubi, & Nussbaum, & Sheklabadi, 2022) به بررسی تغییرات SOC زیست‌بوم‌های زراعی و جنگلی در یک منطقه نیمه‌خشک در ایران پرداختند. اما موردی که علی‌رغم موفقیت تحقیقات قبلی در نقشه برداری ویژگی‌های خاک بویژه SOC همچنان در کشورهای در حال توسعه از قبیل ایران بسیار جدی و تأثیرگذار می‌باشد بررسی روند تغییرات زمانی و مکانی این ویژگی‌ها است، که عمدتاً به دلیل نبود مشاهدات کافی در دوره‌های زمانی گذشته (بویژه سال‌های دور) امکان دسترسی به این داده‌های اطلاعاتی برای تصمیم‌گیری‌های بلند مدت و برنامه‌ریزی‌های آتی را تحت تأثیر قرار داده است (Afshar, Ayoubi, & Jafari, 2018). یکی از راهکارهای غلبه بر این محدودیت در نقشه برداری SOC استفاده از رویکرد مدلسازی معکوس^۶ می‌باشد (Fathizad et al., 2020). در این فرآیند از داده‌های مشاهداتی خاک در زمان حاضر و متغیرهای محیطی برای پیش‌بینی متغیر هدف استفاده می‌گردد و در ادامه مدل مکانی ساخته شده با استفاده از متغیرهای محیطی مانند داده‌های سنجش از دور (RS)، کاربری اراضی و پارامترهای اقلیم که در طی زمان متغیر هستند برای دست یافتن به روند تغییرات گذشته متغیر هدف (به‌عنوان نمونه: SOC) مورد استفاده قرار می‌گیرند. در همین راستا در معدود مطالعاتی که محققین به این موضوع پرداخته‌اند، می‌توان به پژوهش‌های مربوط به مدلسازی زمانی مکانی SOC (در سال‌های ۱۹۸۶، ۱۹۹۹، ۲۰۱۰ و ۲۰۱۶) و شاخص کیفیت خاک (از سال‌های ۱۹۸۶ تا ۲۰۳۰) با استفاده از مدل RF در دشت اردکان یزد اشاره نمود (Fathizad et al., 2020; 2022). نتایج ایشان بیانگر این بود که تعمیم مدل واسنجی‌شده در سال ۲۰۱۶ به دوره‌های زمانی گذشته ۱۹۸۶، ۱۹۹۹ و ۲۰۱۰ حاکی از کاهش قابل توجه در محتوای SOC به‌ویژه در خاک‌هایی که محتوای SOC قابل توجهی در گذشته داشته و در شرایط فعلی از مساحت آنها کاهش یافته است. این کاهش‌ها را عمدتاً به تغییرات کاربری اراضی و توسعه‌ی فعالیت‌های کشاورزی نسبت داده شد. بنابراین، بر مدیریت بهینه منابع خاک به‌منظور کاهش این اثرات تأکید شده است (Fathizad et al.,

4- Soil organic carbon stocks
5- Extreme Learning Model
6- Inverse Modeling Approach

1- Conventional methods
2- Geostatistics
3- Digital Soil Mapping

(2020).

آنها مزیک^۴ و ترمیک^۵ از غرب به شرق منطقه می‌باشد. مواد مادری تشکیل دهنده‌ی خاک‌های منطقه بر اساس نقشه‌ی زمین‌شناسی با مقیاس (۱:۱۰۰۰۰۰) نیز از نوع رسوبات آبرفتی، کفه‌های رسی، شیل خاکستری و ماسه سنگ است. خاک‌های منطقه در سطح تاکسونومیک رده بر اساس سیستم طبقه‌بندی آمریکایی شامل اینسپتی‌سولز، انتی سولز و اردی‌سولز (Soil Survey Staff, 2022) به‌همراه برخی مناطق با بیرون‌زدگی سنگی می‌باشند.

نمونه‌برداری و اندازه‌گیری آزمایشگاهی

تعداد ۳۰۰ نمونه خاک سطحی (۳۰-۰ سانتی‌متری) در حدواسط شهریور تا آبان ماه سال ۱۴۰۳ بر اساس روش نمونه‌برداری ابرمکعب لاتین مشروط^۶ (clhs) در حین مطالعه میدانی برداشت گردید. در روش نمونه‌برداری ابرمکعب لاتین که در نقشه‌برداری رقومی خاک به‌کار می‌رود، هدف این است که توزیع متغیرهای ورودی به‌طور مؤثری نمونه‌برداری شود. در این پژوهش متغیرهای محیطی مورد استفاده شامل مشتقات پستی و بلندی از قبیل (عمق دره، موقعیت نسبی شیب و ...)، سنجش از دور (مانند شاخص‌های سبزی‌نگی، روشنایی و ...) به‌همراه فاکتورهای اقلیمی بارش و دما بودند. این روش شامل تقسیم‌بندی هر یک از متغیرها به π بخش مساوی است و سپس به‌طور تصادفی از هر بخش یک نمونه انتخاب می‌شود. با این کار، تمامی ناحیه‌های ممکن از فضای متغیرها پوشش داده می‌شود و از بروز نمونه‌های تکراری جلوگیری می‌شود. این روش به‌ویژه در مدل‌سازی‌های پیچیده و در شرایطی که داده‌های کافی در دسترس نیست، به کار می‌رود و به محققان کمک می‌کند تا پیش‌بینی‌های دقیق‌تری از ویژگی‌های خاک و تغییرات آن در مناطق مختلف داشته باشند. نمونه‌ها پس از انتقال به آزمایشگاه، هوا خشک و به نرمی کوبیده شده و با عبور از الک ۲ میلی‌متری برای تجزیه‌های آزمایشگاهی آماده شدند. برای اندازه‌گیری SOC، در گام بعد نمونه‌های خاک به‌طور کامل از الک با اندازه ۰/۵ میلی‌متر عبور داده شدند. درنهایت بر اساس روش ارائه شده توسط والکلی-بلاک (Walkley & Black, 1934) مقدار SOC اندازه‌گیری شد.

حوضه آبخیز زاینده رود، به‌عنوان یکی از مهمترین حوضه‌های آبریز در مرکز ایران، از تنوع بالای کاربری اراضی و شرایط محیطی برخوردار است. این حوضه به‌دلیل تأمین آب آشامیدنی، کشاورزی و صنعتی برای میلیون‌ها نفر، از اهمیت ویژه‌ای برخوردار است (Madani, 2007). با این حال، تغییرات اقلیمی، خشکسالی‌های متوالی و فعالیت‌های انسانی در این منطقه، باعث تغییرات قابل توجهی در کیفیت خاک و ذخایر کربن آلی شده است. بنابراین، بررسی تغییرات زمانی و مکانی SOC در این حوضه از اهمیت بالایی برخوردار است. در این مطالعه، از داده‌های RS و سایر نمایندگان عوامل خاکسازي شامل داده‌های اقلیمی، متغیرهای پستی و بلندی، کاربری اراضی و زمین‌شناسی برای بررسی تغییرات زمانی و مکانی SOC استفاده شده است. بنابراین پژوهش حاضر با اهداف، (۱) بررسی روند تغییرات زمانی و مکانی SOC در لایه سطحی خاک، (۲) شناسایی مهمترین عوامل محیطی پیش‌ران و کلیدی بر پیش‌بینی SOC، (۳) ارزیابی کارایی سه مدل یادگیری ماشین RF، رگرسیون بردار پشتیبان (SVR) و درخت تقویت شده با گرادیان افراطی (XGBoost tree) در پیش‌بینی SOC در دوره‌های زمانی مدنظر شامل سال‌های ۲۰۰۰، ۲۰۱۰، ۲۰۱۵ و ۲۰۲۴ و (۴) مقایسه کارایی دو روش عدم قطعیت بوتسراپ و k-فولد برای کمی‌سازی دقت پیش‌بینی مناسبترین مدل یادگیری ماشین انجام گردید.

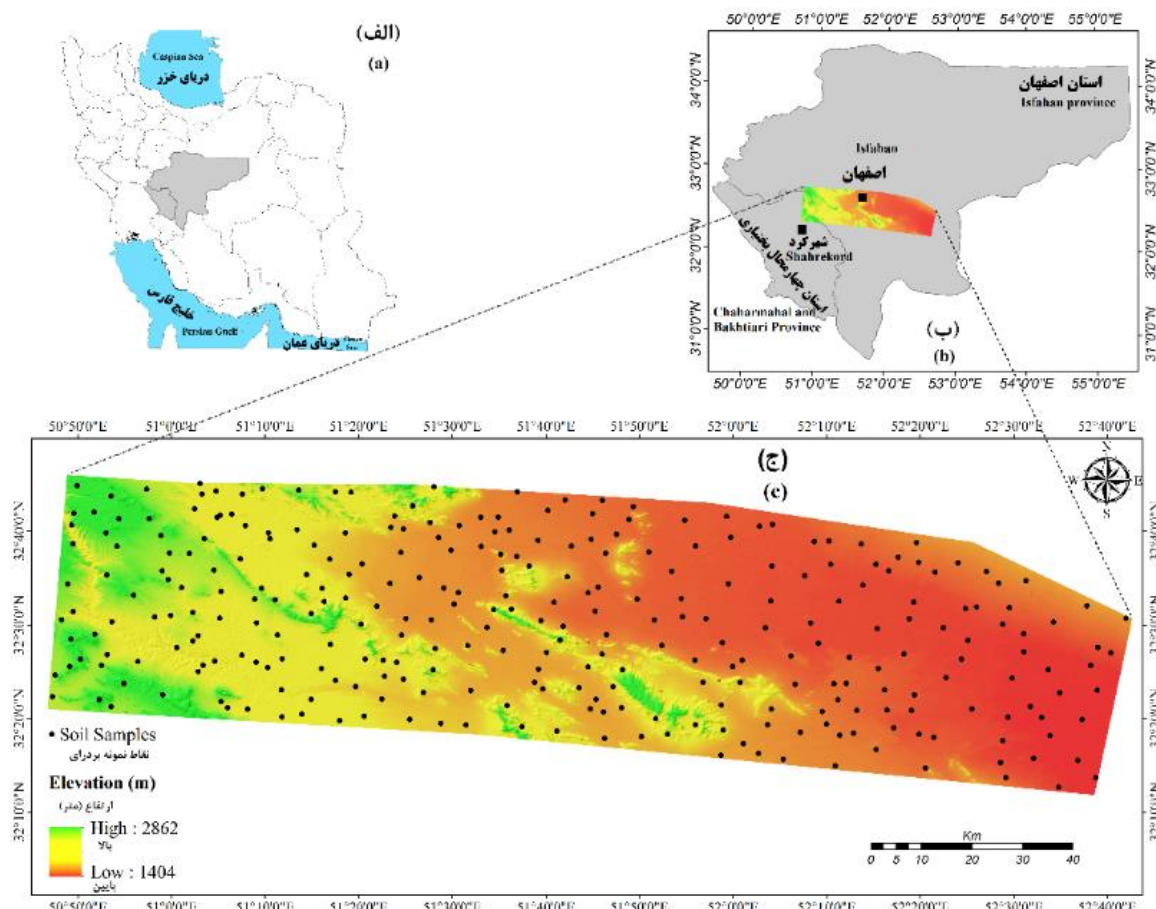
مواد و روش‌ها

منطقه مورد مطالعه

در این پژوهش منطقه مورد مطالعه بخشی از اراضی حوضه آبخیز زاینده رود است، که بین عرض‌های جغرافیایی $35^{\circ} 21' 50''$ تا $35^{\circ} 53' 0''$ طول غربی و $36^{\circ} 24' 31''$ تا $36^{\circ} 41' 33''$ عرض شمالی قرار گرفته است (شکل ۱). حداقل و حداکثر ارتفاع منطقه از شرق به غرب آن به‌ترتیب ۱۴۰۴ تا ۲۸۶۲ متر متغیر می‌باشد. واحدهای فیزیوگرافی غالب منطقه نیز شامل تپ‌های کوهستان، تپه، واریزه‌های آبرفتی، دشت‌های دامنه‌های و اراضی پست بوده و روند تغییرات شیب منطقه نیز از شرق به غرب منطقه در دامنه‌ی کمتر از ۱٪ تا بیش از ۷۰٪ متغیر می‌باشد. بر اساس داده‌های اقلیمی بلند مدت (۲۵ ساله) برگرفته از نزدیکترین ایستگاه‌های هواشناسی منطقه نیز حداقل و حداکثر بارش و دما به‌ترتیب ۱۲۵ تا ۳۱۰ میلی‌متر و ۱۱ تا ۱۸ درجه‌سانتی‌گراد است و بر اساس روش ارائه شده توسط (Van Wambeke, 2000) در نرم افزار نیوهال^۱ و بررسی نقشه‌های رژیم‌های رطوبتی و حرارتی خاک‌های ایران رژیم رطوبتی خاک‌ها به‌ترتیب زیریک^۲ و اریدیک^۳ و رژیم حرارتی

4- Mesic
5- Thermic
6- Conditional Latin Hypercube Sampling Method

1- New Hall
2- Xeric
3- Aridic



شکل ۱- (الف) موقعیت منطقه مطالعاتی در کشور، (ب) نسبت به استان‌های اصفهان و چهارمحال بختیاری و (ج) پراکنش نقاط نمونه‌برداری بر روی نقشه تغییرات ارتفاعی (متر)

Figure 1- (a) Location of the study area in the country, (b) relative to Isfahan and Chaharmahal Bakhtiari provinces, and (c) distribution of sampling points on the elevation variation map (m)

آگوست تا نوامبر برای سال‌های ۲۰۲۴، ۲۰۱۵، ۲۰۱۰ و ۲۰۰۰ از محیط سامانه متن باز و بر خط گوگل ارث انجین (GEE) استخراج شدند (Lamichhane, Kumar, & Wilson, 2019; Benslama, Lucas, Jordan Vidal, Almendro-Candel, & Navarro-Pedreño, 2024). از این رو، ۷۰ متغیر کمکی شامل شاخص‌های پوشش گیاهی از نسبت‌های باندی داده‌های RS و همچنین مشتق اولیه و ثانویه DEM در نرم‌افزار GIS SAGA^۲ نسخه ۴/۷ تهیه شد. قبل از اجرای فرآیند مدل‌سازی مکانی، فرآیند انتخاب متغیرهای کمکی محیطی برای به حداقل رساندن تعداد متغیرها و انتخاب مرتبط‌ترین آنها توصیه می‌شود (Ben Brahim & Limam, 2018). برای دستیابی به این هدف، از روش اهمیت متغیرهای محیطی^۳

متغیرهای محیطی و انتخاب آنها

متغیرهای کمکی مورد استفاده شامل مجموعه گسترده‌ای از نمایندگان عوامل خاکسازي بودند، که تا حد امکان از منابع با حداقل هزینه و با دسترسی بالا برای توضیح محتوای SOC جمع‌آوری شدند. آنها شامل ویژگی‌های پستی و بلندی، داده‌های RS، متغیرهای اقلیمی (بارش، حداقل، حداکثر، میانگین دما) و نقشه‌های موضوعی شامل زمین شناسی و کاربری زمین بودند. مدل رقومی ارتفاع^۱ (DEM) با وضوح مکانی ۳۰ متر پایگاه داده از SRTM برای تهیه ویژگی‌های پستی و بلندی دانلود شد. برای شاخص‌های طیفی، میانگین باندهای سنجنده-های Landsat-8 (OLI/TIRS) منطبق با دوره نمونه‌برداری از

- 1- Digital Elevation Model
- 2- System for Automated Geoscientific Analysis
- 3- Relative importance method

رقومی خاک هستند.

ارزیابی عملکرد مدل‌ها

به منظور اعتبارسنجی خارجی مدل‌های ماشین یادگیری، از ۲۰٪ از کل داده‌ها برای اعتبارسنجی پیش‌بینی SOC در هر دوره زمانی به عنوان داده‌های مستقل استفاده گردید. شاخص‌های آماری رایج، شامل ضریب تبیین (R^2)، ضریب همبستگی تطابق (CCC)، ریشه میانگین مربعات خطا (RMSE) و درصد ریشه میانگین مربعات خطای نرمال شده (nRMSE) برای ارزیابی کارایی مدل‌ها استفاده شدند. هر چه ریشه میانگین مربعات خطا مقدار کوچکتری را نشان دهند، مدل دارای دقت و اعتبار بیشتری است و هر چه مقادیر به صفر نزدیکتر باشد دقت بالاتر است (Ma et al., 2017; Mantero, Moser, & Serpico, 2005). ریشه میانگین مربعات خطا نشان‌دهنده این است که مقادیر پیش‌بینی شده چقدر به مقادیر مشاهده شده نزدیک هستند. مقادیر بالای R^2 و CCC بالاتر منعکس کننده این است که مدل عملکرد بهتری را در برآورد مقادیر پیش‌بینی شده در مقابل مقادیر مشاهداتی ارائه کرده است (Voltz et al., 2020). روابط ریاضی شاخص‌های آماری مورد استفاده در روابط ۱ تا ۴ ارائه شده‌اند.

$$R^2 = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \quad (1)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n ((P_i - O_i)^2)} \quad (2)$$

$$CCC = \frac{2r\sigma_{Pi} \sigma_{Oi}}{\sigma_{Pi}^2 + \sigma_{Oi}^2 + (\bar{P}_i - \bar{O}_i)^2} \quad (3)$$

$$nRMSE = \frac{RMSE}{\bar{O}} \times 100 \quad (4)$$

که در معادلات بالا P_i و O_i مقادیر پیش‌بینی شده و اندازه‌گیری شده هستند، $\bar{O}$ میانگین مقادیر مشاهده شده نسبت به اندازه‌گیری‌های n ، ضریب همبستگی بین مقادیر پیش‌بینی شده و اندازه‌گیری شده، و σ_{Pi}^2 ، σ_{Oi}^2 انحراف معیار مقادیر پیش‌بینی شده و مشاهده شده است (Muslim, Albayati, & Al-Nasri, 2022).

تحلیل عدم قطعیت

در این مطالعه برای تعیین عدم قطعیت پیش‌بینی SOC از روش بوت استرپ k و فولد استفاده شد. روش بوت‌سرایت یکی از روش‌های غیرپارامتریک و شناخته شده برای محاسبه عدم قطعیت مدل‌سازی است (Tibshirani & Efron, 1993). کارکرد روش شامل یک نمونه‌برداری تصادفی همراه با قابلیت جایگزینی از داده‌های موجود می‌باشد. بر اساس این روش هر بار انتخاب تصادفی از داده‌های موجود در بخش واسنجی یک مدل برازش می‌گردد که منجر به پیش‌بینی متغیر

(Mallah, Delsouz Khaki, Davatgar, Poppiel, & Dematté, 2022)

برای تعیین مناسب‌ترین نمایندگان عوامل خاکسازی استفاده شد. در این پژوهش به ترتیب تعداد ۴۶ و ۲۴ متغیر مستخرج از داده‌های پستی و بلندی و طیفی برای پیش‌بینی SOC مورد ارزیابی قرار گرفتند، و آنهایی که مقادیر شاخص عمق حداقل^۱ آنها در مقابل تعداد درختان رگزیون برازش داده شده (۰ تا ۱۰۰۰ عدد) کمتر و همچنین و دارای مقادیر $pvalue$ آنها در همبستگی با متغیر هدف معنی‌دار بود (کمتر از ۰/۰۵) بودند انتخاب شدند (شکل ۲ الف و ب). همچنین نقشه کاربری اراضی برای چهار دوره‌ی زمانی ۲۰۱۰، ۲۰۱۵، ۲۰۲۴ و ۲۰۲۴ از محیط سامانه‌ی متن باز گوگل ارث انجین GEE با هدف ارزیابی تأثیر کاربری اراضی بر روی پیش‌بینی SOC تهیه شدند. داده‌های اقلیم و نقشه زمین‌شناسی منطقه نیز به ترتیب از پایگاه داده جهانی Word Clime و رقومی‌سازی نقشه‌های زمین‌شناسی منطقه (۱:۱۰۰۰۰۰) به عنوان نمایندگان فاکتورهای خاکسازی اقلیم و مواد مادری بر اساس نظر کارشناس استفاده شدند. در نهایت ۳۲ متغیر محیطی به عنوان نمایندگان عوامل خاکسازی برای پیش‌بینی SOC انتخاب گردیدند (جدول ۲).

مدل‌سازی مکانی و زمانی کربن آلی خاک

رویکرد مدل‌سازی مورد استفاده در این پژوهش مدل‌سازی معکوس می‌باشد که از آن به عنوان یک ابزار تحلیلی برای برآورد و تجزیه و تحلیل توزیع SOC استفاده شد. در این روش به منظور شبیه‌سازی و برآورد مقادیر SOC در دوره‌های زمانی که داده‌های خاک در دسترس نیست می‌تواند یک رویکرد کارا و مفید باشد (Fathy et al., 2025). برای برازش این رویکرد از دسته‌ای از متغیرهای محیطی به عنوان نمایندگان عوامل خاکسازی بویژه آنهایی که در طی زمان دارای تغییرات مکانی هستند به همراه سه مدل یادگیری ماشین شامل (Behrens, Zhao et al., SVR, Zhu, Schmidt, & Scholten, 2010) RF (2024) و XGBoost tree (Chen & Guestrin, 2016) برای نقشه برداری تغییرات مکانی SOC در دوره‌های زمانی (۲۰۱۰، ۲۰۱۵، ۲۰۲۴) و ۲۰۰۰) مورد استفاده قرار گرفتند. در این رویکرد از مدل مکانی و زمانی ساخته شده در سال مبنا (۲۰۲۴) که داده‌های مشاهداتی و اندازه‌گیری شده آن موجود می‌باشد برای برآورد SOC در سال‌های ماقبل ۲۰۱۵، ۲۰۱۰ و ۲۰۰۰ استفاده گردید. کارایی این رویکرد در مطالعات قبلی توسط سایر محققین در کشور ایران از قبیل (Fathzad et al., 2022) و (Fathy et al., 2025) به ترتیب در نقشه‌برداری SOC و شاخص کیفیت خاک مورد تأیید قرار گرفته است. تمام مدل‌های ماشین یادگیری با استفاده از بسته‌های تخصصی موجود در نرم‌افزار تجزیه و تحلیل آماری R (نسخه ۴،۲،۱) پیاده‌سازی شدند. الگوریتم‌های مورد استفاده از محبوب‌ترین مدل‌های یادگیری ماشین برای نقشه‌برداری

است. بر اساس نتایج میانگین مقادیر SOC برابر ۱/۲۹٪ است. ضریب تغییرات (CV) نقش مهمی در ارزیابی تنوع ویژگی‌های خاک در مجموعه داده ایفا می‌کند. در اینجا، طبقه‌بندی ارائه شده توسط ویلدینگ (۱۹۸۵) برای تعیین میزان تغییرپذیری SOC استفاده شد. با توجه به آماره CV مقادیر بیش از ۳۵٪ آن در کلاس تغییرپذیری بالا، مقادیر حدواسط ۱۵ تا ۳۵ در کلاس متوسط و مقادیر کمتر از ۱۵٪ در کلاس تغییرپذیری کم قرار دادند. بر همین اساس مقدار CV برای SOC بیش‌تر از ۳۵٪ بود که بیانگر تنوع بالای مکانی آن در منطقه مطالعاتی است. تنوع بالا در مقادیر اندازه‌گیری شده SOC به کاربری اراضی، تنوع در پوشش گیاهی و پستی و بلندی، همچنین تنوع در توزیع اندازه ذرات خاک نسبت داده می‌شود (Zaheri Abdehvand, Karimi, Rangzan, & Mousavi, 2024; Song et al., 2017; Rabbi, Roy, Miah, Amin, & Khandakar, 2014). تغییرات ویژگی‌های خاک را در مزارع کشاورزی را به عوامل زمین‌شناسی و پدولوژیکی تشکیل دهنده خاک و عملیات خاک‌ورزی نسبت دادند.

هدف می‌گردد؛ بنابراین با تکرار هر یک از مدل‌ها در هر مرحله در نهایت امکان دستیابی به نقشه احتمال برای هر یک از پیکسل‌ها فراهم می‌گردد. در روش k-فولد نیز کل داده‌های موجود به k زیر بخش با اندازه مساوی تقسیم می‌شوند که به آن fold یا چینه نیز می‌گویند. در هر مرحله یک چینه کنار گذاشته شده و مدل با توجه به سایر چینه‌ها آموزش می‌بیند که در نهایت منجر به ساخت یک مدل یادگیرنده می‌شود. سپس دقت این مدل بر روی هر دسته از داده‌های کنار گذاشته شده بررسی شده و در نهایت میانگین تمام این دقت‌ها به عنوان تخمین مقدار متغیر هدف در هر پیکسل محاسبه می‌شود (Vanwinckelen & Blockeel, 2012).

نتایج و بحث

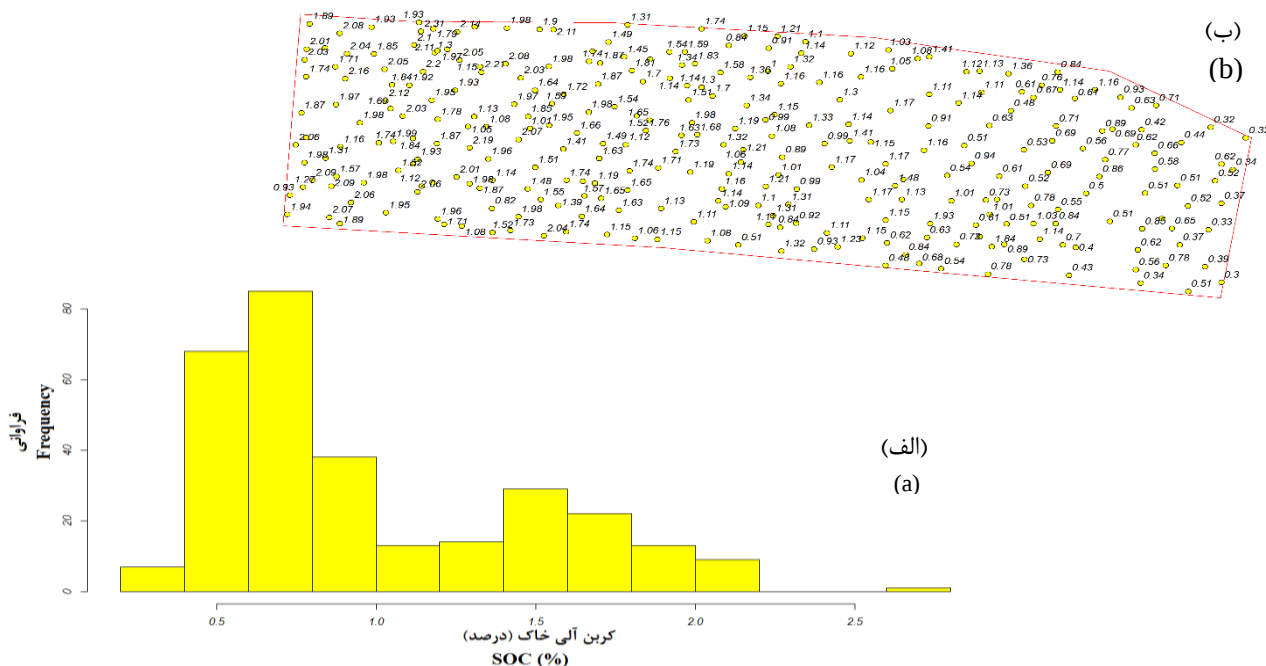
تجزیه و تحلیل آماری وضعیت کربن آلی خاک

خلاصه آماری محتوای SOC در جدول ۱ و نمودار توزیع فراوانی و مقادیر متناظر آن در محل نقاط نمونه‌برداری در شکل ۲ ارائه شده

جدول ۱- آمار توصیفی کربن آلی خاک در خاک‌های مورد مطالعه (n=۳۰۰) در سال ۲۰۲۴

Table 1- Descriptive statistics of soil organic carbon in the studied soils (n=300) in the year 2024

ویژگی	واحد	میانگین	بیشینه	کمینه	میان	انحراف معیار	ضریب تغییرات
Variable	unit	Mean	Maximum	Minimum	Median	Standard deviation	Coefficient of variation
SOC	%	1.29	2.31	0.3	1.17	0.52	40.31



شکل ۲- توزیع فراوانی کربن آلی خاک (الف)، مقادیر متناظر SOC در محل نقاط مشاهده‌ای خاک (ب)

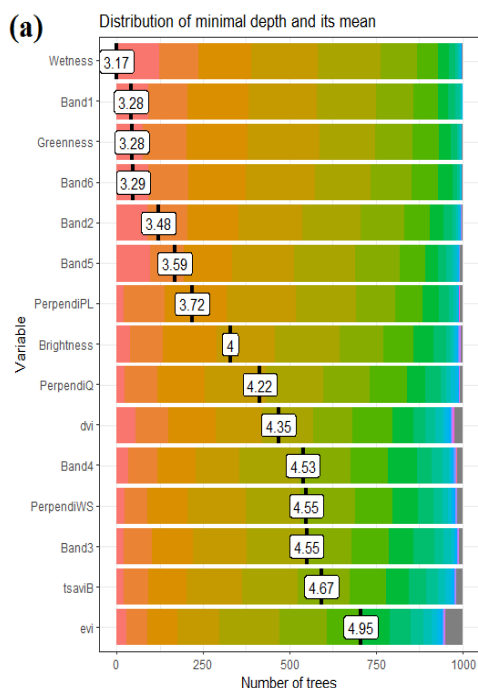
Figure 2- Distribution of SOC frequency (a), and Corresponding SOC values at soil observation points (b)

مشتقات RS و پستی و بلندی بود. در همین راستا (Mallah et al., 2022; Mousavi, Jahandideh Mahjenabadi, Khoshru, & Rezaei, 2024) بر کارایی موفق روش مورد استفاده در این تحقیق برای انتخاب متغیرهای محیطی در مطالعات نقشه‌برداری رقومی خاک تأکید نمودند. همچنین چهار متغیر اقلیمی به همراه دو نقشه زمین شناسی و کاربری اراضی نیز براساس نظر کارشناس انتخاب شدند. در مجموع ۳۱ متغیر محیطی به نمایندگی از عوامل خاکساز پستی و بلندی، اقلیم، RS و مواد مادری برای مدل‌سازی زمانی و مکانی SOC استفاده شدند (جدول ۲). نقشه پراکنش مکانی شش عدد از این متغیرها نیز در شکل ۴ ارائه شده است. به‌طور جزئی‌تر در مورد وضعیت کاربری اراضی بر اساس جدول ۳ تغییراتی در مورد کاهش سهم سطح زیرکشت اراضی کشاورزی و باغی (تحت کشت) در طی زمان اتفاق افتاده به‌نحوی که سهم این اراضی در سال ۲۰۰۰ از ۱۲۴۸۵۵ هکتار به ۹۶۴۵۲ هکتار در سال ۲۰۲۴ معادل ۳/۳۸٪ کاهش یافته است که تغییرات در کاربری اراضی، به‌ویژه تبدیل اراضی زراعی به مناطق شهری، تأثیرات قابل توجهی بر محتوای SOC و خدمات زیست‌بومی آن در منطقه دارد.

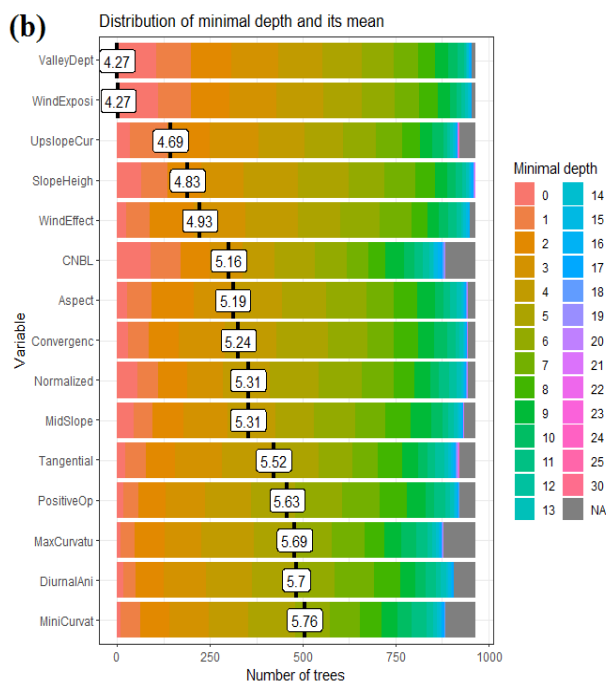
متغیرهای محیطی منتخب

چندین متغیر محیطی بر توزیع ویژگی‌های خاک از جمله کربن آلی تأثیر می‌گذارند، که این متغیرها شامل شاخص‌های RS، پستی و بلندی، آب و هوا و مواد مادری هستند. با این حال، این نوع داده‌ها در بسیاری از مناطق در دسترس نیستند و کاربرد روش‌های DSM را پیچیده می‌کنند (Canero, Rodriguez-Galiano, & Aragones, 2024; Kasraei et al., 2024). بر اساس نتایج انتخاب متغیرهای محیطی در نهایت ۹ و ۱۶ متغیر به‌ترتیب مربوط به داده‌های RS و پستی و بلندی انتخاب شدند (شکل ۳ الف و ب) که همزمان کمترین مقادیر شاخص حداقل عمق و pvalue معنی‌داری با متغیر هدف (SOC) ارائه نمودند انتخاب گردیدند. شایان ذکر است که محور عمودی و افقی در شکل ۳ به‌ترتیب نشان‌دهنده شاخص حداقل عمق و تعداد درختان رگرسیونی است. بر اساس شکل (۲الف) مقادیر کمتر از ۷۵۰ درخت رگرسیونی و در شکل (۲ب) و حداکثر ۵۰۰ درخت رگرسیونی بهینه‌ترین حالت برازش مدل RF در طبق روش انتخاب متغیر اهمیت نسبی در گزینش مناسب

(الف)



(ب)



شکل ۳- فرآیند انتخاب متغیرهای محیطی برای پیش‌بینی SOC. داده‌های RS (الف) و عوامل پستی و بلندی (ب)

Figure 3- Environmental variables selection process for SOC prediction. RS data (a) and, topographic attributes (b)

جدول ۲- متغیرهای محیطی منتخب برای پیش‌بینی کربن آلی خاک

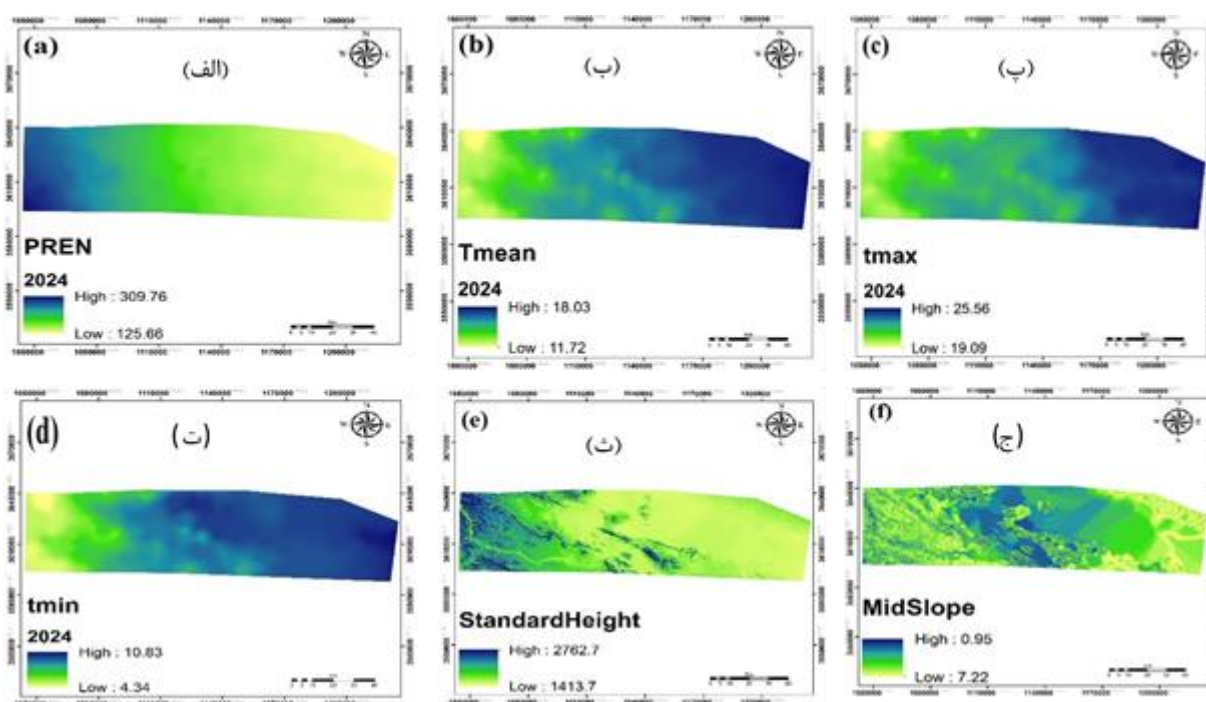
Table 2- Selected environmental variables for predicting soil organic carbon

منبع Source	متغیر کمکی Auxiliary variable	تعریف Definition
پستی و بلندی Topography	Valley depth	عمق دره
	Aspect	جهت شیب
	Convergence index (Convergence)	شاخص همگرایی
	Mid Slope position	موقعیت شیب میانی
	Positive Openness (NegOpp)	باز بودن پستی و بلندی یک مکان منظر را بیان می کند
	Wind Exposition	شاخص بدون بعد مناطق در معرض باد را نشان می دهد
	Plan Curvature (PlanCur)	انحنای در سطح افقی ناحیه شیب دار
	Profile Curvature (ProfileCur)	انحنای در سطح پروفیل
	Wind Effect	شاخص بدون بعد مناطق در معرض باد را نشان می دهد
	Topographic wetness index (Topo Wetness)	شاخص خیسگی پستی و بلندی
سنجش از دور (ماهواره لندست ۸) Remote sensing (Landsat 8 satellite)	Height Normalized	ارتفاع نرمال شده
	Channel Network base Level (CNBL)	فاصله عمودی تا شبکه آبراهه
	Multiresolution Index of Valley Bottom Flatness (MRVBF)	شاخص همواری دره با درجه وضوح بالا
	Slope Height	ارتفاع شیب
	Total Catchment Area	مساحت حوضه آبریز
	Multiresolution Index of the ridge top flatness(MRRTF)	شاخص چند تفکیک پذیری صافی بالای پشته
	Brightness	$((Red)^2 + (NIR)^2)^{0.5}$
	Greenness	$Greenness = -2848 (B2) - 0.2435(B3) - 0.5436(B4) + 0.7243(B5) + 0.0840(B6) - 0.1800(B7)$
	Wetness	$WI = 0.26 \times B2 + 0.21 \times B3 + 0.09 \times B4 + 0.06 \times B5 - B6 \times 0.76 - B7 \times 0.53$
	Bands	Single Bands of landsat_8 (Band 2,3,5,6)
کاربری / پوشش Usage /Cover	) Prependiclr vegetation index (PrependiPl	شاخص عمود پوشش گیاهی
	Prependiclr vegetation index(PrependiQ)	شاخص عمود پوشش گیاهی
	Land use/cover map	نقشه کاربری /پوشش زمین Usage /Cover
	Geology map (1:100,000)	نقشه زمین شناسی Geology map
اقليم Climate	PREN	بارندگی Precipitation
	Tmean	میانگین دمای سالانه Mean Annual Temperature
	Tmax	بیشترین دمای سالانه Maximum Annual Temperature
	Tmin	کمترین دمای سالانه Minimum Annual Temperature

جدول ۳- تغییرات زمانی و مکانی مساحت هر یک از کلاس‌های کاربری/پوشش اراضی از سال‌های ۲۰۰۰ تا ۲۰۲۴ در حوضه آبخیز زاینده رود

Table 3- Temporal and spatial variations in the area of each land use/land cover class from 2000 to 2024 in the Zayandeh Rud watershed

سال Year	اراضی تحت کشت Agricultural lands	اراضی مرتعی و پوشش گیاهی کوتاه Short vegetation cover or Herbaceous vegetation	اراضی مرتعی و پوشش درختی پراکنده Pastureland and scattered tree cover	مناطق مسکونی (Urban)	پهنه‌های آبی (Water body)	پوشش علفی مناطق باتلاقی Herbaceous cover of wetland areas	پوشش درختی مناطق باتلاقی Tree cover of wetland areas
2000	124855	642864	69	68790	229	2523	32
2010	123235	625191	92	88325	211	2266	41
2015	111554	624813	101	100421	325	2086	61
2024	96452	619903	117	120531	299	1997	62



شکل ۴- (الف) بارندگی سالانه برای سال ۲۰۲۴، (ب) میانگین دمای سالانه برای سال ۲۰۲۴، (پ) بیشترین دمای سالانه برای سال ۲۰۲۴، (ت) کمترین دمای سالانه برای سال ۲۰۲۴، (ث) ارتفاع استاندارد شده، (ج) موقعیت نسبی شیب.

Figure 4- (a) Annual precipitation for 2024, (b) Average annual temperature for 2024, (c) Maximum annual temperature for 2024, (d) Minimum annual temperature for 2024, (e) Standardized elevation, (f) Relative slope position

که نشان‌دهنده عملکرد برتر آن در پیش‌بینی SOC در مقایسه با دو مدل دیگر است. بالاترین مقادیر R^2 برای مدل RF برابر با ۰/۷۸ در سال ۲۰۱۵ و ۰/۸۲ در سال ۲۰۲۴ مشاهده گردید. این روند بهبود عملکرد نشان می‌دهد که دقت مدل RF با در دسترس قرار گرفتن

مدل‌سازی زمانی و مکانی کربن آلی خاک

نتایج کارایی مدل‌های یادگیری ماشین مورد استفاده در پیش‌بینی SOC بر اساس آماره‌های اعتبارسنجی مورد استفاده در جدول ۴ ارائه شده است. به‌طور کلی مدل RF بالاترین مقادیر R^2 را ارائه نموده است،

داده‌های مرتبط با زمان نمونه‌برداری افزایش نشان داد. مدل SVR نیز روند دقت کاهشی در پیش‌بینی SOC با 0.73 در سال 2024 تا 0.66 در سال 2000 نشان داد که بیانگر کاهش کارایی مدل با فاصله از داده‌های مشاهده‌ای در زمان نمونه‌برداری است. در مقابل، مدل XGBoost tree عملکرد نسبتاً پایین‌تری نسبت به دو مدل دیگر با مقادیر R^2 برابر 0.58 در سال 2000 به 0.65 در سال 2024 رسیده است.

این نتیجه نشان‌دهنده‌ی کارایی کمتر مدل XGBoost در پیش‌بینی SOC در مقایسه با مدل‌های RF و SVR بود. البته شایان ذکر است که بر اساس طبقه‌بندی ارائه شده توسط راسل و مک براتنی (Rossel & McBratney, 2008) کارایی هر سه مدل از نظر کیفیت پیش‌بینی به‌جز در مورد سال 2000 توسط مدل XGBoost خوب می‌باشد و نشان از انتخاب صحیح و کارایی این مدل‌ها در برقراری ارتباط بین SOC و متغیرهای محیطی پیش‌بینی‌کننده در طی دوره‌های مطالعاتی می‌باشند. در همین راستا نتایج ارائه شده در (جدول ۴)، نیز مؤید این است که مدل RF به‌طور کلی بهترین عملکرد را در پیش‌بینی محتوای SOC در طول زمان نشان داده است. این مدل در همه سال‌های مورد بررسی (2000 ، 2010 ، 2015 و 2024) بالاترین مقادیر CCC را داشته، به‌طوری‌که در سال 2024 مقدار این شاخص به

0.85 رسیده است. همچنین، مدل RF کمترین مقادیر RMSE و nRMSE را در مقایسه با دو مدل دیگر SVR و XGBoost داشت. این نتایج در شرایطی است که براساس طبقه‌بندی موجود برای nRMSE کارایی هر سه مدل خوب تا قابل قبول طبقه‌بندی می‌شود. متین‌فر و همکاران (Matinfar et al., 2021) در تحقیقی در مجاورت شهر خرم‌آباد در استان لرستان ضمن مقایسه کارایی چهار مدل یادگیری ماشین، RF، جنگل تصادفی چندگانه، کیوبیست، رگرسیون حداقل مربعات جزئی با رویکرد فازی (SoLIM) در پیش‌بینی SOC بیان داشتند که مدل RF در تلفیق با داده‌های پستی و بلندی و RS کارایی بیشتری را نسبت به دیگر مدل‌های مورد استفاده ارائه نمود. مشابه با تحقیق حاضر در یک مطالعه سری زمانی (1960 تا 2016) به منظور بررسی تغییرات زمانی و مکانی SOC در دشت یزد-اردکان، ایران نیز نشان داد که مدل RF ($R^2 = 0.53$) دقیق‌تر از SVR و شبکه‌های عصبی مصنوعی برای پیش‌بینی SOC در بازه‌های زمان‌های مختلف عمل نمود. همچنین محققین به اثربخشی RF در پیش‌بینی SOC در مناطق خشک و نیمه‌خشک بویژه در تحلیل‌های مربوط به بررسی روندها در طول زمان اشاره کردند (Fathizad et al., 2022).

جدول ۴- نتایج کارایی مدل‌های یادگیری ماشین در پیش‌بینی SOC در در طی دوره‌های زمانی 2000 تا 2024 در حوضه آبخیز زاینده‌رود
Table 4- Results of the efficiency of machine learning models in predicting SOC during the time periods from 2000 to 2024 in the Zayandeh Rood watershed

مدل یادگیری ماشین (۲۰۲۴) Machine Learning Model (2024)	ضریب تبیین R^2	ضریب همبستگی تطابق CCC	ریشه دوم میانگین مربعات خطا RMSE	ریشه دوم میانگین مربعات خطا نرمال شده nRMSE
RF (جنگل تصادفی)	0.82	0.85	0.25	19.0
SVR (رگرسیون بردار پشتیبان)	0.73	0.81	0.29	22.48
(درخت تقویت شده با گرادیان افراطی) XGBoost	0.65	0.76	0.31	24.80
مدل یادگیری ماشین (۲۰۱۵) Machine Learning Model (2015)	R^2	CCC	RMSE	nRMSE
RF (جنگل تصادفی)	0.78	0.84	0.26	20.15
SVR (رگرسیون بردار پشتیبان)	0.68	0.82	0.28	21.37
(درخت تقویت شده با گرادیان افراطی) XGBoost	0.64	0.76	0.32	24.80
مدل یادگیری ماشین (۲۰۱۰) Machine Learning Model (2010)	R^2	CCC	RMSE	nRMSE
RF (جنگل تصادفی)	0.72	0.80	0.28	21.70
SVR (رگرسیون بردار پشتیبان)	0.66	0.80	0.29	22.48
(درخت تقویت شده با گرادیان افراطی) XGBoost	0.60	0.72	0.35	27.34
مدل یادگیری ماشین (۲۰۰۰) Machine Learning Model (2000)	R^2	CCC	RMSE	nRMSE
RF (جنگل تصادفی)	0.72	0.77	0.28	21.87
SVR (رگرسیون بردار پشتیبان)	0.66	0.75	0.31	24.21
(درخت تقویت شده با گرادیان افراطی) XGBoost	0.58	0.70	0.33	25.78

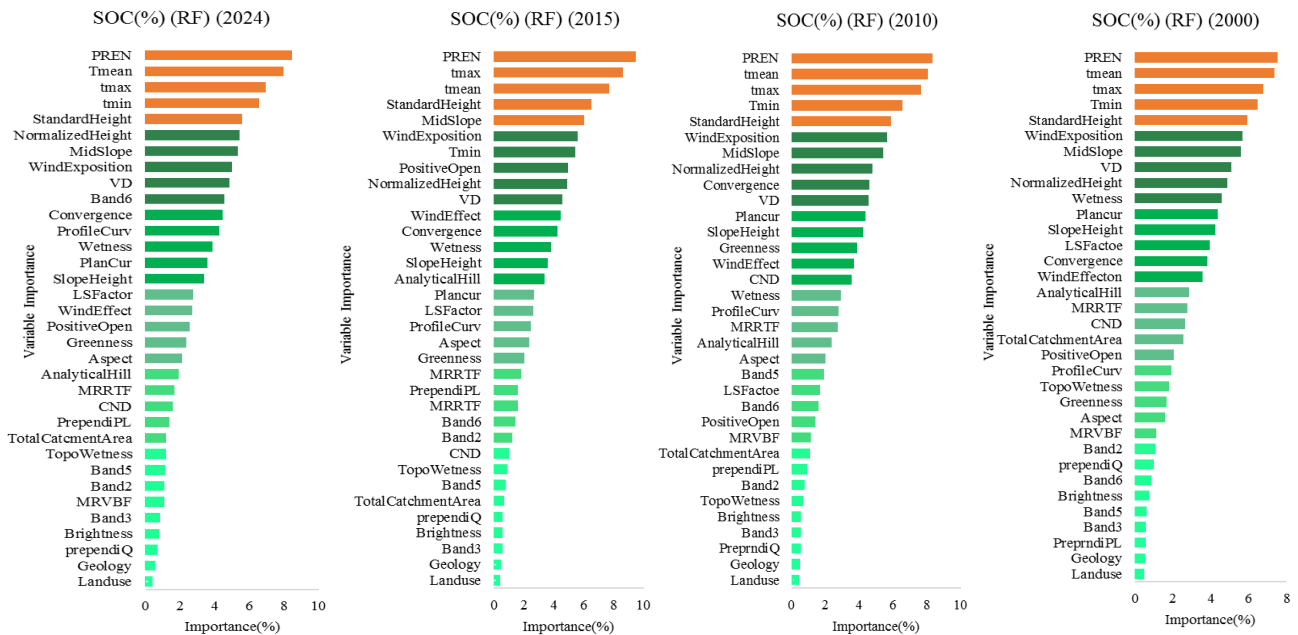
اهمیت نسبی متغیرهای محیطی

اهمیت نسبی متغیرهای محیطی در پیش‌بینی SOC بر اساس تجزیه تحلیل انجام شده توسط مدل‌های یادگیری ماشین RF در شکل ۵ ارائه شده است. با توجه به اینکه مدل RF بیشترین دقت را در پیش‌بینی SOC در همه دوره‌های زمانی ارائه نمود، بنابراین اولویت‌بندی درجه اهمیت متغیرهای محیطی بر اساس این مدل انجام گردید. به‌طور کلی نتایج اهمیت نسبی متغیرهای محیطی نشان داد که در سال ۲۰۲۴ پنج متغیر محیطی PREN، Tman، Tmax، Tmin و StandardHeight با توجیه حدود ۳۵٪ از کل واریانس SOC بیشترین نقش را در پیش‌بینی آن داشتند. برای سال ۲۰۱۵، Tmax، StandardHeight، tmean و Midslope با توجیه ۳۶٪ از تغییرات و برای سال‌های ۲۰۱۰ و ۲۰۰۰ روند مشابه با سال ۲۰۲۴ از نظر درجه‌بندی اهمیت متغیرها مشاهده شد و به‌ترتیب ۳۵ و ۳۳٪ از تغییرات را توجیه نمودند. در همین راستا می‌توان اشاره نمود که شناسایی مهمترین متغیرهای محیطی نسبتاً مشابه در فرآیند مدلسازی SOC می‌تواند به کارایی و قابلیت تعمیم مدل زمانی و مکانی مورد استفاده بیافزاید (Fathizad et al., 2022). بنابراین، چنین بنظر می‌رسد متغیرهای وابسته به عامل خاکساز اقلیم بیشترین نقش را در توجیه تغییرات مکانی SOC داشتند و پس از آن عامل پستی و بلندی Standard Height در درجه اهمیت بعدی قرار داشت. محققین در مطالعات خود وجود رابطه مثبت و معنی‌دار بین میزان بارش با محتوای SOC گزارش نمودند. در همین راستا محققان بر اهمیت نقش عوامل اقلیمی مانند میانگین دما و بارش سالانه بر ذخیره‌سازی میزان کربن آلی در خاک‌های مناطق جنگل تأکید نمودند و دریافتند که این نقش از طریق تأثیر بر میزان فتوسنتز گیاهان و تجزیه مواد آلی خاک اعمال می‌شود (Zhou et al., 2020). به‌طور مشابه (Hoyle, O'Leary, & Murphy, 2016) در مدل‌سازی مکانی SOC مشاهده نمودند که به‌ترتیب متغیرهای اقلیمی و مدیریت اراضی با توجیه ۴۳ و ۳۲٪ از کل تغییرات SOC بیشترین اهمیت را در پیش‌بینی آن داشتند. همچنین در مطالعه‌ای در دشت قزوین موسوی و همکاران (Mousavi, Sarmadian, Omid, & Bogaert, 2022) بیان داشتند که عامل اقلیمی میانگین دمای سالانه به همراه ویژگی‌های خاک (درصد رس، آهک، جرم مخصوص ظاهری) مهمترین متغیرهای موثر در پیش‌بینی آن بودند. همچنین ایشان مشاهده نمودند که متغیر پستی و بلندی StandardHeight از درجه اهمیت بالایی به‌همراه ویژگی‌های خاک و داده‌های اقلیمی در پیش‌بینی SOC برخوردار بود. سایر محققین نیز در مطالعات خود به کارایی مدل RF در مقایسه با سایر مدل‌های یادگیری ماشین تأکید نمودند (Zhao et al., 2024; Masoudi, Mousavi, Rahimabadi,)

(Panahi, & Rahmani, 2023).

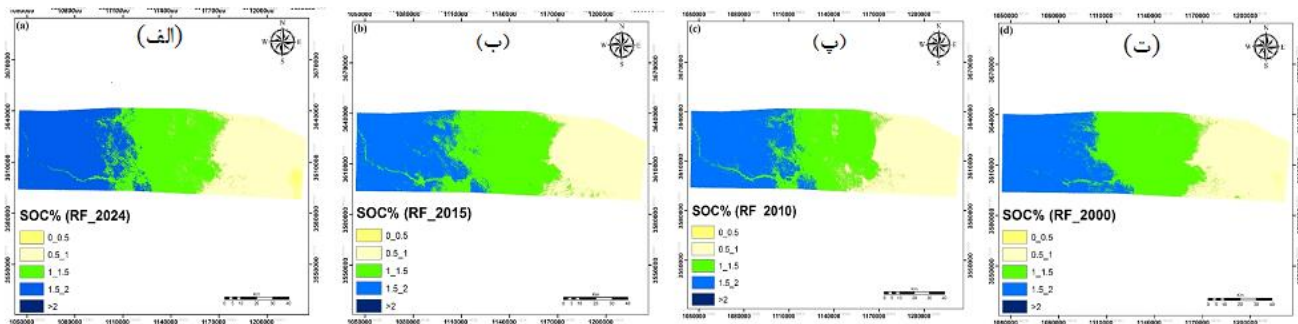
تغییرپذیری مکانی کربن آلی

درک تغییرات مکانی و زمانی SOC برای ارزیابی ذخایر جهانی کربن، ارزیابی سیاست‌های استفاده از زمین و حفظ حاصلخیزی خاک ضروری است (Schillaci et al., 2017). در همین راستا نقشه‌های رقومی تغییرات زمانی و مکانی SOC مربوط به بازه‌های زمانی مورد نظر (۲۰۰۰ تا ۲۰۲۴) تهیه و در شکل ۶ ارائه شده است. همچنین جهت تسهیل در درک بهتر تغییرات SOC در طول زمان نقشه‌های پیش‌بینی مکانی بر اساس محتوای SOC به پنج کلاس با دامنه‌های ۰/۵-، ۰/۵-۱/۵، ۱/۵-۲/۵، ۲/۵-۳/۵ و ۳/۵-۴/۵ باز طبقه‌بندی گردیدند (شکل ۶) و مساحت و درصد مربوط به هر یک از آنها در جدول ۵ ارائه شده است. همان‌گونه که در جدول ۵ مشاهده می‌گردد، بر اساس داده‌های ارائه شده، کلاس پراکنش مکانی ۰/۵-٪ از ۲۰٪ در سال ۲۰۰۰ به ۱/۱٪ در سال ۲۰۱۰ کاهش یافت، اما در سال ۲۰۲۴ به ۲/۲۴٪ افزایش یافت که نشان‌دهنده گسترش مناطق با کربن آلی بسیار پایین است. کلاس پراکنش مکانی ۱/۵-۰/۵٪ نیز از ۳/۴٪ در سال ۲۰۰۰ به ۳۱/۹٪ در سال ۲۰۱۰ افزایش یافت، اما در سال‌های ۲۰۱۵ و ۲۰۲۴ به‌ترتیب به ۳۰/۲٪ و ۳۰/۸٪ کاهش و سپس افزایش اندکی داشت، که به نوسانات کیفیت خاک اشاره دارد. کلاس‌های پراکنش مکانی ۱/۵-۱/۰ و ۱/۵-۲/۵ نیز روندهای متفاوتی را نشان دادند. کلاس پراکنش مکانی ۱/۵-۱/۰٪ از ۱/۰٪ در سال ۲۰۰۰ به ۳۵/۲٪ در سال ۲۰۱۰ کاهش یافت و سپس در سال ۲۰۱۵ به ۳۷/۰٪ افزایش یافت، اما در سال ۲۰۲۴ به ۳۳/۴۰٪ کاهش یافت، که نشان‌دهنده نوسانات و عدم ثبات در مدیریت SOC است، هرچند که در شرایط کنونی نسبت به دوره بلند مدت روند کاهشی است و بیانگر افزایش خطر تخریب و کاهش حاصلخیزی خاک می‌باشد. کلاس پراکنش مکانی ۱/۵-۲/۵٪ نیز از ۳۴/۱٪ در سال ۲۰۰۰ به ۳۶/۰٪ در سال ۲۰۱۰ افزایش یافت، اما در سال‌های ۲۰۱۵ و ۲۰۲۴ به ترتیب به ۳۲/۷٪ و ۳۳/۵٪ کاهش و سپس اندکی افزایش داشت. در نهایت، کلاس پراکنش مکانی ۲/۵-۳/۵٪ مقادیر تغییرات اندکی را نشان می‌دهد، اما از آنجاییکه افزایش SOC بیش از ۱/۵٪ بر اساس طبقه‌بندی ارائه شده توسط (Charman & Roper, 2007) حاکی از شرایط مطلوب در خاک می‌باشد در روند بلند مدت کاهشی بوده است و کلاس با مقادیر SOC بیش از ۲٪ نیز به طور کلی سهم بسیار ناچیزی در کل مساحت منطقه دارد.



شکل ۵- اهمیت نسبی متغیرهای محیطی در پیش‌بینی تغییرات مکانی و زمانی SOC خاک سطحی در حوضه آبخیز زاینده رود به‌ترتیب از چپ به راست مربوط به سال‌های ۲۰۲۴، ۲۰۱۵، ۲۰۱۰ و ۲۰۰۰ است.

Figure 5- Relative importance of environmental variables in predicting spatial and temporal changes in topsoil SOC in the Zayandeh Road watershed
From left to right, it corresponds to the years 2024, 2015, 2010, and 2000.



شکل ۶- نقشه‌های پیش‌بینی مکانی SOC در بازه‌های زمانی مورد مطالعه با استفاده از مدل F
(الف) ۲۰۲۴، (ب) ۲۰۱۵، (پ) ۲۰۱۰ و (ت) ۲۰۰۰

Figure 6- Spatial SOC prediction maps in the studied time periods (2000-2024) using the RF model
a) 2024, b) 2015, c) 2010, and d) 2000

روند دقیقاً با الگوی تغییرات مهمترین متغیرهای محیطی پیش‌بینی کننده از قبیل PREN، Tmax، tmean و tmin همراستا می‌باشد و همچنین با تغییرات تراز ارتفاعی (۱۴۰۴-۲۸۶۲ متر) در منطقه نیز همخوانی دارد، به‌طوری‌که در بخش‌های غربی با حداکثر ارتفاع،

از نظر تغییرات مکانی نیز بیشترین مقدار SOC (۵/۰-۱/۲)؛ تن (رنگ آبی) در بخش‌های غربی، مقادیر متوسط محتوای SOC در مرکز منطقه (۵/۰-۱/۱)؛ تن (رنگ سبز) می‌باشد و مقادیر کم و خیلی کم در بخش‌های شرقی (کمتر از ۱/۰)؛ تن (رنگ زرد) متمرکز می‌باشند. این

بیشترین میزان بارش (۱۲۵-۳۰۹ میلی‌متر) و حداقل دما مقادیر بالای SOC و در بخش‌های مرکزی با کاهش ارتفاع، بارش و افزایش دما

جدول ۵- طبقه‌بندی مقادیر درصد و مساحت تغییرات مکانی SOC در طی سال‌های مورد بررسی (۲۰۲۴-۲۰۰۰)

Table 5- Classification of percentage and area values of spatial changes in SOC during the years under study (2000-2024)

سال Year	محتوای کربن آلی خاک SOC content (%)									
	(0-0.5)		(0.5-1)		(1-1.5)		(1.5-2)		(2-2.5)	
	هکتار	درصد	هکتار	درصد	هکتار	درصد	هکتار	درصد	هکتار	درصد
	Hectare	%	Hectare	%	Hectare	%	Hectare	%	Hectare	%
2024	18801.7	2.24	258523.2	30.8	280346.6	33.4	281185.9	33.5	503.60	0.06
2015	467.4	0.1	253532.1	30.2	310658.3	37.0	274235.3	32.7	467.9	0.056
2010	703.2	0.1	267898.3	31.9	268002.2	31.9	302248.1	36.0	509.2	0.061
2000	1954.41	0.2	254804	30.4	295704.9	35.2	286430.3	34.1	467.46	0.06

از ۰/۰ تا ۰/۲۳٪ متغیر است، در سال ۲۰۱۵ این مقادیر بین ۰/۰ تا ۰/۲۷٪، در سال ۲۰۱۰ از ۰/۰ تا ۰/۳۳٪ و برای سال ۲۰۰۰ از ۰/۰ تا ۰/۳۸٪ متغیر می‌باشد که بیانگر افزایش تدریجی عدم قطعیت پیش‌بینی از سال ۲۰۲۴ تا ۲۰۰۰ است که یکی از دلایل آن را می‌توان به افزایش فاصله زمانی بین داده‌های مشاهداتی و متغیرهای محیطی مرتبط دانست (Fathizad et al., 2020) و با نتایج اعتبارسنجی مدل‌های یادگیری ماشین در پیش‌بینی SOC همراستا می‌باشد. به‌طور کلی بیشترین میزان عدم قطعیت مربوط به مقادیر بالای SOC در طول دوره‌های زمانی مورد بررسی بود، که تعداد داده‌های مشاهداتی و نیز متغیرهای محیطی مورد استفاده می‌تواند از دلایل آن باشد (Zaho et al., 2024). در مطالعه‌ی مقایسه‌ای رحمانی و همکاران (Rahmani, Sarmadian, & Arefi, 2022) نیز گزارش کردند که روش بوتسراپ نسبت به روش k-فولد از کارایی بیشتری و نتایج منطبق با واقعیت را در پیش‌بینی ضخامت خاک سطحی در دشت قزوین گزارش نمودند. لطف‌اللهی و همکاران (Lotfollahi, Delavar, Biswas, Fatehi, & Scholten, 2023) همچنین بر این باور بودند که روش بوتسراپ تخمین‌های قابل اعتمادی از عدم قطعیت‌های مرتبط با این پیش‌بینی‌ها ارائه می‌دهد. منابع عدم قطعیت می‌تواند مرتبط به خطاهای متغیرهای محیطی، مدل‌یادگیری ماشین مورد استفاده، خطاهای مربوط به نمونه برداری میدانی و تجزیه و تحلیل‌های آزمایشگاهی است (Nelson, Bishop, Triantafilis, & Odeh, 2011).

نتیجه‌گیری

بر اساس اهداف تحقیق، مدلسازی معکوس تغییرات زمانی و مکانی SOC در لایه سطحی خاک با استفاده از متغیرهای محیطی و سه مدل یادگیری ماشین RF، SVR و XGBoost در حوضه آبخیز زاینده رود

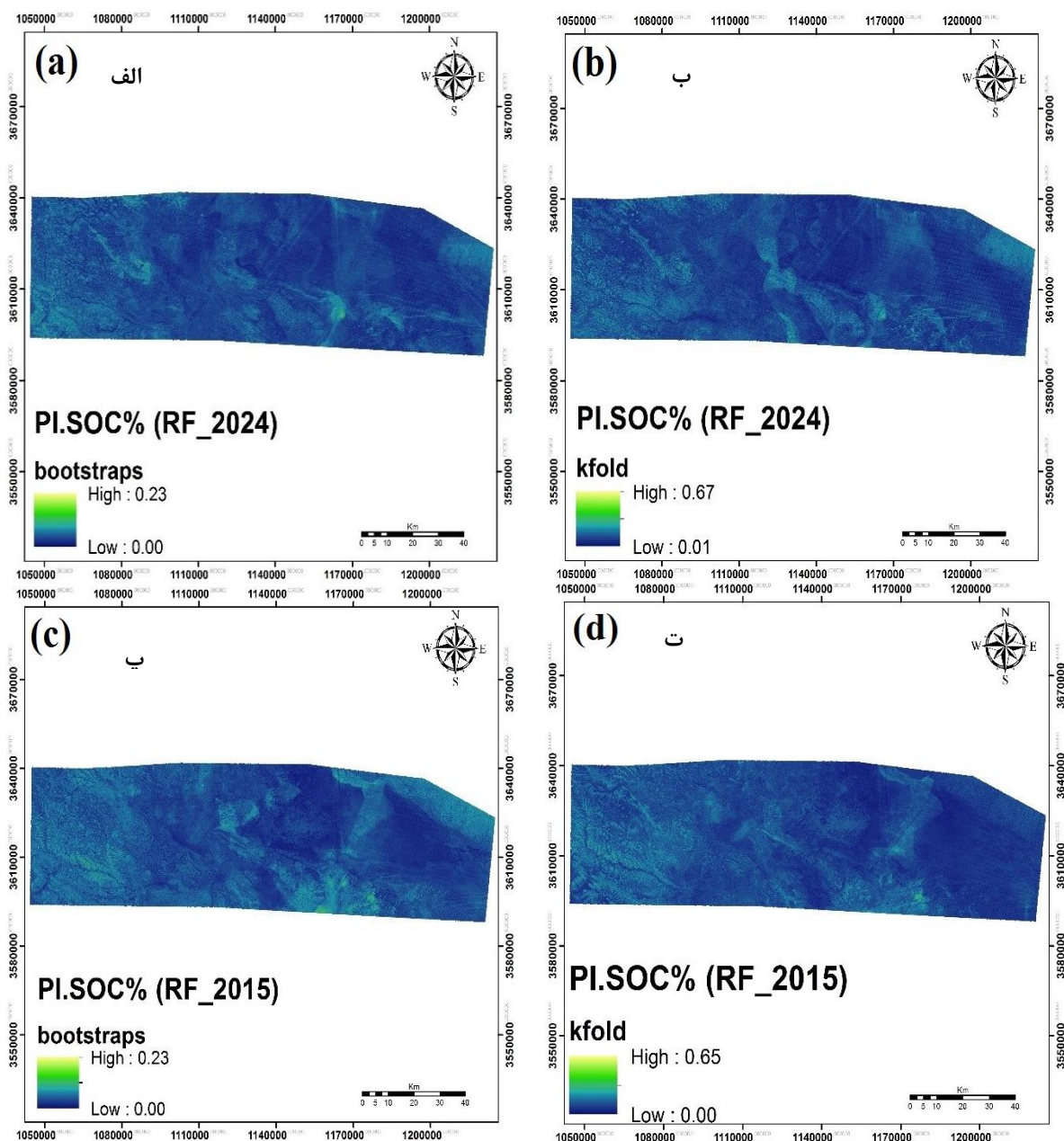
از نظر کاربری اراضی بیشترین تراکم پوشش گیاهی از جمله اراضی زراعی، باغات و مناطق مرتعی با پوشش علفی به‌طور عمده در غرب، جنوب غرب و بصورت یک نوار یا زون در مجاورت رودخانه زاینده رود از غرب به مرکز منطقه می‌باشد. در مقابل از مرکز منطقه به سمت شرق با کاهش سطح اراضی کشاورزی و افزایش مناطق بدون پوشش و یا با پوشش گیاهی بسیار کم مقدار SOC کاهش قابل توجهی نشان داد. همچنین بنظر می‌رسد بدلیل عدم رعایت اصول حفاظت خاک و کاهش سطح اراضی زراعی و توسعه مناطق مسکونی در بلند مدت، همراه با تغییر کاربری اراضی یکی دیگر از دلایل کاهش مناطق با SOC بالا می‌باشد. در در همین راستا همزه پور و همکاران (Hamzehpour, Shafizadeh-Moghadam, & Valavi, 2019) بیشترین محتوای کربن آلی را در مناطق جنگلی و مراتع (علفزارها) گزارش نمودند، همچنین بیان داشتند که در اراضی زراعی محتوای SOC می‌تواند بر اساس نوع مدیریت زمین توسط کشاورزان متغیر باشد.

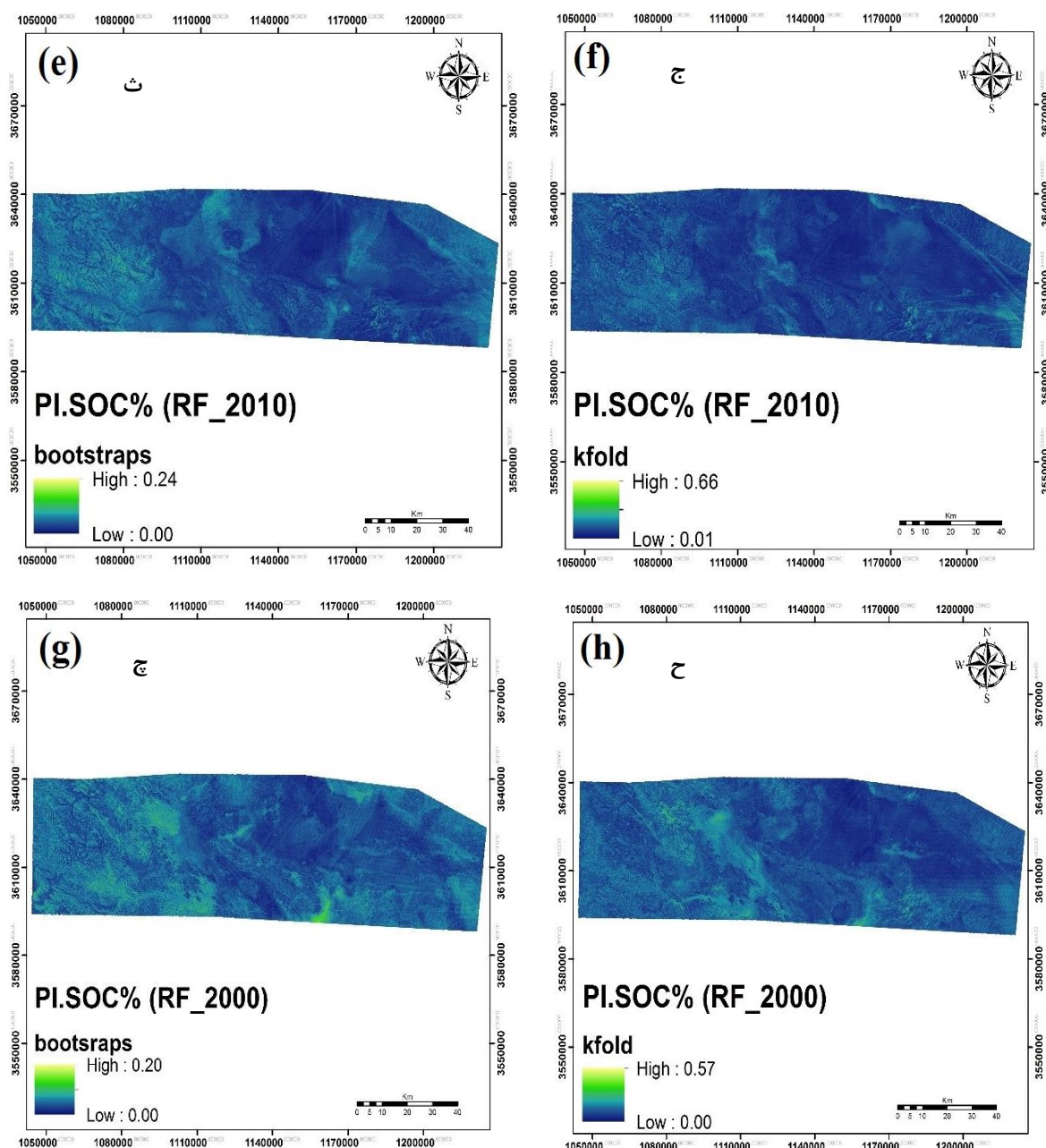
تحلیل عدم قطعیت پیش‌بینی کربن آلی خاک

بر اساس اهداف تحقیق برای بررسی دامن‌های اطمینان پیش‌بینی مدل RF به‌عنوان موفق‌ترین الگوریتم یادگیری ماشین در پیش‌بینی تغییرات زمانی و مکانی SOC بود از دو روش برآورد نقشه‌ی عدم قطعیت پیش‌بین شامل بوتسراپ و k-فولد استفاده گردید (شکل ۷). همچنین نتایج مقایسه‌ی کارایی این دو روش نیز بر اساس دو آماره انحراف معیار (SD) و میانگین عدم قطعیت^۱ کمی‌سازی گردید، و نتایج نشان داد که روش بوتسراپ با مقادیر کمتر هر دو آماره SD و میانگین عدم قطعیت کمتر کارایی بالاتری را نسبت به روش k-فولد ارائه نمود (شکل ۸). همچنین بر اساس نتایج کمی‌سازی شده عدم قطعیت به روش بوتسراپ به‌طور کلی مقادیر حداقل و حداکثر آن در سال ۲۰۲۴

ارائه بالاترین مقادیر R^2 عملکرد بهتری نسبت به دو مدل دیگر ارائه نمود.

انجام گردید. نتایج ارزیابی کارایی مدل‌های یادگیری ماشین در پیش‌بینی SOC بر اساس آماره R^2 نشان داد که، کارایی کلیه مدل‌ها از سال ۲۰۲۴ به سمت ۲۰۰۰ کاهش یافت. با این حال مدل RF



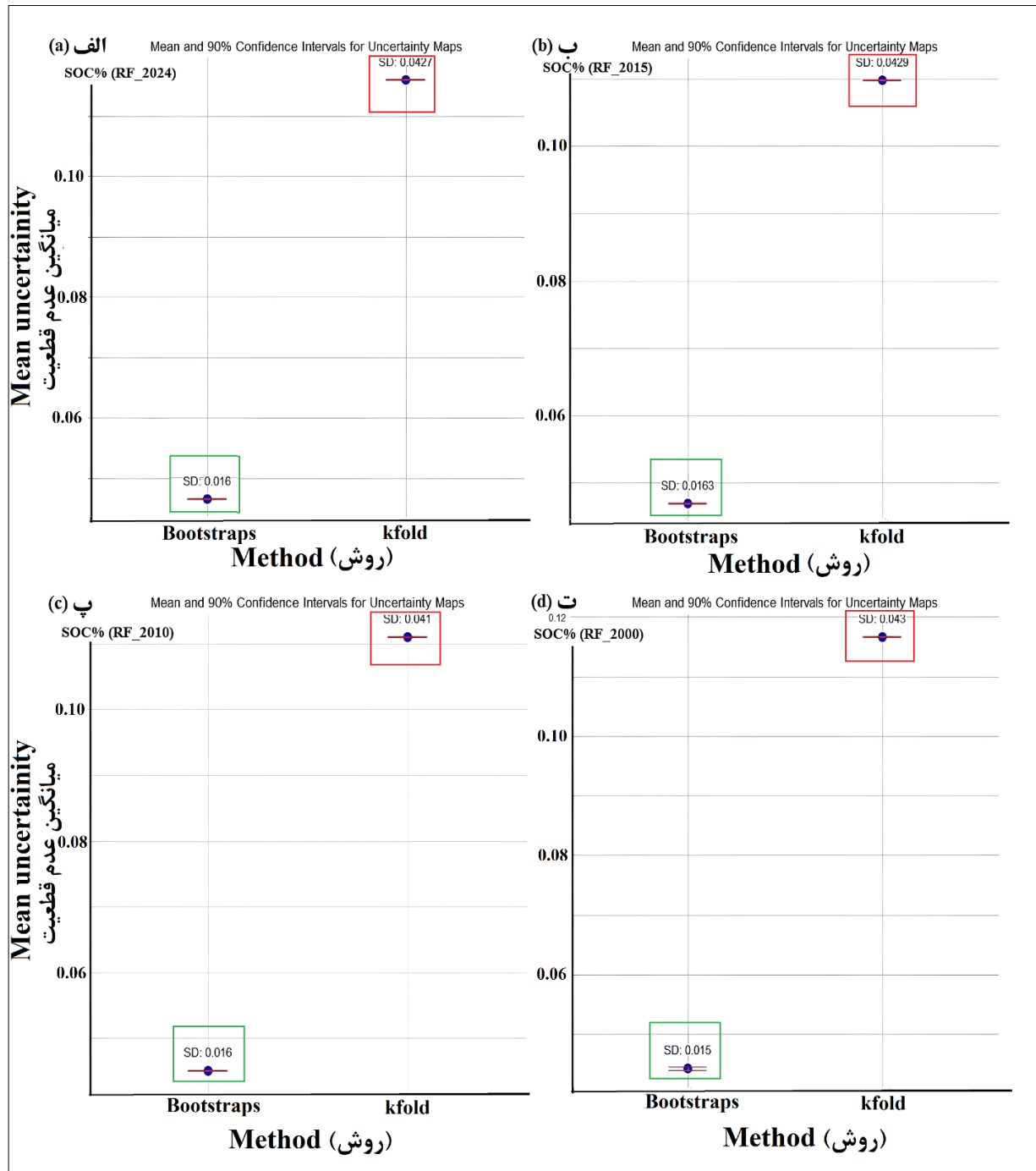


شکل ۷- نقشه‌های عدم قطعیت SOC با دو روش بوتسترپت و k-فولد در بازه‌های زمانی مورد مطالعه با استفاده از مدل RF

بترتیب الف) نقشه عدم قطعیت پیش‌بینی SOC به روش بوتسترپت سال ۲۰۲۴، ب) نقشه عدم قطعیت پیش‌بینی SOC به روش k-فولد سال ۲۰۲۴، ج) نقشه عدم قطعیت پیش‌بینی SOC به روش بوتسترپت سال ۲۰۱۵، د) نقشه عدم قطعیت پیش‌بینی SOC به روش k-فولد سال ۲۰۱۵، ه) نقشه عدم قطعیت پیش‌بینی SOC به روش بوتسترپت سال ۲۰۱۰، و) نقشه عدم قطعیت پیش‌بینی SOC به روش k-فولد سال ۲۰۱۰، ز) نقشه عدم قطعیت پیش‌بینی SOC به روش بوتسترپت سال ۲۰۰۰، ح) نقشه عدم قطعیت پیش‌بینی SOC به روش k-فولد سال ۲۰۰۰.

Figure 7- Uncertainty maps of SOC using two methods: bootstrap and k-fold over the studied time periods using the RF model

(a) Uncertainty map of SOC prediction using the bootstrap method for 2024, (b) Uncertainty map of SOC prediction using k-fold for 2024, (c) Uncertainty map of SOC prediction using the bootstrap method for 2015, (d) Uncertainty map of SOC prediction using k-fold for 2015, (e) Uncertainty map of SOC prediction using the bootstrap method for 2010, (f) Uncertainty map of SOC prediction using k-fold for 2010 (g) Uncertainty map of SOC prediction using the bootstrap method for 2000, (h) Uncertainty map of SOC prediction using k-fold for 2000.



شکل ۸- مقایسه‌ی کارایی این دو روش بوتسراپت و کافولد برای پیش‌بینی SOC بر اساس دو آماره انحراف معیار (SD) و میانگین عدم قطعیت در بازه‌های زمانی مورد مطالعه با استفاده از مدل RF

باکس سبز و قرمز به‌ترتیب نتایج مدل بوتسراپت و کافولد. به‌ترتیب الف) مقایسه کارایی دو رویکرد عدم قطعیت برای سال ۲۰۲۴، ب) مقایسه کارایی دو رویکرد عدم قطعیت برای سال ۲۰۱۵، پ) مقایسه کارایی دو رویکرد عدم قطعیت برای سال ۲۰۱۰، ت) مقایسه کارایی دو رویکرد عدم قطعیت برای سال ۲۰۰۰.

Figure 8- Comparison of the performance of the two methods, bootstrap and k-fold, for predicting SOC based on two statistics: standard deviation (SD) and mean uncertainty over the studied time periods using the RF model

The green and red boxes represent the results of the bootstrap and k-fold models, respectively. (a) Comparison of the performance of the two uncertainty approaches for the year 2024, (b) Comparison of the performance of the two uncertainty approaches for the year 2015, (c) Comparison of the performance of the two uncertainty approaches for the year 2010, (d) Comparison of the performance of the two uncertainty approaches for the year 2000.

نشان‌دهنده نوسانات و تغییرات قابل توجهی در کیفیت خاک در طول زمان است. به‌طور کلی، افزایش مساحت مناطق با محتوای SOC بسیار پایین (۵۰-٪) و تخریب خاک، در کنار نوسانات در کیفیت خاک، حاکی از عدم ثبات در مدیریت کربن آلی و خطرات ناشی از تخریب و کاهش حاصلخیزی خاک می‌باشد. همچنین، تغییرات مکانی در محتوای SOC با الگوهای تغییرات مکانی فاکتورهای اقلیمی و پستی و بلندی همخوانی داشت، به‌طوری‌که مناطق با بارش بیشتر و دمای پایین‌تر، SOC بالاتری را نشان می‌دهند. در نهایت، کاربری اراضی و کاهش سطح اراضی زراعی نیز به‌عنوان دیگر عوامل مؤثر که حاکی از نقش انسان در کاهش محتوای SOC در عمق سطحی خاک‌های مورد مطالعه در حوضه آبخیز زاینده رود شناسایی شده‌اند. در پایان پیشنهاد می‌گردد رویکرد مورد استفاده در این پژوهش برای یافتن روند تغییرات مکانی سایر ویژگی‌های کلیدی و مؤثر بر کیفیت خاک با استفاده از متغیرهای محیطی و مدل RF در طول زمان مورد استفاده قرار گیرد.

بالاترین مقادیر R^2 برای مدل RF به‌ترتیب مربوط به سال‌های ۲۰۲۴ و ۲۰۱۵ بود. مدل SVR نیز عملکردی حدواسط را ارائه نمود اگرچه مدل XGBoost tree عملکرد نسبتاً ضعیف‌تری را نسبت به دو مدل دیگر داشت. با این حال، هر سه مدل یادگیری ماشین مورد استفاده از نظر کیفیت پیش‌بینی عملکرد خوبی داشتند و این نتیجه نشان‌دهنده انتخاب مناسب آنها در برقراری ارتباط بین SOC و متغیرهای محیطی پیش‌بینی‌کننده هستند. همچنین نتایج ارزیابی کمی عدم قطعیت پیش‌بینی SOC نشان داد که روش بوتسرایت با مقادیر کمتر هر دو آماره SD و میانگین عدم قطعیت کارایی بالاتری را نسبت به روش k-فولد ارائه نمود. در مورد اهمیت نسبی متغیرهای محیطی در پیش‌بینی SOC نتایج تجزیه و تحلیل در تمامی بازه‌های زمانی نشان داد عوامل اقلیمی و سپس عوامل وابسته به پستی و بلندی بیشترین نقش را نسبت به دیگر متغیرهای محیطی (کاربری اراضی، زمین‌شناسی و RS) در توجیه تغییرات مکانی SOC داشتند. نتایج تحلیل تغییرات مکانی SOC نیز

References

- Adedeji, O.H. (2019). Geospatial information as a tool for soil resource information, management and decision support in Nigeria. *Journal of Applied Sciences and Environmental Management*, 23(12), 2107-2116. <https://doi.org/10.4314/jasem.v23i12.5>
- Adeniyi, O.D., Brenning, A., & Maerker, M. (2024). Spatial prediction of soil organic carbon: Combining machine learning with residual kriging in an agricultural lowland area (Lombardy region, Italy). *Geoderma*, 448, 116953. <https://doi.org/10.1016/j.geoderma.2024.116953>
- Afshar, F.A., Ayoubi, S., & Jafari, A. (2018). The extrapolation of soil great groups using multinomial logistic regression at regional scale in arid regions of Iran. *Geoderma*, 315, 36-48. <https://doi.org/10.1016/j.geoderma.2017.11.030>
- Behrens, T., Zhu, A.X., Schmidt, K., & Scholten, T. (2010). Multi-scale digital terrain analysis and feature selection for digital soil mapping. *Geoderma*, 155(3-4), 175-185. <https://doi.org/10.1016/j.geoderma.2009.07.010>
- Ben Brahim, A., & Limam, M. (2018). Ensemble feature selection for high dimensional data: a new method and a comparative study. *Advances in Data Analysis and Classification*, 12(4), 937-952. <https://doi.org/10.1007/s11634-017-0285-y>
- Benslama, A., Lucas, I.G., Jordan Vidal, M.M., Almendro-Candel, M.B., & Navarro-Pedreño, J. (2024). Carbon and nitrogen stocks in topsoil under different land use/land cover types in the southeast of Spain. *AgriEngineering*, 6(1), 396-408. <https://doi.org/10.3390/agriengineering6010024>
- Canero, F.M., Rodriguez-Galiano, V., & Aragonés, D. (2024). Machine learning and feature selection for soil spectroscopy. An evaluation of random forest wrappers to predict soil organic matter, clay, and carbonates. *Heliyon*, 10(9). <https://doi.org/10.1016/j.heliyon.2024.e30228>
- Charman, P.E.V., & Roper, M.M. (2007). *Soil organic matter*. In: Charman, P.E.V. and Murphy, B.W., Eds., *Soils—Their Properties and Management*, 3rd Edition, Oxford University Press, Melbourne, 276-285.
- Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794). <https://doi.org/10.1145/2939672.2939785>
- Fathizad, H., Ardakani, M.A.H., Heung, B., Sodaiezadeh, H., Rahmani, A., Fathabadi, A., & Taghizadeh-Mehrjardi, R. (2020). Spatio-temporal dynamic of soil quality in the central Iranian desert modeled with machine learning and digital soil assessment techniques. *Ecological Indicators*, 118, 106736. <https://doi.org/10.1016/j.ecolind.2020.106736>
- Fathizad, H., Taghizadeh-Mehrjardi, R., Hakimzadeh Ardakani, M.A., Zeraatpisheh, M., Heung, B., & Scholten, T. (2022). Spatiotemporal assessment of soil organic carbon change using machine-learning in arid regions. *Agronomy*, 12(3), 628. <https://doi.org/10.3390/agronomy12030628>
- Fathy, H., Heydari, M., Fathizad, H., Hosseinzadeh, J., Najafifar, A., Mousavi, S.R., & Heung, B. (2025). From

- forest to farmland: Tracking time series variations in soil quality in semiarid oak forest. *Geoderma Regional*, 42, e00974. <https://doi.org/10.1016/j.geodrs.2025.e00974>
13. Garosi, Y., Ayoubi, S., Nussbaum, M., & Sheklabadi, M. (2022). Effects of different sources and spatial resolutions of environmental covariates on predicting soil organic carbon using machine learning in a semi-arid region of Iran. *Geoderma Regional*, 29, e00513. <https://doi.org/10.1016/j.geodrs.2022.e00513>
 14. Hoyle, F.C., O'Leary, R.A., & Murphy, D.V. (2016). Spatially governed climate factors dominate management in determining the quantity and distribution of soil organic carbon in dryland agricultural systems. *Scientific Reports*, 6(1), 31468. <https://doi.org/10.1038/srep31468>
 15. Hamzehpour, N., Shafizadeh-Moghadam, H., & Valavi, R. (2019). Exploring the driving forces and digital mapping of soil organic carbon using remote sensing and soil texture. *Catena*, 182, 104141. <https://doi.org/10.1016/j.catena.2019.104141>
 16. Kasraei, B., Schmidt, M.G., Zhang, J., Bulmer, C.E., Filatow, D.S., Pennell, T., & Heung, B. (2024). A framework for optimizing environmental covariates to support model interpretability in digital soil mapping. *Geoderma*, 445, 116873. <https://doi.org/10.1016/j.geoderma.2024.116873>
 17. Kulikova, A.I., Chechenkov, P.D., Osipova, M.S., Shopina, O.V., & Semenov, I.N. (2023). Comparative analysis of the results of traditional and digital large-scale soil mapping on the example of a key site in the Smolenskoe Poozerye National Park. *Eurasian Soil Science*, 56(3), 271-277. <https://doi.org/10.1134/s1064229322602281>
 18. Laban, P., Metternicht, G., & Davies, J. (2018). Soil biodiversity and soil organic carbon: keeping drylands alive. *Gland, Switzerland: IUCN*, 10. <https://doi.org/10.2305/IUCN.CH.2018.03.en>
 19. Lamichhane, S., Kumar, L., & Wilson, B. (2019). Digital soil mapping algorithms and covariates for soil organic carbon mapping and their implications: A review. *Geoderma*, 352, 395-413. <https://doi.org/10.1016/j.geoderma.2019.05.031>
 20. Lotfollahi, L., Delavar, M.A., Biswas, A., Fatehi, S., & Scholten, T. (2023). Spectral prediction of soil salinity and alkalinity indicators using visible, near-, and mid-infrared spectroscopy. *Journal of Environmental Management*, 345, 118854. <https://doi.org/10.1016/j.jenvman.2023.118854>
 21. Ma, Z., Shi, Z., Zhou, Y., Xu, J., Yu, W., & Yang, Y. (2017). A spatial data mining algorithm for downscaling TMPA 3B43 V7 data over the Qinghai-Tibet Plateau with the effects of systematic anomalies removed. *Remote Sensing of Environment*, 200, 378-395. <https://doi.org/10.1016/j.rse.2017.08.023>
 22. Madani, K. (2007, May). Water Transfer and watershed development: A system dynamics approach. In World Environmental and Water Resources Congress 2007: Restoring Our Natural Habitat (pp. 1-15). [https://doi.org/10.1061/40927\(243\)551](https://doi.org/10.1061/40927(243)551)
 23. Mantero, P., Moser, G., & Serpico, S.B. (2005). Partially supervised classification of remote sensing images through SVM-based probability density estimation. *IEEE Transactions on Geoscience and Remote Sensing*, 43(3), 559-570. <http://doi.org/10.1109/TGRS.2004.842022>
 24. Mallah, S., Delsouz Khaki, B., Davatgar, N., Poppiel, R.R., & Demattê, J.A. (2022). Digital mapping of topsoil texture classes using a hybridized classical statistics-artificial neural networks approach and relief data. *AgriEngineering*, 5(1), 40-64. <https://doi.org/10.3390/agriengineering5010004>
 25. Masoudi, R., Mousavi, S.R., Rahimabadi, P.D., Panahi, M., & Rahmani, A. (2023). Assessing data mining algorithms to predict the quality of groundwater resources for determining irrigation hazard. *Environmental Monitoring and Assessment*, 195(2), 319. <https://doi.org/10.1007/s10661-022-10909-9>
 26. Matinfar, H.R., Maghsodi, Z., Mousavi, S.R., & Rahmani, A. (2021). Evaluation and prediction of topsoil organic carbon using machine learning and hybrid models at a Field-scale. *Catena*, 202, 105258. <https://doi.org/10.1016/j.catena.2021.105258>
 27. 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. <https://doi.org/10.1016/j.catena.2023.107392>
 28. 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. <https://doi.org/10.1016/j.measurement.2022.111706>
 29. Mousavi, S.R., Jahandideh Mahjenabadi, V.A., Khoshru, B., & Rezaei, M. (2024). Spatial prediction of winter wheat yield gap: agro-climatic model and machine learning approaches. *Frontiers in Plant Science*, 14, 1309171. <https://doi.org/10.3389/fpls.2023.1309171>
 30. Muslim, W.A., Albayati, T.M., & Al-Nasri, S.K. (2022). Decontamination of actual radioactive wastewater containing 137Cs using bentonite as a natural adsorbent: equilibrium, kinetics, and thermodynamic studies. *Scientific Reports*, 12(1), 13837. <https://doi.org/10.1038/s41598-022-18202-y>
 31. Nelson, M.A., Bishop, T.F.A., Triantafilis, J., & Odeh, I.O.A. (2011). An error budget for different sources of error

- in digital soil mapping. *European Journal of Soil Science*, 62(3), 417-430. <https://doi.org/10.1111/j.13652389.2011.01365.x>
32. Parras-Alcántara, L., Díaz-Jaimes, L., & Lozano-García, B. (2015). Management effects on soil organic carbon stock in Mediterranean open rangelands—treeless grasslands. *Land Degradation & Development*, 26(1), 22-34. <https://doi.org/10.1002/ldr.2269>
 33. Rabbi, S.M.F., Roy, B.R., Miah, M.M., Amin, M.S., & Khandakar, T. (2014). Spatial variability of physical soil quality index of an agricultural field. *Applied and Environmental Soil Science*, 2014(1), 379012. <https://doi.org/10.1155/2014/379012>
 34. Rahmani, A., Sarmadian, F., & Arefi, H. (2022). Digital mapping of top-soil thickness and associated uncertainty using machine learning approach in some part of arid and semi-arid lands of Qazvin plain. *Iranian Journal of Soil and Water Research*, 53(3), 585-602. <https://doi.org/10.47176/jwss.24.2.32993>
 35. Rossel, R.V., & McBratney, A.B. (2008). Diffuse reflectance spectroscopy as a tool for digital soil mapping. In *Digital soil mapping with limited data* (pp. 165-172). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-1-4020-8592-5_13
 36. Rostaminia, M., Rahmani, A., Mousavi, S.R., Taghizadeh-Mehrjardi, R., & Maghsodi, Z. (2021). Spatial prediction of soil organic carbon stocks in an arid rangeland using machine learning algorithms. *Environmental Monitoring and Assessment*, 193, 1-17. <https://doi.org/10.1007/s10661-021-09543-8>
 37. Schillaci, C., Acutis, M., Lombardo, L., Lipani, A., Fantappie, M., Märker, M., & Saia, S. (2017). Spatio-temporal topsoil organic carbon mapping of a semi-arid Mediterranean region: The role of land use, soil texture, topographic indices and the influence of remote sensing data to modelling. *Science of the Total Environment*, 601, 821-832. <https://doi.org/10.1016/j.scitotenv.2017.05.239>
 38. Song XiaoDong, S.X., Liu Feng, L.F., Zhang GanLin, Z.G., Li DeCheng, L.D., Zhao YuGuo, Z.Y., & Yang JinLing, Y.J. (2017). Mapping soil organic carbon using local terrain attributes: a comparison of different polynomial models. [https://doi.org/10.1016/S1002-0160\(17\)60445-4](https://doi.org/10.1016/S1002-0160(17)60445-4)
 39. Su, L., Heydari, M., Jaafarzadeh, M.S., Mousavi, S.R., Rezaei, M., Fathizad, H., & Heung, B. (2024). Incorporating forest canopy openness and environmental covariates in predicting soil organic carbon in oak forest. *Soil and Tillage Research*, 244, 106220. <https://doi.org/10.1016/j.still.2024.106220>
 40. Taghizadeh-Mehrjardi, R., Neupane, R., Sood, K., & Kumar, S. (2017). Artificial bee colony feature selection algorithm combined with machine learning algorithms to predict vertical and lateral distribution of soil organic matter in South Dakota, USA. *Carbon Management*, 8(3), 277-291. <https://doi.org/10.1080/17583004.2017.1330593>
 41. Tibshirani, R.J., & Efron, B. (1993). An introduction to the bootstrap. *Monographs on Statistics and Applied Probability*, 57(1), 1-436.
 42. Van Wambeke, A.R. (2000). The Newhall simulation model for estimating soil moisture and temperature regimes. Department of Crop and Soil Sciences. Cornell University, Ithaca, NY.
 43. Vanwinckelen, G., & Blockeel, H. (2012). On estimating model accuracy with repeated cross-validation. In *BeneLearn 2012: Proceedings of the 21st Belgian-Dutch conference on machine learning* (pp. 39-44).
 44. Vasiliniuc, I., Patriche, C.V., Pîrnău, R., & Roșca, B. (2013). Statistical spatial models of soil parameters. An approach using different methods at different scales. *Environmental Engineering and Management Journal*, 12(3), 457-464. <https://doi.org/10.30638/eemj.2013.057>
 45. Voltz, M., Arrouays, D., Bispo, A., Lagacherie, P., Laroche, B., Lemerrier, B., & Schnebel, N. (2020). Possible futures of soil-mapping in France. *Geoderma Regional*, 23, e00334. <https://doi.org/10.1016/j.geodrs.2020.e00334>
 46. Walkley, A., & Black, I.A. (1934). An examination of the Degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. *Soil Science*, 37(1), 29-38. <https://doi.org/10.1097/00010694-193401000-00003>
 47. Yang, L., Jia, W., Shi, Y., Zhang, Z., Xiong, H., & Zhu, G. (2020). Spatiotemporal differentiation of soil organic carbon of grassland and its relationship with soil physicochemical properties on the northern slope of Qilian mountains, China. *Sustainability*, 12(22), 9396. <https://doi.org/10.3390/su12229396>
 48. Zaheri Abdehvand, Z., Karimi, D., Rangzan, K., & Mousavi, S.R. (2024). Assessment of soil fertility and nutrient management strategies in calcareous soils of Khuzestan province: a case study using the Nutrient Index Value method. *Environmental Monitoring and Assessment*, 196(6), 503. <https://doi.org/10.1007/s10661-024-12665-4>
 49. Zeraatpisheh, M., Jafari, A., Bodaghabadi, M.B., Ayoubi, S., Taghizadeh-Mehrjardi, R., Toomanian, N., & Xu, M. (2020). Conventional and digital soil mapping in Iran: Past, present, and future. *Catena*, 188, 104424. <https://doi.org/10.1016/j.catena.2019.104424>
 50. Zhao, S., Ayoubi, S., Mousavi, S.R., Mireei, S.A., Shahpouri, F., Wu, S.X., & Tian, C.Y. (2024). Integrating proximal soil sensing data and environmental variables to enhance the prediction accuracy for soil salinity and

- sodicity in a region of Xinjiang Province, China. *Journal of Environmental Management*, 364, 121311. <https://doi.org/10.1016/j.jenvman.2024.121311>
51. Zhou, T., Geng, Y., Chen, J., Liu, M., Haase, D., & Lausch, A. (2020). Mapping soil organic carbon content using multi-source remote sensing variables in the Heihe River Basin in China. *Ecological Indicators*, 114, 106288. <https://doi.org/10.1016/j.ecolind.2020.106288>