



Predicting and Mapping of Soil Organic Carbon Stock Using Machine Learning Algorithm

Seyyed Erfan Khamoshi¹, Fereydoon Sarmadian^{✉2}, Mahmoud Omid³

1. Department of Soil Science, College of Agriculture and Natural Resources, University of Tehran, Karaj, Iran. Email:

khamoshierfan@ut.ac.ir

2. Corresponding Author, Department of Soil Science, College of Agriculture and Natural Resources, University of Tehran, Karaj, Iran. Email: fsarmad@ut.ac.ir

3. Department of Agricultural Machinery Engineering, College of Agriculture and Natural Resources, University of Tehran, Karaj, Iran. omid@ut.ac.ir

Article Info

ABSTRACT

Article type: Research Article

Article history:

Received: Aug. 29, 2022

Revised: Oct. 2, 2022

Accepted: Nov. 28, 2022

Published online: Jan. 22, 2022

Keywords:

Machine learning,
Soil organic carbon Stock,
Remote sensing,
Environmental covariates.

Investigation of soil organic carbon stock (SOCS) in agricultural lands and the role of factors affecting its variability and digital modeling are important for predicting possible scenarios of future carbon stock. The purpose of this study was to investigate the spatial variability and to estimate SOCS at 0 to 100 cm depth based on two generation of machine learning approaches in a part of Qazvin plain. SOCS of about 211 legacy soil data were prepared. The environmental variables including 11 geomorphometric variables and 25 spectral indices with 10-meter spatial resolution were used. Further, the dataset was divided into two parts: 70% of data were chosen as training and 30% of data for model validation. Two algorithm were used for SOCS modeling in the study area. Validation results indicated that the QRF had a higher coefficient of determination than the RF. According to the results of the relative importance of environmental variables, DEM and Valley depth parameters are more important in the spatial modeling of SOCS than other variables. Generally, it is suggested to investigate hybrid models in the process of modeling secondary soil characteristics.

Cite this article: 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. <https://doi.org/10.22059/ijswr.2022.346976.669339>

© The Author(s).

Publisher: University of Tehran Press.

DOI: <https://doi.org/10.22059/ijswr.2022.346976.669339>



تهیه نقشه رقومی کربن آلی ذخیره شده در خاک با استفاده از روش های یادگیری ماشینی

سیدعرفان خاموشی^۱، فریدون سرمیدیان^۲، محمود امید^۳۱. گروه علوم ومهندسی خاک، دانشکدگان کشاورزی و منابع طبیعی دانشگاه تهران. کرج، ایران، ایمیل: khamoshierfan@ut.ac.ir۲. نویسنده مسئول، گروه علوم ومهندسی خاک، دانشکدگان کشاورزی و منابع طبیعی دانشگاه تهران. کرج، ایران، ایمیل: farmad@ut.ac.ir۳. گروه مهندسی ماشین های کشاورزی، دانشکدگان کشاورزی و منابع طبیعی دانشگاه تهران. کرج، ایران، ایمیل: omid@ut.ac.ir

اطلاعات مقاله چکیده

نوع مقاله: مقاله پژوهشی

بررسی ذخایر کربن آلی خاک (SOCS) در زمین های کشاورزی و نقش عوامل مؤثر بر تغییرپذیری آن و مدل سازی رقومی برای پیش بینی سناریوهای احتمالی ذخایر کربن در آینده مهم است. هدف از این مطالعه بررسی تنوع مکانی و برآورد مقدار کربن آلی ذخیره در عمق ۱۰۰ سانتی متری بر اساس دو نسل از مدل های یادگیری ماشینی در بخشی از دشت قزوین است. محتوای کربن آلی خاک، ۲۱۱ نمونه خاک که اطلاعات آن از قبل جمع آوری شده و موجود بود استخراج گردید. از متغیرهای محیطی، ۱۱ متغیر بر پایه مدل رقومی ارتفاع و ۲۵ شاخص طیفی مستخرج از تصاویر ماهواره های لندست ۸ و سنتینل ۲ با قدرت تفکیک مکانی ۱۰ متر استفاده شد. علاوه بر این، مجموعه داده ها به دو بخش تقسیم شد: ۷۰ درصد از داده ها به عنوان آموزش و ۳۰ درصد از داده ها برای اعتبارسنجی مدل انتخاب شدند. جهت مدل سازی کربن ذخیره آلی در منطقه مورد مطالعه از دو مدل جنگل تصادفی (RF) و جنگل تصادفی کوانتایل (QRF) استفاده شد. نتایج اعتبارسنجی نشان داد که استفاده از مدل QRF ضریب تعیین بالاتری نسبت به مدل RF دارد. با توجه به نتایج اهمیت نسبی متغیرهای محیطی، پارامترهای مدل رقومی ارتفاع و عمق دره نسبت به سایر متغیرها در مدل سازی فضایی SOCS اهمیت بیشتری دارند. به طور کلی، پیشنهاد می شود که در فرآیند مدل سازی ویژگی های ثانویه خاک، به بررسی مدل های هیبریدی پرداخته شود.

تاریخ دریافت: ۱۴۰۱/۶/۷
تاریخ بازنگری: ۱۴۰۱/۷/۱۰
تاریخ پذیرش: ۱۴۰۱/۹/۷
تاریخ انتشار: ۱۴۰۱/۱۱/۱

واژه های کلیدی:

یادگیری ماشینی،
کربن آلی ذخیره خاک،
سنجش از دور،
متغیر محیطی.

استناد: خاموشی؛ سیدعرفان، سرمیدیان؛ فریدون، امید؛ محمود، (۱۴۰۱). تهیه نقشه رقومی کربن آلی ذخیره شده در خاک با استفاده از روش های یادگیری ماشینی، مجله

تحقیقات آب و خاک ایران، ۵۳ (۱۱)، ۲۶۸۱-۲۶۷۱. <https://doi.org/10.22059/ijswr.2022.346976.669339>

ناشر: مؤسسه انتشارات دانشگاه تهران.

© نویسندگان.

DOI: <https://doi.org/10.22059/ijswr.2022.346976.669339>

مقدمه

خاک‌ها بزرگترین منابع زمینی کربن در جهان و حاوی تقریباً ۱۵۰۰ پکاگرم کربن تا عمق یک متری هستند (Batjes, 2014). کربن ذخیره شده در خاک بیشتر از محتوای ذخیره شده در گیاهان و جو زمین است (Vicente-Vicente et al., 2016). از این رو ذخایر کربن آلی خاک یکی از عوامل اصلی تأثیرگذار بر خصوصیات فیزیکی و شیمیایی خاک مانند ظرفیت ذخیره آب خاک، ظرفیت ذخیره مواد مغذی، ظرفیت تبادل کاتیونی و میزان نفوذپذیری خاک است

(Minasny and McBratney, 2018; Scholten et al., 2017; Bangroo et al., 2020). استفاده از اراضی با شیوه‌های مدیریت ضعیف (Smith, 2012) و همچنین تغییر کاربری اراضی، دلایل اصلی کاهش کربن آلی خاک در سطح جهانی هستند. عامل دیگر در کاهش SOCS، جنگل زدایی و زیر کشت بردن این اراضی، به ویژه در مناطق گرمسیری و نیمه گرمسیری است (Canadell et al., 2007). از طرفی خاک‌های مناطق خشک و نیمه خشک در مقایسه با خاک‌های سایر مناطق اقلیمی خواص ویژه‌ای از خود نشان می‌دهند. این خاک‌ها در آب و هوای خشک با محتوای مواد آلی کم، در دسترس بودن کم آب، دماهای بالا و نوسانات دما تشکیل شده‌اند و مطالعات اخیر خاک‌های این مناطق را به عنوان عوامل موثر در تعیین مقدار و وضعیت کربن ذخیره شده در جهان شناسایی کرده‌اند (Ahlström et al., 2015; Kučera et al., 2020; Ma et al., 2016). بنابراین، اطلاعات دقیق در مورد توزیع مکانی SOCS برای تخمین و پیش‌بینی انتشار گازهای گلخانه‌ای و عملکردهای فیزیکی شیمیایی خاک حیاتی است (Adhikari et al., 2019; Minasny et al., 2013). همچنین، برآورد میزان تأثیر SOCS در کشاورزی و تولید محصول، ارزیابی اهمیت محرک‌های کاهش و افزایش آن و آگاهی از نحوه توزیع مکانی کربن ذخیره شده در خاک، برای پیش‌بینی اثرات احتمالی SOCS در آینده کره زمین در اثر تغییرات آب و هوایی و همچنین طراحی استراتژی‌های کاهش اثرات تغییر اقلیم و سازگاری محیط زیست با این تغییرات، حیاتی است (Funes et al., 2019).

نقشه‌های دقیق و بروز رسانی شده خصوصیات خاک مانند نقشه کربن آلی ذخیره شده در خاک جهت ارزیابی و نحوه مدیریت اراضی کشاورزی از جمله معیارهای مهم است. میزان دقت نقشه‌های به دست آمده می‌تواند به توسعه و برنامه ریزی سیاست‌های حفاظتی خاک و اراضی مناسب در سطوح استانی و منطقه‌ای و همچنین در بخش‌های تجاری و علمی کمک کند (Zizala et al., 2021). پیشرفت در روش‌ها و فن‌آوری‌های تجزیه و تحلیل رقومی، عصر مدرنی را در نقشه‌برداری خاک ایجاد کرده است که می‌تواند تقاضای فزاینده برای اطلاعات مکانی خاک را برآورده کند (Lagacherie, 2008). نقشه‌برداری رقومی خاک^۱ (DSM) بر اساس مفهوم SCORPAN (McBratney et al., 2003) استاندارد برای تولید نقشه‌های خصوصیات و کلاس‌های خاک در مقایسه با روش‌های قدیمی در نظر گرفته می‌شود. DSM با استفاده از مدل‌های یادگیری ماشین یک چارچوب تجربی-کمی برای پیش‌بینی ویژگی‌ها و کلاس‌های خاک با استفاده از متغیرهای محیطی ارائه می‌کند (Minasny et al., 2013; Taghizadeh-Mehrjardi et al., 2020; Žizala et al., 2021). بسیاری از مدل‌های یادگیری ماشین روابط قابل قبولی بین محتوای کربن آلی خاک و متغیرهای محیطی ایجاد می‌کنند (Grimm et al., 2008; Hounkpatin et al., 2019; Nabiollahi et al., 2021). ترکیب این عوامل محیطی متنوع شرایط منطقه‌ای منحصر به فردی ایجاد می‌کند که به ما در پیش‌بینی توزیع مکانی میزان ذخیره کربن آلی خاک کمک می‌کند (Gomes et al., 2019).

برخی از مهم‌ترین مدل‌ها عبارتند از: شبکه‌های عصبی مصنوعی کلاسیک، رگرسیون چند متغیره (Behrens and Scholten, 2006)، رگرسیون بردار پشتیبان (Were et al., 2015) ، درختان رگرسیون (Were et al., 2015; Rentschler et al., 2019) و جنگل تصادفی (Grimm et al., 2008; Nabiollahi et al., 2019). بسیاری از مطالعات گزارش کرده‌اند که مدل رگرسیون جنگل تصادفی که یک روش یادگیری ماشینی است، می‌تواند پتانسیل بالایی برای تعیین مناسب روابط غیرخطی بین خاک و متغیرهای کمکی محیطی داشته باشد (Gomes et al., 2019; Hengl et al., 2015; Taghizadeh-Mehrjardi et al., 2016; Wiesmeier et al., 2011).

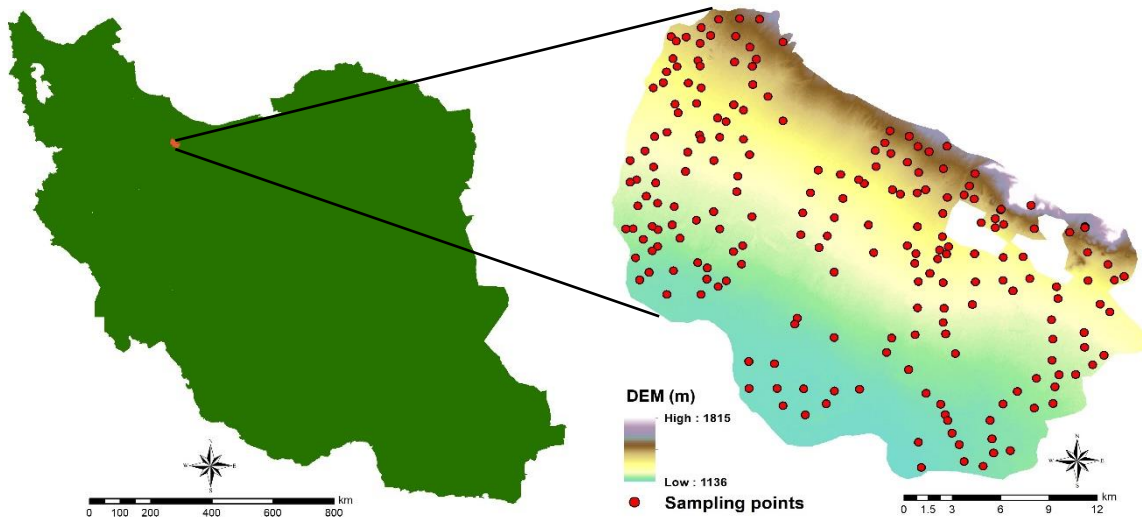
الگوریتم RF یک روش طبقه‌بندی و رگرسیون است که چندین الگوریتم درختی را برای ایجاد پیش‌بینی مکرر از هر مشخصه ترکیب می‌کند. این روش می‌تواند الگوهای پیچیده را یاد بگیرد و رابطه غیرخطی بین متغیرهای مستقل و وابسته را در نظر بگیرد. همچنین می‌تواند انواع مختلف داده‌ها را در تجزیه و تحلیل و ترکیب کند، که به دلیل انتخاب تصادفی داده‌ها در طول مدل‌سازی است (Breiman, 2001). (Meinshausen 2006) یک الگوریتم ارتقا یافته برای جنگل‌های تصادفی به نام جنگل‌های تصادفی کوانتایل معرفی کرد. این الگوریتم یک توزیع کاملاً شرطی برای متغیر پاسخ را استنباط می‌کند. از این اطلاعات می‌توان برای ایجاد چندک‌های پیش‌بینی و شناسایی

نقاط پرت در داده‌ها استفاده کرد. چندک‌های پیش‌بینی مشاهدات تازه را با درجه بالایی از دقت پوشش می‌دهد. در نتیجه، چندکی بودن پیش‌بینی‌ها، نوسان داده‌های را در اطراف مقادیر پیش‌بینی شده آنها نشان می‌دهد و در نهایت دقت پیش‌بینی مشاهدات جدید برای نمونه‌هایی در همان مجموعه داده‌ها مشخص می‌گردد که مقادیر پیش‌بینی شده چندک‌ها توسط QRF قابل مشاهده است. در نتیجه، توزیع شرطی برآورد شده، معمولاً، مکمل میانگین شرطی استنباط شده متغیر پاسخ است. نشان داده شد که QRF برای تخمین کمیت شرطی تحت برخی مفروضات قابل قبول سازگار است. همچنین، عملکرد این الگوریتم در مقایسه با رویکردهای خطی و درختی کاملاً رقابتی است. در تحقیقات مدل‌سازی و تخمین محتوای SOCS (Adhikari et al., 2019; Funes et al., 2019; Kučera et al., 2020; Ottoy et al., 2017) مقادیر SOCS برای هر نمونه خاک محاسبه می‌شود، سپس از مقادیر برای مدل‌سازی استفاده می‌گردد. در این تحقیق با استفاده از مدل‌های RF و QRF، داده‌های خاک و متغیرهای محیطی با وضوح ۱۰ متر، الف) مقایسه الگوریتم‌های RF و QRF در پیش‌بینی SOCS و ب) نحوه توزیع مکانی و مقادیر کربن آلی ذخیره خاک تا عمق ۱۰۰ سانتی متری مورد بررسی قرار گرفت.

مواد و روش‌ها

منطقه مورد مطالعه

منطقه مورد مطالعه بین ۳۶/۱۸ تا ۳۵/۹۰ درجه عرض شمالی و ۵۰/۳۴ تا ۵۰/۵۲ درجه طول شرقی در آبیک، استان قزوین، در شمال غربی ایران قرار دارد (شکل ۱). میانگین بارندگی و دما سالانه بر اساس داده‌های هواشناسی ۲۲ ساله (سازمان هواشناسی ایران (IRIMO) به ترتیب ۲۵۷ میلی متر و ۱۴/۳ درجه سانتی گراد است. بر این اساس، رژیم رطوبتی و دمایی خاک به ترتیب آریدیک ضعیف، زریک خشک، آبی و ترمیک بود. خاک‌های غالب عبارتند از Entisols، Inceptisols و Aridisols. این منطقه مساحتی به وسعت ۵۶۰ کیلومتر مربع با ارتفاع بین ۱۱۳۶ تا ۱۸۱۵ از سطح دریا را پوشش می‌دهد (شکل ۱). انواع کاربری اراضی منطقه مورد مطالعه شامل کشاورزی دیم، کشاورزی آبی و مرتع ضعیف است. بیش از ۸۰ درصد مساحت در کلاس شیب ۵-۰ درصد قرار دارد و جهت اصلی آن جنوبی با ۲۷/۱۵ درصد مساحت است.



شکل ۱. محدوده مورد مطالعه و نقاط نمونه‌برداری

داده‌ها و متغیرهای محیطی

در مجموع ۲۱۱ پروفایل موجود در بانک اطلاعاتی آزمایشگاه سنجش از دور گروه علوم و مهندسی خاک دانشگاه تهران انتخاب شدند. سپس مقادیر کربن آلی ذخیره شده در خاک تا عمق ۱۰۰ سانتی متری بر اساس معادله پیشنهادی به صورت زیر محاسبه شد (Martín et al., 2016):

$$SOCS = OC \times BD \times D \times (1 - G/100)$$

رابطه ۱)

در رابطه ۱: SOCS نشان دهنده کربن ذخیره آلی خاک (تن در هکتار)، OC درصد کربن آلی، BD نشان دهنده چگالی ظاهری (گرم بر سانتی متر مکعب)، D عمق خاک (سانتی متر) و G درصد حجمی سنگریزه خاک است.

در مدل‌سازی، از متغیرهای کمکی مربوط به معادله SCORPAN یعنی خاک، آب و هوا، موجودات زنده، توپوگرافی، مواد مادری و موقعیت مکانی استفاده شد (McBratney et al., 2003). بر این اساس، از مدل رقومی ارتفاع (DEM) با تفکیک مکانی ۱۲/۵ متر استفاده شد. در مجموع، ۱۱ متغیر ژئومورفومتریک به عنوان مشتقات اولیه و ثانویه DEM در نرم افزار SAGA GIS 7.3 تهیه شد. لایه‌های آماده شده در جدول ۱ گزارش شده است. علاوه بر این، ۲۵ شاخص طیفی از تصاویر ماهواره‌های لندست ۸ و سنتینل ۲ با استفاده از نرم افزار TerrSet 2020 تهیه شد که در جدول ۱ ارائه شده است. در نهایت، تمامی متغیرهای محیطی اندازه پیکسل‌های ۱۰ متری تبدیل شدند. مناسب‌ترین شاخص‌های محیطی برای مدل‌سازی بر اساس همبستگی پیرسون انتخاب شدند.

جدول ۱. متغیرهای محیطی استفاده شده

ردیف	نام	علامت اختصاری در این تحقیق
	متغیرهای مستخرج از مدل رقومی ارتفاع	
۱	مدل رقومی ارتفاع	dem
۲	سطح پایه شبکه کانال	Channel_Network_Bsae_Level (cnbl)
۳	شاخص توپوگرافی خیزی	Topographic_Wetness_Index (twi)
۴	شاخص عمق دره	Valley_Depth
۵	شاخص صافی پایین دره با وضوح چندگانه	MRVBF
۶	شاخص طول شیب	LS_Factor
۷	جهت شیب	Aspect
۸	شاخص همگرایی	Convergence_Index
۹	شاخص موقعیت نسبی شیب	RSP
۱۰	فاصله عمودی تا شبکه کانال	VDCN
۱۱	شیب	Slope
	متغیرهای مستخرج از تصاویر ماهواره	
۱۲	باند آبی لندست ۸ (0.482 μm)	b02l
۱۳	باند سبز لندست ۸ (0.561 μm)	b03l
۱۴	باند قرمز لندست ۸ (0.654 μm)	B04L
۱۵	باند مادون قرمز نزدیک لندست ۸ (0.864 μm)	B05L
۱۶	موج کوتاه ۱ مادون قرمز لندست ۸ (1.608 μm)	B06L
۱۷	موج کوتاه ۲ مادون قرمز لندست ۸ (2.200 μm)	B07L
۱۸	باند آبی سنتینل ۲ (0.490 μm)	b2s
۱۹	باند سبز سنتینل ۲ (0.560 μm)	b3s
۲۰	باند قرمز سنتینل ۲ (0.665 μm)	b4s
۲۱	مرز قرمز گیاهی سنتینل ۲ (0.705 μm)	b5s
۲۲	مرز قرمز گیاهی سنتینل ۲ (0.740 μm)	b6s
۲۳	مرز قرمز گیاهی سنتینل ۲ (0.783 μm)	b7s
۲۴	باند مادون قرمز نزدیک سنتینل ۲ (0.842 μm)	b8s
۲۵	مرز قرمز گیاهی سنتینل ۲ (0.865 μm)	b8as
۲۶	موج کوتاه ۱ مادون قرمز سنتینل ۲ (1.610 μm)	b11s
۲۷	موج کوتاه ۲ مادون قرمز سنتینل ۲ (2.190 μm)	b12s
۲۸	شاخص پوشش گیاهی نرمال شده (لندست ۸)	ndvi.ls8
۲۹	شاخص پوشش گیاهی نرمال شده (سنتینل ۲)	ndvi_sen
۳۰	شاخص شوری نرمال شده (لندست ۸)	ndsi
۳۱	شاخص شوری نرمال شده (سنتینل ۲)	ndsi_sen
۳۲	شاخص شوری (لندست ۸)	si
۳۳	شاخص شوری (سنتینل ۲)	si_sen
۳۴	Tasseled Cab brightness (لندست ۸)	tasseled_cab_b
۳۵	Tasseled Cab greenness (لندست ۸)	tasseled_cab_g
۳۶	Tasseled Cab wetness (لندست ۸)	tasseled_cab_w

مدل سازی

الگوریتم جنگل تصادفی یک رویکرد ناپارامتریک است که ارتباط نسبی هر متغیر پیش‌بینی کننده را در متغیر پاسخ شاهد، رتبه‌بندی می‌کند. همچنین این الگوریتم می‌تواند روابط غیرخطی و افزایشی را مدیریت کند. (Wang et al., 2018). الگوریتم رندم فارست کوانتایل با تعمیم مدل کلاسیک RF به طوری توسعه یافت که مقادیر پیش‌بینی شده همه درختان را برای هر نقطه پیش‌بینی میانگین دهی کند (Biau and Scornet, 2016; Breiman, 2001). علاوه بر این مدل QRF می‌تواند شرط برقرار شده در هریک از متغیرهای مدل شده را در دیگری تخمین بزند (Meinshausen, 2006). الگوریتم‌های رندم فارست فقط نیاز به تنظیم دو پارامتر برای تولید یک مدل پیش‌بینی دارند: (۱) تعداد درختان رگرسیون در حال رشد در جنگل‌ها (ntree) و (۲) تعداد متغیرهای محیطی به طور تصادفی در هر گره (mtry). به طور پیش فرض، اندازه زیرمجموعه تصادفی جذر تعداد کل پیش‌بینی کننده‌ها در مدل است. اهمیت متغیرها بر اساس خطاهای پیش‌بینی رگرسیون خارج از کیسه (OOB) است که در نمونه آموزشی گنجانده نشده است (Wang et al., 2018). در این تحقیق از اعتبار سنجی متقاطع K-fold برای یافتن بهترین تعداد mtry استفاده شد. الگوریتم QRF و RF در نرم افزار R نسخه ۴٫۱٫۰ اجرا شدند. ۷۰ درصد از نمونه‌های خاک (n=۱۴۸) به عنوان داده آموزشی و ۳۰ درصد به عنوان داده اعتبارسنجی مدل (n=۶۳) از طریق نمونه‌گیری تصادفی انتخاب شدند.

اعتبار سنجی

عملکرد مدل‌ها با استفاده از شاخص‌های آماری شامل ریشه میانگین مربعات خطا (RMSE)، ضریب تعیین (R^2) از طریق پکیج "caret" نرم افزار R نسخه ۴٫۱٫۰ به دست آمد. (Kuhn, 2008). RMSE (معادله ۲) دقت برآورد را نشان می‌دهد. اگر نزدیک به صفر باشد، تخمین قابل اعتمادتر است:

$$RMSE = \sqrt{\frac{1}{N} \sum (K_p - K_m)^2} \quad \text{رابطه ۲}$$

همچنین ضریب تعیین (R^2) با استفاده از فرمول زیر محاسبه شد:

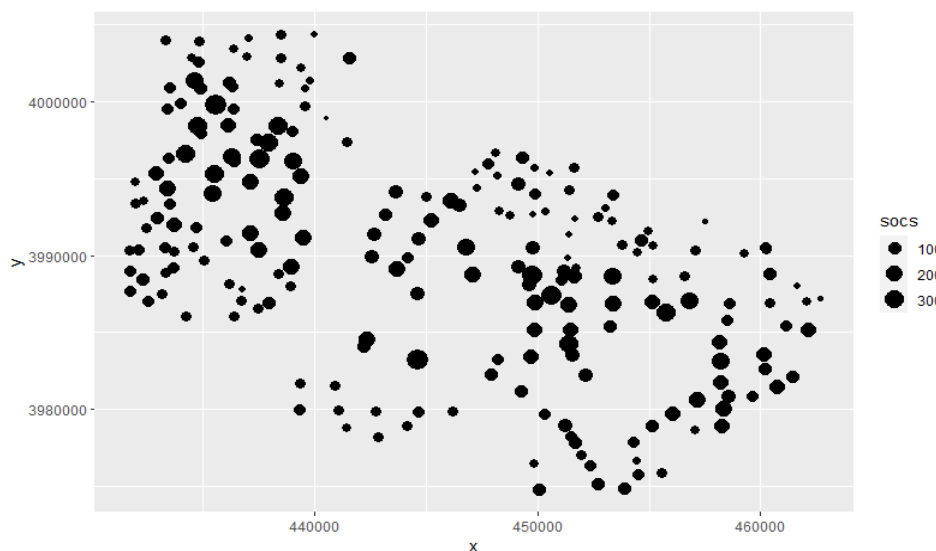
$$R^2 = 1 - \frac{\sum (K_m - K_p)^2}{\sum (K_m - \bar{K}_m)^2} \quad \text{رابطه ۳}$$

در معادلات فوق K_p مقادیر پیش‌بینی شده توسط مدل و K_m مقادیر مشاهده شده و N تعداد کل مشاهدات است.

نتایج و بحث

آنالیزهای آماری

نمودار پراکنش مکانی مقادیر کربن آلی ذخیره شده خاک در شکل ۲ نشان داده شده است. کمترین و بیشترین محتوای کربن آلی ذخیره خاک، ۰ و ۳۹۵ تن در هکتار با میانگین ۱۰۷ تن در هکتار بود (جدول ۲).



شکل ۲. پراکنش مکانی مقادیر اندازه‌گیری شده کربن آلی ذخیره شده در خاک

جدول ۲. آمار توصیفی مقادیر کربن آلی ذخیره شده در خاک تا عمق ۱۰۰ سانتی متری

کمترین	بیشترین	میانگین	ضریب تغییرات (%)	انحراف از استاندارد
۰	۳۹۵	۱۰۷	۷۶/۸۱	۸۱/۹۳
کربن آلی ذخیره شده (تن در هکتار)				

نتایج ضرایب همبستگی پیرسون بین کربن ذخیره خاک و متغیرهای کمکی محیطی در جدول ۳ نشان داده شده است. براین اساس مقدار کربن ذخیره آلی با متغیرهای محیطی dem و cnbl، b021، B07L، b2s، b3s، b4s، b11s، b12s، ndsi_sen و si_sen همبستگی منفی معنی داری داشت. در حالی که با شاخص‌های Valley depth، tasseled_cab_g و ndvi_sen همبستگی مثبت و معناداری نشان داد. به طور کلی، شاخص پوشش گیاهی یکی از عوامل موثر در تخمین میزان کربن آلی خاک است (Devine et al., 2020)، به طوری که هرچه شاخص پوشش گیاهی مقادیر بالاتر و نزدیک به ۱ را نشان دهد بیانگر تراکم بالای پوشش گیاهی و به طبع آن فعالیت بیشتر ریشه و ریز موجودات است. بنابراین انتظار می‌رود در رابطه با کربن آلی روندی مثبت داشته باشد (جدول ۳). در نهایت از متغیرهای محیطی که همبستگی معنی داری با SOCS داشتند در مدل سازی استفاده شد.

جدول ۳. نتایج تحلیل همبستگی پیرسون بین کربن ذخیره خاک تا عمق ۱۰۰ سانتی متری و متغیرهای محیطی

متغیر	SOCS
dem	-۰/۲۸*
cnbl	-۰/۲۷*
MRVBF	۰/۰۵
Valley Depth	۰/۲۱*
twi	-۰/۱۴
LS_Factor	-۰/۱۹
b021	-۰/۱۶*
b031	-۰/۱۳
B04L	-۰/۱۱
B05L	۰/۱۹
B06L	-۰/۰۳
B07L	-۰/۱۹**
si	-۰/۱۴
ndsi	-۰/۲۹
ndvi.ls8	-۰/۲۹*
tasseled_cab_b	-۰/۰۳
tasseled_cab_g	۰/۳۲**
b2s	-۰/۳۲**
b3s	-۰/۳۱**
b4s	-۰/۳۲**
b5s	۰/۲۸
b6s	۰/۰۹
b7s	۰/۲۳
b8s	۰/۲۲
b8as	۰/۲۴
b11s	-۰/۳*
b12s	-۰/۳۵**
ndsi_sen	-۰/۳۶**
ndvi_sen	۰/۳۶**
si_sen	-۰/۳۲**

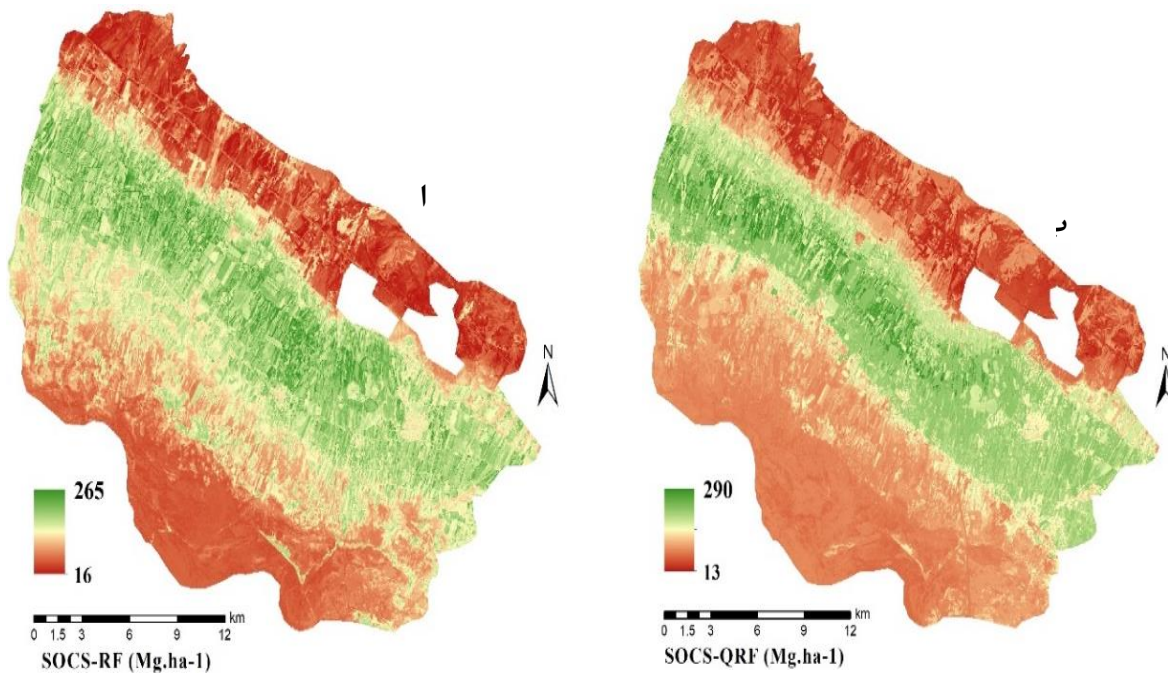
**همبستگی معنی دار در سطح ۰/۰۱ درصد و * همبستگی معنی دار در سطح ۰/۰۵ درصد

نتایج مدل سازی

نقشه رقومی حاصل از استفاده الگوریتم‌های یادگیری ماشین در شکل ۳ آمده است. همچنین عملکرد مدل‌های RF و QRF برای پیش‌بینی مقادیر کربن آلی ذخیره در منطقه مورد مطالعه با استفاده از R^2 ، RMSE و همبستگی استخراج گردید. با توجه به نتایج اعتبار سنجی ارائه شده در جدول ۴، الگوریتم QRF دقت بالاتری را با R^2 بالاتر و RMSE کمتر ارائه می‌دهد. این نتایج نشان داد که عموماً روش‌های رگرسیون درختی در توابع پیش‌بینی خود، مقادیر نهایی را بر اساس رویکرد محاسباتی خود به‌عنوان میانگین‌هایی از مقادیر در حد بالا و پایین ارائه می‌کنند. علاوه بر این، با توجه به اهمیت نسبی داده شده متغیرهای محیطی، نتایج مدل QRF نشان می‌دهد که (شکل ۴) مدل رقومی ارتفاع مهم‌ترین متغیری است که بر توزیع مکانی SOCS

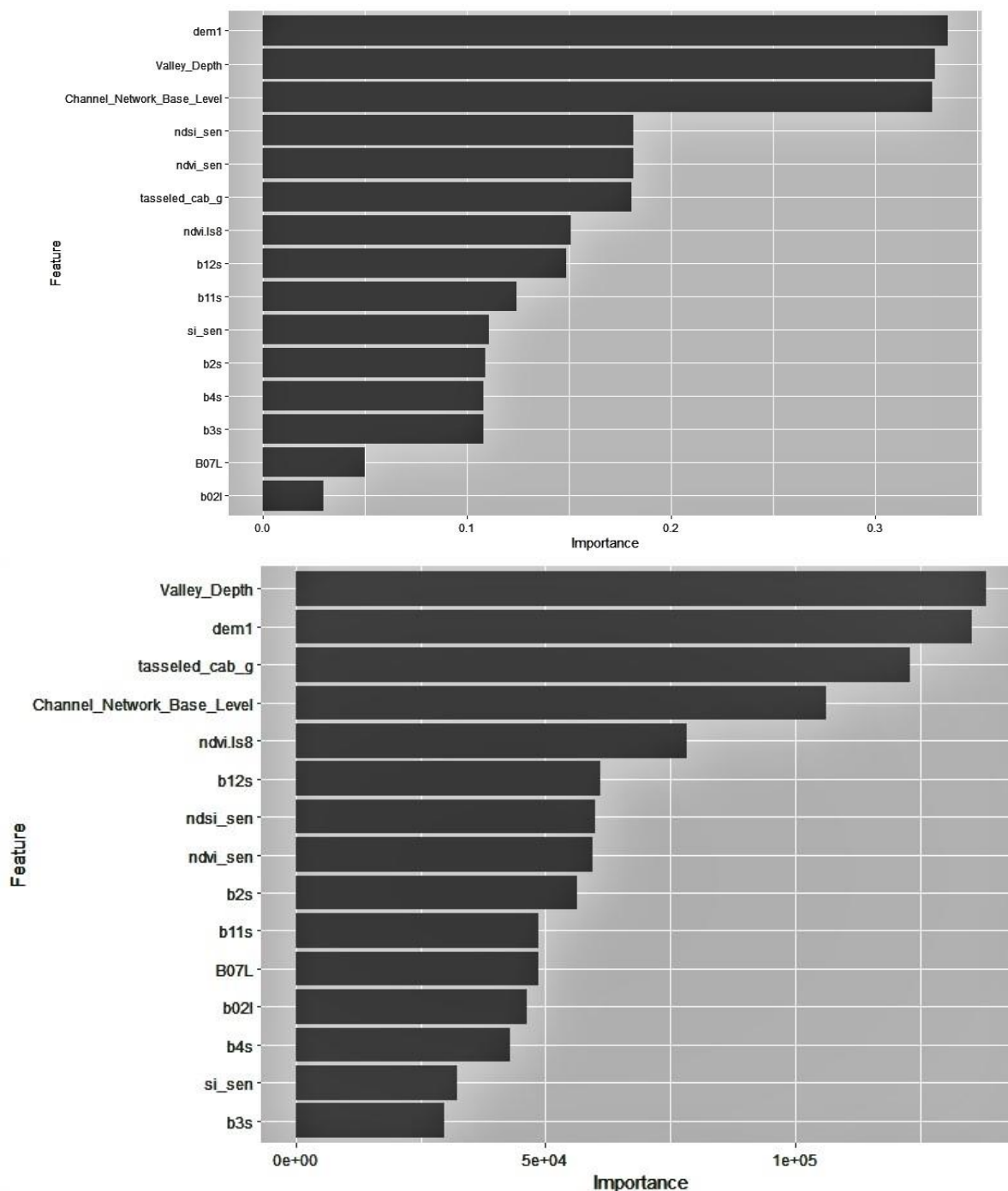
جدول ۴. نتایج اعتبار سنجی پیش‌بینی مقادیر کربن آلی ذخیره شده در خاک

RMSE		Correlation		R^2		SOCS
QRF	RF	QRF	RF	QRF	RF	
۵۲/۲	۶۱/۶	۰/۶۹	۰/۶۶	۰/۴۷	۰/۴۳	



شکل ۳. نقشه رقومی نحوه توزیع کربن آلی ذخیره شده در خاک: الف) الگوریتم رندم فارست، ب) الگوریتم رندم فارست کوانتایل

تأثیر می‌گذارد، در حالی که اهمیت مدل رقومی ارتفاع در الگوریتم RF در رتبه سوم قرار گرفت. این نتایج با نتایج Dharumarajan et al. (2020) مطابقت داشت. بر اساس نتایج مدل RF، در بین متغیرهای مبتنی بر RS، متغیرهای استخراج شده از ماهواره لندست ۸ (Tasseled_cab_g و ndvi.ls8) نسبت به متغیرهای مشتق شده از ماهواره سنتینل ۲ (ndvi_sen و ndsi_sen) اهمیت بیشتری داشتند. این در حالی بود که نتایج مدل QRF خلاف این را نشان داد و متغیرهای مستخرج از ماهواره سنتینل ۲ دارای اهمیت بیشتری در پیش‌بینی مکانی مقادیر کربن آلی ذخیره شده در خاک دارا بودند. اما به طور کلی می‌توان بیان کرد که در مدل سازی مکانی SOCS، عوامل مرتبط با پوشش گیاهی و نوع آن (DEM و NDVI) بیشترین تأثیر را دارند (Gomez et al., 2019). بنابراین در منطقه مورد مطالعه متغیرهای ژئومورفومتریک کنترل کننده تغییرات SOCS و پارامترهای مدیریتی مانند شاخص پوشش گیاهی مهم‌ترین عوامل در کنترل تغییرپذیری مکانی SOCS هستند.



شکل ۴. نمودار اهمیت نسبی متغیرهای مورد استفاده در مدل سازی. الف) رندم فارست، ب) رندم فارست کوانتایل

نتیجه گیری

بررسی وضعیت SOCS یکی از عوامل مهم در مدیریت پایدار اراضی به شمار می آید. در این تحقیق از یک مجموعه داده خاص بر اساس دو الگوریتم از دو نسل مدل های یادگیری ماشین برای محاسبه و پیش بینی مکانی و نحوه توزیع مکانی مقدار کربن ذخیره شده خاک استفاده شد. که نتایج آن به طور کلی به شرح زیر می باشد:

۱. الگوریتم QRF دقت و صحت بالاتری نسبت به مدل RF در تخمین مقادیر و توزیع مکانی SOCS داشت.
۲. مهم ترین متغیرهای کمکی که بر کربن آلی ذخیره شده در خاک تأثیر می گذارد مربوط به نوع پوشش گیاهی است.
۳. در منطقه مورد مطالعه تقریباً ۱۵ میلیون تن کربن آلی در ۱۰۰ سانتی متر بالا ذخیره شده است که نحوه مدیریت و افزایش آن می بایست یکی از مهمترین برنامه های مدیریت اراضی در منطقه باشد.



همچنین با توجه به نتایج، علی رقم اینکه SOCS یکی از ویژگی‌های خاک است که از ترکیب چندین ویژگی دیگر خاک (داده‌های ثانویه خاک) حاصل می‌شود، متغیرهای کمکی محیطی به نحو مطلوبی ارتباط معنی داری را با مقدار کربن ذخیره در خاک نشان دادند که در منطقه مورد مطالعه متغیرهای مستخرج از تصاویر ماهواره‌ای نسبت به متغیرهای مستخرج از مدل رقومی ارتفاع ارتباط معنی داری با متغیر هدف (SOCS) برقرار کردند. البته می‌توان گفت با توجه به اینکه خاک محیطی پویا و تحت تاثیر شرایط محیطی است ممکن است این ارتباط در مکان جغرافیایی متفاوت تغییر کند.

"هیچ‌گونه تعارض منافع بین نویسندگان وجود ندارد"

REFERENCES

- Adhikari, K., Owens, P. R., Libohova, Z., Miller, D. M., Wills, S. A., & Nemecek, J. (2019). Assessing soil organic carbon stock of Wisconsin, USA and its fate under future land use and climate change. *Science of the Total Environment*, 667, 833-845.
- Ahlström, A., Raupach, M. R., Schurgers, G., Smith, B., Arneeth, A., Jung, M., ... & Zeng, N. (2015). The dominant role of semi-arid ecosystems in the trend and variability of the land CO₂ sink. *Science*, 348(6237), 895-899.
- Bangroo, S. A., Najar, G. R., Achin, E., & Truong, P. N. (2020). Application of predictor variables in spatial quantification of soil organic carbon and total nitrogen using regression kriging in the North Kashmir forest Himalayas. *Catena*, 193, 104632.
- Batjes, N. H. (2014). Total carbon and nitrogen in the soils of the world. *European Journal of Soil Science*, 65(1), 10-21.
- Behrens, T., & Scholten, T. (2006). Digital soil mapping in Germany—a review. *Journal of Plant Nutrition and Soil Science*, 169(3), 434-443.
- Biau, G., Scornet, E., (2016). A random forest guided tour. *Test* 25, 197–227.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32 .
- Canadell, J. G., Pataki, D. E., Gifford, R., Houghton, R. A., Luo, Y., Raupach, M. R., ... & Steffen, W. (2007). Saturation of the terrestrial carbon sink. In *Terrestrial ecosystems in a changing world* (pp. 59-78). Springer, Berlin, Heidelberg.
- Dharumarajan, S., Kalaiselvi, B., Suputhra, A., Lalitha, M., Hegde, R., Singh, S. K., & Lagacherie, P. (2020). Digital soil mapping of key GlobalSoilMap properties in Northern Karnataka Plateau. *Geoderma Regional*, 20, e00250.
- Funes, I., Savé, R., Rovira, P., Molowny-Horas, R., Alcañiz, J. M., Ascaso, E., ... & Vayreda, J. (2019). Agricultural soil organic carbon stocks in the north-eastern Iberian Peninsula: Drivers and spatial variability. *Science of the Total Environment*, 668, 283-294.
- Gomes, L. C., Faria, R. M., de Souza, E., Veloso, G. V., Schaefer, C. E. G., & Fernandes Filho, E. I. (2019). Modelling and mapping soil organic carbon stocks in Brazil. *Geoderma*, 340, 337-350.
- Grimm, R., Behrens, T., Märker, M., & Elsenbeer, H. (2008). Soil organic carbon concentrations and stocks on Barro Colorado Island—Digital soil mapping using Random Forests analysis. *Geoderma*, 146(1-2), 102-113.
- Hengl, T., Heuvelink, G. B., Kempen, B., Leenaars, J. G., Walsh, M. G., Shepherd, K. D., ... & Tondoh, J. E. (2015). Mapping soil properties of Africa at 250 m resolution: Random forests significantly improve current predictions. *PloS one*, 10(6), e0125814.
- Houkpatin, K. O., Stendahl, J., Lundblad, M., & Karlun, E. (2021). Predicting the spatial distribution of soil organic carbon stock in Swedish forests using a group of covariates and site-specific data. *Soil*, 7(2), 377-398.
- Kučera, A., Skene, K. R., & Kupec, P. (2020). Soil hydric properties and carbon stock in a semi-arid region of Iraqi Kurdistan: The importance of historical pedogenesis, climate and locality. *Ecological Indicators*, 119, 106813.
- Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, 28, 1-26.
- Lagacherie, P. (2008). *Digital soil mapping with limited data*, Springer: Dordrecht.
- Martín, J. R., Álvaro-Fuentes, J., Gonzalo, J., Gil, C., Ramos-Miras, J. J., Corbí, J. G., & Boluda, R. (2016).

- Assessment of the soil organic carbon stock in Spain. *Geoderma*, 264, 117-125.
- Ma, X., Huete, A., Cleverly, J., Eamus, D., Chevallier, F., Joiner, J., ... & Ponce-Campos, G. (2016). Drought rapidly diminishes the large net CO₂ uptake in 2011 over semi-arid Australia. *Scientific Reports*, 6(1), 1-9.
- McBratney, A. B., Santos, M. M., & Minasny, B. (2003). On digital soil mapping. *Geoderma*, 117(1-2), 3-52.
- Meinshausen, N., & Ridgeway, G. (2006). Quantile regression forests. *Journal of Machine Learning Research*, 7 (6).
- Minasny, B., & McBratney, A. B. (2018). Limited effect of organic matter on soil available water capacity. *European Journal of Soil Science*, 69(1), 39-47.
- Minasny, B., McBratney, A. B., Malone, B. P., & Wheeler, I. (2013). Digital mapping of soil carbon. *Advances in agronomy*, 118, 1-47.
- Nabiollahi, K., Eskandari, S., Taghizadeh-Mehrjardi, R., Kerry, R., & Triantafylis, J. (2019). Assessing soil organic carbon stocks under land-use change scenarios using random forest models. *Carbon Management*, 10(1), 63-77.
- Ottoy, S., De Vos, B., Sindayihebura, A., Hermy, M., & Van Orshoven, J. (2017). Assessing soil organic carbon stocks under current and potential forest cover using digital soil mapping and spatial generalisation. *Ecological indicators*, 77, 139-150.
- Rentschler, T., Gries, P., Behrens, T., Bruelheide, H., Kühn, P., Seitz, S., ... & Schmidt, K. (2019). Comparison of catchment scale 3D and 2.5 D modelling of soil organic carbon stocks in Jiangxi Province, PR China. *Plos one*, 14(8), e0220881.
- Scholten, T., Goebes, P., Kühn, P., Seitz, S., Assmann, T., Bauhus, J., ... & Schmidt, K. (2017). On the combined effect of soil fertility and topography on tree growth in subtropical forest ecosystems—a study from SE China. *Journal of Plant Ecology*, 10(1), 111-127.
- Smith, P. (2012). Agricultural greenhouse gas mitigation potential globally, in Europe and in the UK: what have we learnt in the last 20 years?. *Global Change Biology*, 18(1), 35-43.
- Taghizadeh-Mehrjardi, R., Nabiollahi, K., & Kerry, R. (2016). Digital mapping of soil organic carbon at multiple depths using different data mining techniques in Baneh region, Iran. *Geoderma*, 266, 98-110.
- Taghizadeh-Mehrjardi, R., Schmidt, K., Amirian-Chakan, A., Rentschler, T., Zeraatpisheh, M., Sarmadian, F., ... & Scholten, T. (2020). Improving the spatial prediction of soil organic carbon content in two contrasting climatic regions by stacking machine learning models and rescanning covariate space. *Remote Sensing*, 12(7), 1095.
- Vicente-Vicente, J. L., García-Ruiz, R., Francaviglia, R., Aguilera, E., & Smith, P. (2016). Soil carbon sequestration rates under Mediterranean woody crops using recommended management practices: A meta-analysis. *Agriculture, Ecosystems & Environment*, 235, 204-214.
- Wang, B., Waters, C., Orgill, S., Cowie, A., Clark, A., Li Liu, D., ... & Sides, T. (2018). Estimating soil organic carbon stocks using different modelling techniques in the semi-arid rangelands of eastern Australia. *Ecological Indicators*, 88, 425-438.
- Were, K., Bui, D. T., Dick, Ø. B., & Singh, B. R. (2015). A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afrotropical landscape. *Ecological Indicators*, 52, 394-403.
- Wiesmeier, M., Barthold, F., Blank, B., & Kögel-Knabner, I. (2011). Digital mapping of soil organic matter stocks using Random Forest modeling in a semi-arid steppe ecosystem. *Plant and soil*, 340(1), 7-24.
- Žížala, D., Minařík, R., Skála, J., Beitlerová, H., Juřicová, A., Rojas, J. R., ... & Zádorová, T. (2022). High-resolution agriculture soil property maps from digital soil mapping methods, Czech Republic. *Catena*, 212, 106024.