



پیش‌بینی مکانی برخی از ویژگی‌های خاک سطحی با استفاده از مدل‌های درون‌یابی و یادگیری ماشین

سیدروح‌اله موسوی^۱، فرزانه پارسائی^۲، اصغر رحمانی^۱، *محمد حسین سدري^۳ و محمد کوهسار بستانی^۴

^۱ دانشجوی دکتری مدیریت منابع خاک، گروه علوم و مهندسی خاک، دانشگاه تهران، دانشجوی دکتری مدیریت منابع خاک، گروه علوم و مهندسی خاک، دانشگاه شهید چمران اهواز، ^۲ استادیار بخش تحقیقات خاک و آب، مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی استان کردستان، سازمان تحقیقات، آموزش و ترویج کشاورزی، سنندج، ایران، ^۳ محقق بخش تحقیقات خاک و آب، مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی استان کردستان، سازمان تحقیقات، آموزش و ترویج کشاورزی، سنندج، ایران

تاریخ دریافت: ۱۳۹۸/۰۷/۲۹؛ تاریخ پذیرش: ۱۳۹۹/۰۴/۲۲

چکیده

سابقه و هدف: دست‌یابی به اطلاعات مکانی دقیق و با جزئیات بیشتر از پراکنش مکانی ویژگی‌های خاک بر روی سیمای اراضی برای پیش‌دقیق منابع اراضی، کاربردهای هیدرولوژیکی، مدیریت کاربری اراضی و سایر مدل‌سازی‌های محیطی ضروری است و نقش بنیادینی را در فرآیندهای اکو هیدرولوژی، اراضی مستعد کشاورزی و مدیریت پایدار اراضی در مناطق نیمه‌خشک ایفا می‌نماید. پژوهش حاضر با هدف پیش‌بینی مکانی درصد کربن آلی خاک (SOC)، کربنات کلسیم معادل (CCE)، رس، سیلت و ماسه در منطقه قروه-دهگلان کردستان صورت پذیرفت.

مواد و روش‌ها: منطقه قروه دهگلان با میانگین دمای سالانه ۱۲ درجه سلسیوس و میانگین بارش سالانه ۳۴۸ میلی‌متر (دوره آماری ۲۰ ساله) به‌ترتیب دارای رژیم‌های رطوبتی و حرارتی زیریک و ترمیک است. حدود ۱۴۵ نمونه خاک از عمق ۰-۳۰ سانتی‌متر بر اساس الگوی نمونه‌برداری تصادفی برداشت گردید. سپس نمونه‌های خاک برای انجام آزمایش‌های فیزیکی و شیمیایی لازم به آزمایشگاه منتقل گردید. مدل جنگل تصادفی (RF) به‌عنوان نماینده روش‌های غیرپارامتریک و دو روش کریگینگ معمولی (OK) و وزن‌دهی عکس فاصله (IDW) برای مدل‌سازی تغییرات مکانی ویژگی‌های خاک و خودهمبستگی مکانی بین آن‌ها استفاده گردیدند. تمامی مراحل مدل‌سازی روش RF در نرم‌افزار RStudio و روش‌های درون‌یابی (IDW و OK) در نرم‌افزارهای ArcGIS و GS⁺ صورت پذیرفت. ۳۰ متغیر محیطی شامل مشتقات مدل رقومی ارتفاع (DEM) در نرم‌افزار SAGA GIS 7.3 و داده‌های باند انعکاسی ماهواره لندست ۸ به‌عنوان متغیرهای محیطی تهیه شدند. تمامی متغیرهای محیطی مورد استفاده با تفکیک مکانی ۳۰ متر بازنمونه‌گیری گردیدند. انتخاب متغیرهای محیطی بهینه طبق شاخص تورم واریانس (VIF) انجام گردید. مدل‌سازی ویژگی‌ها طبق دو دسته داده ۸۰ و ۲۰ درصد، به‌ترتیب برای واسنجی و اعتبارسنجی انجام گردید و از دو آماره میانگین ریشه مربعات خطا (RMSE) و ضریب تبیین (R^2) برای تعیین دقت مدل‌ها استفاده شد.

یافته‌ها: بر اساس شاخص تورم واریانس از مجموع ۳۰ متغیر محیطی تهیه‌شده در نهایت هفت متغیر کمکی شامل، چهار متغیر سنجش‌ازدور شاخص پوشش گیاهی تعدیل‌یافته خاک (SAVI)، سبزی‌نگی شاخص گیاهی تفاضلی

* مسئول مکاتبه: sedri_mh@yahoo.com

نرمال شده (GNDVI)، شاخص پوشش گیاهی نسبی (RVI) و شاخص پوشش گیاهی بهبودیافته (EVI) و سه ویژگی ژئومورفومتری شامل مدل رقومی ارتفاع (DEM)، فاصله اقلیدسی از شبکه آبراهه و شاخص موقعیت توپوگرافی (TPI) انتخاب گردیدند. نتایج مدل‌سازی نشان داد که مدل RF برای متغیر کربن آلی خاک ($R^2=0/5$) و $R^2=0/4$ ، متغیر کربنات کلسیم معادل ($R^2=0/4$ و $RMSE=11/61$ ٪)، متغیر رس ($R^2=0/21$) و $R^2=0/65$ و متغیر سیلت ($R^2=0/15$ و $RMSE=7/24$ ٪)، مناسب‌ترین مدل و روش کریگینگ معمولی برای متغیر ماسه با ($R^2=0/14$ و $RMSE=10/26$) نسبت به دو مدل RF و IDW دارای دقت بالاتری بود. از میان نیم‌تغییرنماهای برازش داده‌شده مدل نمایی برای متغیرهای کربن آلی خاک، رس، سیلت و ماسه به‌جز کربنات کلسیم معادل دارای مناسب‌ترین برازش بود. نتایج خودهمبستگی مکانی نشان داد که دو متغیر کربنات کلسیم معادل و ماسه دارای کلاس همبستگی مکانی قوی بود و مابقی دارای کلاس متوسط می‌باشند. بالاترین مقادیر سقف نیم‌تغییرنما مربوط به کربنات کلسیم معادل و رس و کم‌ترین مقدار آن مربوط به کربن آلی خاک و ماسه بود. این نتایج بیانگر وجود الگوی تصادفی یا ساختار مکانی ضعیف در نمونه‌های مورد استفاده برای محاسبه نیم‌تغییرنما است. از میان متغیرهای کمکی مورد استفاده برای مدل‌سازی مکانی سه ویژگی کربن آلی خاک، رس و کربنات کلسیم معادل متغیرهای ژئومورفومتری مدل رقومی ارتفاع شاخص موقعیت توپوگرافی دارای بیش‌ترین اهمیت و برای متغیرهای ماسه و سیلت پارامترهای NDVI، SAVI و RVI دارای اهمیت بیش‌تری بودند.

نتیجه‌گیری: به‌طور کلی ویژگی‌های سطحی مورد بررسی دارای کلاس خودهمبستگی مکانی متوسط تا قوی بودند، اما نتایج مدل‌سازی دارای دقت بالایی نبودند؛ بنابراین برای مطالعات آتی پیشنهاد می‌گردد که از سایر روش‌های نمونه‌برداری مانند ابر مکعب لاتین یا طبقه‌بندی تصادفی و نقشه‌های موضوعی مثل زمین‌شناسی، ژئومورفولوژی و نقشه خاک موجود منطقه به‌عنوان ورودی‌های مدل‌سازی مکانی به‌منظور ارتقاء نتایج مدل‌سازی استفاده گردد.

واژه‌های کلیدی: تغییرپذیری مکانی، جنگل تصادفی، داده‌کاوی، نقشه‌برداری رقومی خاک

مقدمه

رقومی خاک برای پیش‌بینی الگوی مکانی تغییرات خاک‌ها در سراسر زمین‌نما، میان اطلاعات مربوط به ویژگی‌ها یا کلاس‌های خاک با داده‌های محیطی بر اساس یکسری روش‌های ریاضیاتی و آماری ارتباط برقرار می‌کند. روش‌های زیادی در نقشه‌برداری رقومی خاک به کار می‌رود که عمدتاً در گروه روش‌های یادگیری ماشین^۱ قرار دارند (۵۶). مکاران (۲۰۰۳) با مروری بر مدل‌های مختلف تخمین و برآورد مکانی خاک اقدام به ارائه یک چارچوب تعمیم‌یافته پیش‌بینی مکانی نمودند؛

با توجه به جایگاه خاک در طبیعت، آگاهی از پراکنش مکانی و مدیریت درست خاک‌ها برای پایداری تولید محصولات ضروری است. در این راستا، دسترسی به نقشه‌هایی با قدرت تفکیک مکانی بالا برای برنامه‌ریزی کاربری زمین و دیگر فعالیت‌های اصلی هم‌چون جنگلداری، کشاورزی و حفاظت محیط‌زیست اهمیت بنیادین دارند. یک رویکرد مفید برای نقشه‌برداری خاک‌ها در سطح وسیعی از اراضی که با روش‌های معمول عمدتاً فعالیتی زمان‌بر است، استفاده از نقشه‌برداری رقومی خاک است. نقشه‌برداری

به دلیل این که نسبت به سایر روش های درون یابی به طور گسترده توسط نرم افزارهای مختلف قابل اجرا می باشند و از طرفی نیاز کم تر کاربران به دانش محاسباتی در زمینه اجرا نمودن این روش ها به دلیل انجام بخش غالب پردازش های مورد نیاز به صورت خودکار، مانند بخش تحلیل زمین آماری^۴ در محیط نرم افزار ArcGIS به عنوان یکی از اصلی ترین مزیت های استفاده از این روش ها محسوب می شود (۱۱). یکی از فرضیات روش کریگینگ و بسیاری از روش های زمین آماری تبعیت داده ها از توزیع نرمال است و در صورتی که این فرض محقق نشود باید عملیات تبدیل^۵ و باز تبدیل^۶ صورت پذیرد. یکی دیگر از فرضیات کریگینگ وجود حداقل ۱۰۰ نقطه مشاهداتی برای برازش مدل نیم تغییرنا می باشد (۵۳). روش جنگل تصادفی با ایجاد الگوهای غیرخطی بین داده ها می تواند روابط پیچیده بین متغیرهای مستقل و وابسته را مدل سازی نماید و موجب افزایش دقت پیش بینی ویژگی های خاک در مناطق نمونه برداری نشده می گردد. البته از طرفی ممکن است نتایج روش جنگل تصادفی دست خوش بیش برازش گردد. متغیرهای کمکی محیطی با توزیع تصادفی خود موجب ایجاد نویز شده که در نهایت منجر به کاهش دقت مدل سازی می گردند (۴۹). ما و همکاران (۲۰۱۷) در شرق کشور چین با استفاده از ۲۸۹۵ نمونه خاک سطحی و متغیرهای محیطی کمکی از روش های رگرسیون درختی و رگرسیون کریگینگ برای مدل سازی کربن آلی خاک استفاده نمودند نتایج مطالعه ایشان نشان داد که روش رگرسیون کریگینگ دقت بیشتری را نسبت به سایر روش های مورد استفاده داشت. (۳۳). در مطالعات پیشین همواره روش های یادگیری ماشین به خصوص روش جنگل تصادفی به دلیل عملکرد غیرخطی و مدل

چارچوبی که بعدها تحت عنوان نقشه برداری رقومی خاک نامیده شد (۳۷). نقشه برداری رقومی خاک به عنوان یکی از زیرشاخه های موفق علوم خاک مطرح شده است، به طوری که تعداد مقالات و ارجاعات مربوط به آن در طی سال های اخیر به سرعت رو به افزایش است (۳۹). یکی از اجزاء اصلی نقشه برداری رقومی خاک مدل ارتباط دهنده بین متغیرهای محیطی و خاک است که در این راستا می توان از انواع روش های داده کاوی^۱ بهره گرفت. در حقیقت داده کاوی یک حوزه میان رشته ای و با رشد سریع است که حوزه های مختلفی هم چون پایگاه داده، آمار، یادگیری ماشین و سایر زمینه های مرتبط را با هم تلفیق کرده است تا اطلاعات و دانش ارزشمند نهفته در حجم بزرگی از داده ها را استخراج نماید. مدل های مختلفی در نقشه برداری رقومی خاک هم چون جنگل تصادفی، روش های فازی، شبکه عصبی مصنوعی، رگرسیون درختی، رگرسیون لجستیک به کار برده می شوند (۱۸ و ۲۰). الگوریتم جنگل تصادفی یک مدل توسعه یافته از مدل طبقه بندی و رگرسیون درختی^۲ است که داده های مشاهداتی و متغیرهای کمکی را به طور تکراری برای به دست آوردن ارتباط بهینه بین متغیر پاسخ و متغیرهای مستقل و انجام تخمین جداسازی می کند (۴۷). درون یابی با استفاده از روش های زمین آماری، امروزه به طور گسترده ای برای تهیه نقشه تغییرپذیری مکانی ویژگی های فیزیکوشیمیایی خاک (درصد رس، آهک، کربن آلی و واکنش خاک) مورد استفاده واقع می گردد (۳). به کمک زمین آمار^۳ علاوه بر توصیف الگوهای تغییرپذیری مکانی و زمانی داده ها، می توان اقدام به تخمین و تهیه نقشه های کمی پراکنش خصوصیات با حداقل واریانس نمود (۲۶). روش های زمین آماری کریگینگ معمولی و جهانی

4- Geostatistical analysis
5- Transportation
6- Back transportation

1- Data mining
2- Classification and regression tree
3- Geostatistic

زریک^۲ و حرارتی مزیک^۳ است (۴۸). واحدهای دشت‌های دامنه‌ای، فلات و تپه‌ها، اصلی‌ترین واحدهای سیمای اراضی در منطقه می‌باشند. خاک‌های مورد مطالعه بر اساس سامانه آمریکایی رده‌بندی خاک‌ها (۴۶) در دو رده اینسپتی سولز^۴ و انتی سولز^۵ قرار دارند.

مطالعات میدانی و تجزیه و تحلیل آزمایشگاهی: موقعیت نمونه‌های خاک (عمق ۰ تا ۳۰ سانتی‌متر) با استفاده از روش نمونه‌برداری تصادفی و افزونه ET Geowizard در محیط نرم‌افزار Arc-GIS 10.6.1 طراحی و محل ۱۴۵ نقطه مشاهداتی با سیستم موقعیت‌یاب جهانی^۶ برداشت شدند (شکل ۱). پس از هوا خشک کردن نمونه‌ها، کوبیدن و عبور از الک ۲ میلی‌متری، تجزیه‌های فیزیکی و شیمیایی بر روی آن‌ها در آزمایشگاه انجام گردید. بافت خاک به روش هیدرومتری (۱۴)، درصد کربنات کلسیم معادل به روش تیتراسیون (۴۱) درصد کربن آلی به روش والکل- بلک یا اکسیداسیون تر (۵۲) اندازه‌گیری گردید. انتخاب مناسب‌ترین متغیرهای محیطی با استفاده از کدنویسی در محیط نرم‌افزار RStudio نسخه 1.0.136 و بسته "VIF" و روش شاخص تورم واریانس^۷ صورت پذیرفت.

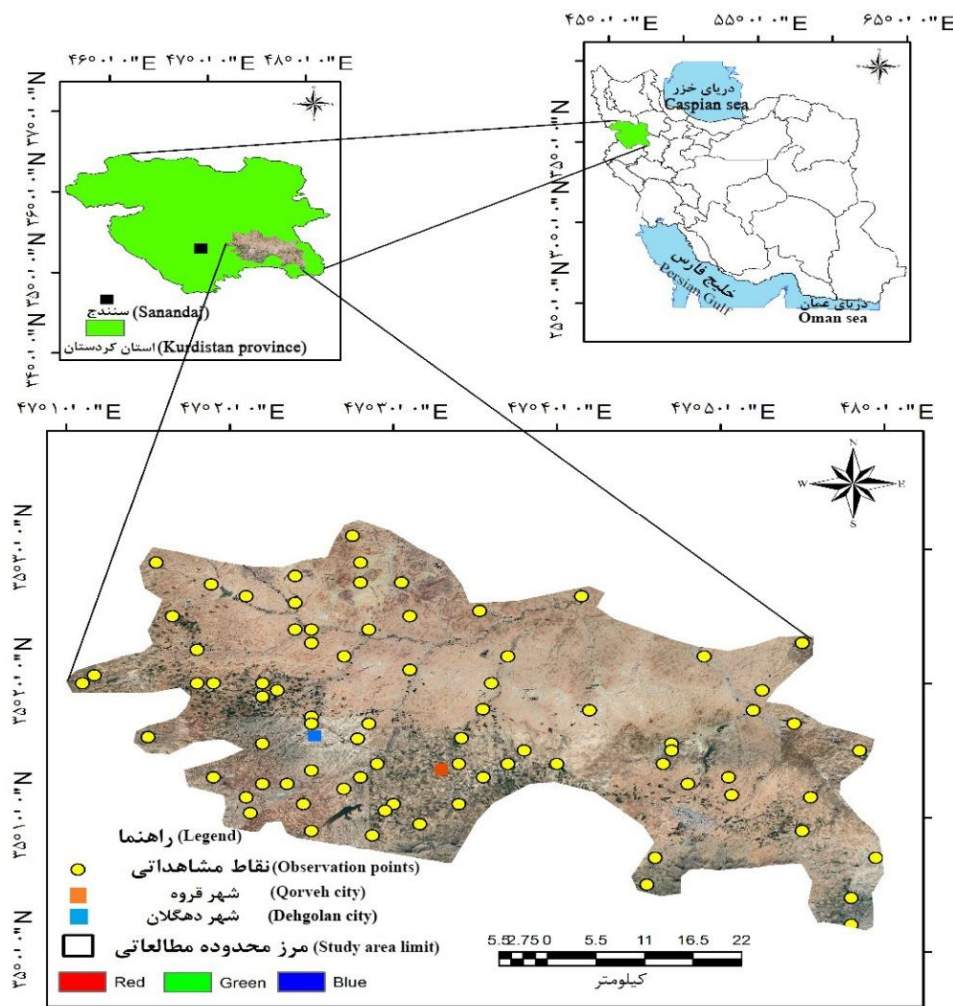
نمودن پیچیدگی توزیع مکانی متغیرهای خاک نسبت به روش‌های درون‌یابی خطی زمین‌آماري نتایج دقیق‌تر را پیش‌بینی می‌نمودند و تعداد اندکی مطالعات تأثیر تراکم داده‌های مشاهداتی بر نتایج مدل‌های خطی و غیرخطی را بررسی نموده‌اند، بنابراین مطالعه حاضر با هدف بررسی تأثیرپذیری پیش‌بینی مکانی ویژگی‌های کربن آلی، درصد رس، سیلت، ماسه و کربنات کلسیم معادل خاک سطحی از تراکم نقطه مشاهداتی مناسب با استفاده روش‌های مبتنی بر یادگیری ماشین با استفاده از متغیرهای کمکی سنجش‌ازدور و ژئومورفومتری و زمین‌آماري (بررسی خودهمبستگی مکانی) و همچنین تأثیر روش نوین انتخاب متغیر کمکی شاخص تورم واریانس (VIF) بر نتایج نهایی پیش‌بینی مکانی ویژگی‌های ذکر شده خاک سطحی در بخشی از اراضی دشت قروه دهگلان استان کردستان است.

مواد و روش‌ها

منطقه مورد مطالعه: منطقه مورد مطالعه با مساحت حدود ۲۵۰ هزار هکتار و متوسط ارتفاع ۱۸۵۶ متر از سطح دریای آزاد بخشی از اراضی دشت قروه- دهگلان واقع در استان کردستان است که بین طول جغرافیایی ۴۷ درجه و ۳۱ دقیقه تا ۴۸ درجه شرقی و عرض جغرافیایی ۳۵ درجه و ۱۰ دقیقه تا ۳۵ درجه و ۳۰ دقیقه شمالی قرار دارد (شکل ۱). کاربری غالب اراضی منطقه شامل مراتع و کشاورزی به صورت زراعت آبی و دیم است. این منطقه با متوسط درجه حرارت سالیانه هوای ۱۲ درجه سلسیوس و میانگین بارندگی سالیانه ۳۴۸ میلی‌متر در دوره آماری ۲۰ ساله (۱۳۷۷-۱۳۹۷) به همراه پردازش داده‌های اقلیمی در نرم‌افزار jNSM^۱ به ترتیب دارای رژیم‌های رطوبتی

2- Xeric
3- Thermic
4- Inceptisols
5- Entisols
6- Global Position System
7- Variance Inflation Factor

1- Java Newhall Simulation Model

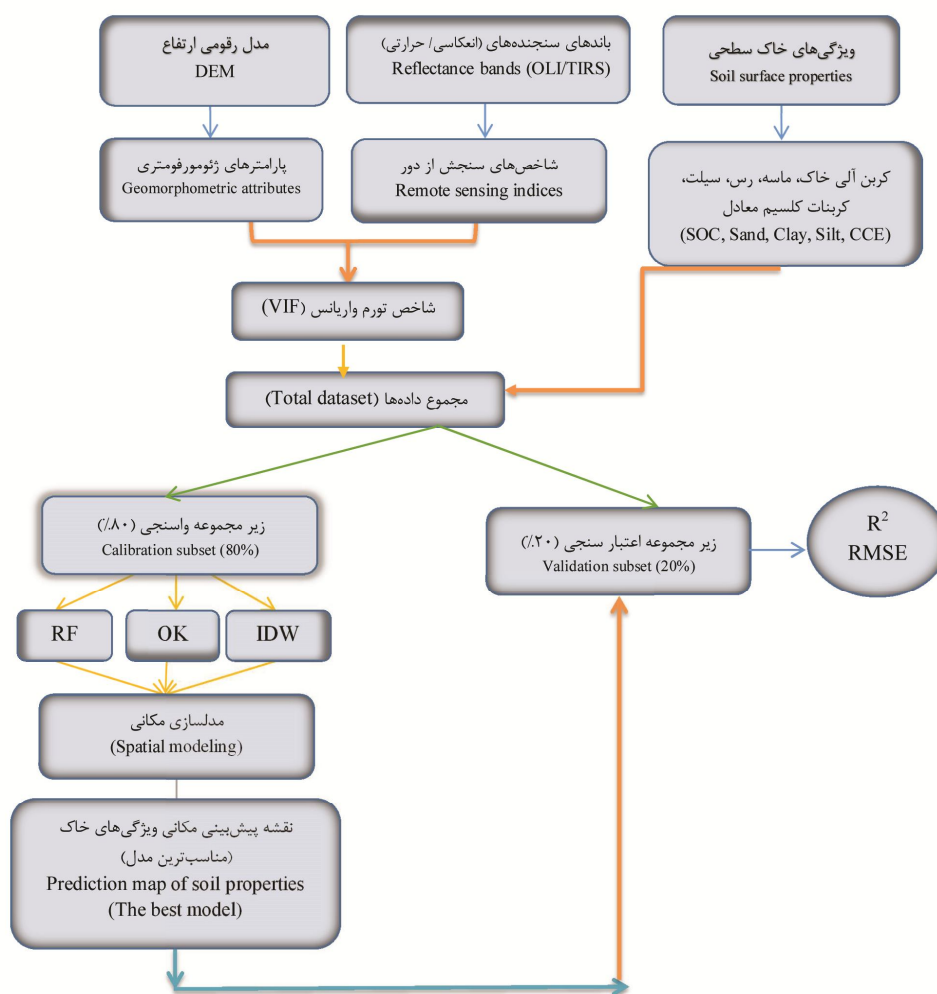


شکل ۱- موقعیت منطقه مورد مطالعه و نقاط نمونه برداری.

Figure 1. Location of the study area and observation points.

مدل رقومی ارتفاع با تفکیک مکانی ۱۲/۵ متر ماهواره ALOS، (۳) انتخاب بهینه ترین متغیرهای محیطی، (۴) مدل سازی مکانی داده ها بر اساس سه مدل RF، OK و IDW (۵) تهیه نقشه پیش بینی مکانی بر اساس مناسب ترین مدل پیش بینی کننده مکانی و اعتبار سنجی نتایج مدل سازی.

روندنمای پژوهش: در شکل ۲ روندنمای کلی این پژوهش ارائه شده است. این مراحل به ترتیب شامل موارد زیر می باشند: (۱) جمع آوری داده های خاک از طریق نمونه برداری با الگوی تصادفی، (۲) تهیه شاخص های سنجش ازدور و پارامترهای ژئومورفومتریک بر اساس تصویر ماهواره لندست ۸ مربوط به تاریخ ۱۸ خرداد ۱۳۹۷ که منطقه دارای پوشش گیاهی بود و



شکل ۲- روندنمای کلی پژوهش در منطقه مورد مطالعه (توسعه یافته توسط: نویسندگان).

Figure 2. Flowchart of the study in the study area (developed by :authors).

بر روی مدل رقومی ارتفاع در نهایت ۲۰ پارامتر زمین‌نما، هیدرولوژی و بازتابشی با استفاده از نرم‌افزار SAGA GIS 7.3 استخراج گردیدند. همچنین از تصویر لندست ۸ (۱۸ خرداد ۱۳۹۷) سنجنده‌های (باندهای انعکاسی / حرارتی) با قدرت تفکیک مکانی ۳۰×۳۰ متر استفاده گردید. در مجموع ۱۰ متغیر محیطی سنجنش از دور از طریق نسبت‌گیری بانندی حاصل گردید. سپس تمامی متغیرهای محیطی به فرمت رستری از طریق تابع نمونه‌گیری مجدد^۱ به

متغیرهای کمکی

مدل رقومی ارتفاعی: در نقشه‌برداری رقومی خاک برای مدل‌سازی تغییرات خصوصیات یا کلاس‌های خاک علاوه بر داده‌های مشاهداتی، داده‌های کمکی نیز نیاز است (۱۷). توپوگرافی یکی از مهم‌ترین فاکتورهای خاک‌ساز در مناطق خشک و نیمه‌خشک است (۳۸). متغیرهای کمکی مورد استفاده در این مطالعه از مدل رقومی ارتفاعی (DEM) با قدرت تفکیک مکانی ۱۲/۵×۱۲/۵ متر (ALOSPALSAR, 2006-2011) استفاده گردید (۴۴). پس از اعمال تصحیحات لازم

1- Resampling

روش‌های درون‌یابی: در این پژوهش برای پیش‌بینی تغییرپذیری مکانی ویژگی‌های خاک در نقاط نمونه‌برداری نشده از روش‌های درون‌یابی نیز استفاده گردید. روش‌های درون‌یابی مورد استفاده در این پژوهش شامل وزن‌دهی معکوس فاصله^۴ (رابطه ۱) و کریجینگ معمولی (رابطه ۲)^۵ می‌باشند. برای محاسبه واریانس به صورت پیوسته، مدل‌های تجربی مختلفی مثل مدل گوسی، کروی و نمایی به نیم‌تغییرنما برازش داده شد و بهترین مدل بر اساس ضریب تبیین (R^2) بالاتر و RMSE کمتر (رابطه ۳) انتخاب گردید.

$$Z^*(X_0) = \sum_{i=1}^n \left(\frac{1}{d_i}\right)^p Z(X_i) / \sum_{i=1}^n \left(\frac{1}{d_i}\right)^p \quad (1)$$

$$Z^*(X_0) = \sum_{i=1}^n \lambda_i \cdot Z(X_i) \quad (2)$$

در رابطه ۱: d_i فاصله بین نقطه تخمین تا هر کدام از نمونه‌های واقع در همسایگی آن، $Z(X_i)$ مقادیر نمونه‌های واقع در همسایگی محل تخمین و p یک عدد حقیقی مثبت است که پارامتر توان نامیده می‌شود. در رابطه ۲: $Z^*(X_i)$ عبارت است از مقدار متغیر مکانی برآورد شده در نقطه X_0 ، $Z(X_i)$ مقدار متغیر مکانی مشاهده شده در نقطه X_i و λ_i وزنی که به مقدار مشاهده شده در نقطه X_i نسبت داده می‌شود که بیانگر اهمیت نقطه i ام برآورد است.

نیم‌تغییرنما حاصل از زوج نقاط و بردار فاصله (h) ، $Z(x)$ و $Z(x+h)$ به ترتیب مقدار اندازه‌گیری شده متغیر در موقعیت (x) و مقدار اندازه‌گیری شده متغیر در موقعیت $(x+h)$ می‌باشند. با توجه به این‌که کریجینگ به نرمال بودن داده‌ها حساس است بنابراین برای ویژگی‌های درصد کربن آلی خاک و سیلت از

اندازه سلولی ۳۰ متر تبدیل و در نرم‌افزار SAGA GIS به فرمت sdat تبدیل و برای مدل‌سازی به نرم‌افزار RStudio منتقل گردیدند.

انتخاب متغیر کمکی محیطی بهینه

فاکتور تورم واریانس: در این مطالعه بر اساس رویکرد گام به گام فاکتور تورم واریانس و با استفاده از بسته^۱ "vif" در نرم‌افزار RStudio نسخه 1.0.136 و بر اساس حدود آستانه $VIF = 1$ (فاقد اثر هم خطی و همبستگی بین متغیرها)، $1 < VIF < 5$ (اثر هم خطی و همبستگی متوسط)، $5 < VIF < 10$ (دارای اثر هم خطی و همبستگی بالا)، $VIF > 10$ (اثر هم خطی و همبستگی خیلی بالا) مناسب‌ترین متغیرهای کمکی محیطی برای مدل‌سازی مکانی خصوصیات خاک انتخاب گردید. (۹).

مدل‌سازی مکانی

الگوریتم یادگیری ماشین به روش جنگل تصادفی^۲: در این پژوهش از الگوریتم غیرخطی یادگیرنده ماشینی جنگل تصادفی (RF) به‌عنوان یکی از روش‌های داده‌کاوی (۴) برای مدل‌سازی مکانی ویژگی‌های درصد کربن آلی، رس، ماسه، سیلت و کربنات کلسیم معادل خاک سطحی استفاده گردیده است. تمامی مراحل مدل‌سازی پیش‌بینی مکانی ویژگی‌های خاک بر اساس رویکرد داده‌کاوی جنگل تصادفی از بسته "Random Forest" در نرم‌افزار RStudio نسخه 1.0.136 و برای آنالیز اهمیت^۳ متغیرهای محیطی از شاخص درصد افزایش میانگین خطای پیش‌بینی (%IncMSE) از تابع "var Important" استفاده گردید.

4- Invers distance weighting
5- Ordinary kriging

1- Package
2- Random Forest
3- Relative important

نتایج و بحث

خلاصه آماری: خلاصه آماری ویژگی‌های خاک سطحی در عمق ۳۰ سانتی متری در جدول ۱ ارائه شده است. مقادیر کربن آلی خاک در منطقه از ۰/۱۴ تا ۳/۷ درصد متغیر است. میانگین آن در منطقه با ۰/۷۷ درصد نشان‌دهنده مقادیر کم کربن آلی در خاک‌های منطقه می‌باشد. حداقل و حداکثر درصد ذرات سیلت، رس و ماسه به ترتیب ۱۴/۸-۴۶/۳، ۴/۵۷-۱۰/۴ و ۳۱/۱۷-۱۷/۲-۶۳/۶ و میانگین آن‌ها به ترتیب ۳۴/۵۷، ۳۱/۱۷ و ۳۴/۲۳ درصد است. درصد کربنات کلسیم معادل نیز به ترتیب دارای حداقل، حداکثر و میانگین ۱/۲۵، ۴۳/۴ و ۲۳/۶۵ درصد است. بر اساس طبقه‌بندی ارائه شده توسط ویل‌دینگ (۱۹۸۵) کلاس ضریب تغییرات، ذرات تشکیل‌دهنده بافت خاک در کلاس ضریب تغییرات متوسط و دو ویژگی درصد کربن آلی خاک و کربنات کلسیم معادل در کلاس تغییرپذیری زیاد قرار دارند (۵۵). بیش‌ترین و کم‌ترین مقدار چولگی به ترتیب مربوط به دو ویژگی درصد سیلت و کربن آلی مشاهده گردید. هم‌چنین بیش‌ترین و کم‌ترین میزان افزایشگی به ترتیب در مورد درصد کربن آلی و کربنات کلسیم معادل حاصل گردید. جعفری و همکاران (۲۰۰۸) علت وجود چولگی بالا را به ویژگی‌های ذاتی متغیر، شرایط محیطی مانند فعالیت انسانی و نیز نحوه و تعداد نمونه برداشت شده نسبت دادند و استفاده از الگوی نمونه‌برداری شبکه‌ای در جهت کاهش چولگی را توصیه کردند (۲۵).

تبدیل لگاریتمی استفاده گردید و در انتها با استفاده از تابع تبدیل بازگشتی^۱ برای ارائه مقادیر واقعی این دو متغیر استفاده گردید.

ارزیابی صحت؛ واسنجی و اعتبارسنجی: به منظور مدل‌سازی مکانی ویژگی‌های خاک از مجموع داده‌ها (۱۴۵ نمونه)، ۸۰ درصد آن‌ها (۱۱۶ نمونه) برای واسنجی یا آموزش مدل و ۲۰ درصد داده‌ها (۲۹ نمونه) به عنوان داده‌های آزمون مدل انتخاب گردید. تمام مراحل مدل‌سازی با روش جنگل تصادفی در نرم‌افزار R studio و برازش نیم تغییرنا تجربی و اعتبارسنجی مدل‌های ریاضی در نرم‌افزار GS+ نسخه ۹ و درون‌یابی در جعبه‌ابزار تجزیه و تحلیل زمین‌آماري نرم‌افزار ArcGIS نسخه 10.6.1 صورت پذیرفت. در مدل جنگل تصادفی برای انتخاب بهینه دسته داده‌های مرحله واسنجی و آزمون مدل از روش بوت استرپت^۲ استفاده می‌شود (۱۰ و ۵۱).

به منظور ارزیابی صحت روش‌ها در این پژوهش، از دو شاخص آماری ریشه میانگین مربعات خطا (RMSE) و ضریب تبیین (R^2) استفاده شده است (رابطه‌های ۳ و ۴).

$$RMSE = \sqrt{\frac{1}{N} \sum (K_p - K_m)^2} \quad (3)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (K_m - K_p)^2}{\sum_{i=1}^n (K_m - \bar{K}_m)^2} \quad (4)$$

که در آن‌ها، N تعداد کل نمونه‌ها، K_m مقادیر اندازه‌گیری شده، K_p مقادیر پیش‌بینی شده و \bar{K}_m میانگین مقادیر اندازه‌گیری شده می‌باشد.

- 1- Back-transformed
- 2- Bootstrapping

جدول ۱- خلاصه توصیف آماری ویژگی های خاک در منطقه مورد مطالعه.

Table 1. Statistical Summary of Soil Properties in the Study Area.

متغیر Variable	حداقل Min	حداکثر Max	میانگین Mean	انحراف معیار D S.	چولگی Skewness	کشیدگی Kurtosis	ضرب تغییرات C.V (%)
کربن آلی خاک (%) SOC (%)	0.14	3.77	0.77	0.5	0.08	3.8	64
سیلت (%) Silt (%)	14.80	46.3	34.57	7.2	-1.04	3.7	20
رس (%) Clay (%)	10.40	57.40	31.17	6.34	-0.36	2.7	20.3
ماسه (%) Sand (%)	19.00	63.60	34.89	10.72	0.83	3.5	30.72
کربنات کلسیم معادل (%) CCE (%)	1.25	43.4	23.65	11.4	-0.25	2.5	48.2

از ویژگی های شیمیایی خاک است (۲۸ و ۳۱). موسوی و همکاران (۲۰۱۷) در مطالعه ای مدل نمایی را مناسب ترین مدل جهت توصیف پراکنش مکانی درصد سیلت و مدل کروی را برای درصد ماسه و رس گزارش نمودند (۴۰). کم ترین مقدار واریانس قطعه ای مربوط به درصد کربن آلی بود که این نشان دهنده این است که واریانس نسبی و حجم نمونه برداری برای آشکار کردن ساختارهای مکانی مناسب است (۴۰) و بیش ترین اثر قطعه ای در مورد درصد رس مشاهده گردید که نشانگر واریانس تصادفی قوی در فواصل کوتاه است که ناشی از خطای نمونه برداری، اندازه گیری و وجود تغییرات کوتاه دامنه و ویژگی مورد مطالعه در فواصل کوچک تر از کوتاه ترین فاصله نمونه برداری است (۲۱). بیش ترین مقدار آستانه نیم تغییرنا مربوط به درصد کربنات کلسیم معادل، رس و کم ترین مقدار آن در ویژگی های درصد کربن آلی و ماسه مشاهده گردید که نشان دهنده تصادفی بودن یا بدون ساختار بودن تغییرات آن یا واریانس کل تمام نمونه هایی است که در محاسبه نیم تغییرنا آن ها به کار رفته است.

خودهمبستگی مکانی: در این پژوهش نتایج برازش نیم تغییرنمای تجربی برای هر یک از ویژگی های خاک نشان داد که فاقد ناهمسانگردی است، بنابراین نیم تغییرنمای همسانگرد آن ها ترسیم گردید. بهترین مدل برازش داده شده بر روی ویژگی های خاک در جدول ۲ ارائه شده است. نتایج واریوگرافی نشان داد که به جز درصد کربنات کلسیم معادل که از مدل کروی تبعیت می نماید، در مورد سایر ویژگی ها مدل نمایی^۱ مناسب ترین مدل برای توصیف همبستگی مکانی آن ها بود. کدیا و ویرا (۲۰۰۸) بیان نمودند که مدل کروی یکی از مناسب ترین نیم تغییرنماها در تجزیه و تحلیل ویژگی های خاک و گیاه است (۸). فلاین و همکاران (۲۰۱۹) برای ویژگی های درصد سنگریزه، ماسه، سیلت، رس و کربن آلی خاک سطحی مدل نمایی را به عنوان مناسب ترین مدل برای برازش ساختار مکانی این ویژگی ها گزارش نمودند (۱۳). جونیور و همکاران (۲۰۰۷) و لئو و همکاران (۲۰۰۹) بیان نمودند که مدل نمایی یکی از مناسب ترین مدل ها برای بیان تغییرات مکانی بسیاری

1- Exponential

نمودند. نتایج کلاس همبستگی مکانی ویژگی‌ها بر اساس طبقه‌بندی کمبردلا و همکاران (۱۹۹۴) بیانگر این است که کلاس همبستگی مکانی در مورد ویژگی‌های درصد کربنات کلسیم معادل و ماسه (کلاس قوی) و ویژگی‌های کربن آلی، رس و سیلت در کلاس متوسط قرار دارند (۷). قوی بودن ویژگی‌های ساختار مکانی مورد مطالعه بیانگر آن است که استفاده از روش‌های زمین‌آماري به‌خوبی می‌تواند در تجزیه و تحلیل مکانی ویژگی‌های مورد مطالعه مفید باشد (۴۰).

بیش‌ترین و کم‌ترین مقدار دامنه تأثیر به ترتیب مربوط به درصد رس و ماسه بود که نشانگر وسعت تأثیرگذاری متغیرهای ذکر شده بر نقاط مجاورشان است (جدول ۲). پایین بودن دامنه تأثیر ماسه می‌تواند مربوط به ناپایداری ذاتی این ویژگی در طبیعت و یا تغییرات مکانی زیاد آن در فواصل کم‌تر از نقاط نمونه‌برداری باشد (۲۹). با افزایش دامنه تأثیر فواصل نمونه‌برداری افزایش یافته و تعداد نمونه‌های لازم برای نمونه‌برداری و در نتیجه هزینه‌های لازم برای نمونه‌برداری کاهش می‌یابد. بوستانی و همکاران (۳) کم‌ترین مقدار شعاع تأثیر را در مورد سیلت گزارش

جدول ۲- پارامترها و مناسب‌ترین مدل برازش داده شده بر روی نیم‌تغییرنا ویژگی‌های خاک.

Figure 2. Parameters and the most appropriate model fitted to the semi-variance soil properties.

RSS (%)	R ²	همبستگی مکانی Spatial correlation	واریانس	واریانس	دامنه تأثیر (متر) Effective range (m)	مدل Model	ویژگی‌های خاک Soil properties	
			قطعه‌ای / آستانه C ₀ / C %	آستانه Sill (C)				قطعه‌ای Nugget (C ₀)
5.06×10 ⁻⁵	0.73	متوسط Medium	50	0.12	0.06	4110	نمایی Exponential	کربن آلی خاک SOC
0.13	0.7	قوی Strong	0.11	89.10	0.1	8110	کروی Spherical	کربنات کلسیم معادل CCE
0.18	0.50	قوی Strong	15.71	0.70	0.11	1760	نمایی Exponential	ماسه Sand
1543	0.47	متوسط Medium	34.22	63.69	21.80	3130	نمایی Exponential	سیلت Silt
302	0.7	متوسط Medium	29.03	89.94	26.10	15930	نمایی Exponential	رس Clay

کریگینگ برای ویژگی‌های سنگریزه، ماسه، سیلت، رس و کربن آلی خاک سطحی به ترتیب، مقادیر RMSE ۰/۱۱، ۰/۱۴، ۰/۱۷ و ۰/۱۵ را در مقایسه با مدل‌های یادگیری ماشین (جنگل تصادفی و ماشین بردار پشتیبان) گزارش نمودند، هم‌چنین طبق نتایج ایشان، ویژگی درصد ماسه مشابه با نتایج این پژوهش بیش‌ترین تأثیرپذیری از اثرات خودهمبستگی

نتایج اعتبارسنجی: مناسب‌ترین مدل مورد استفاده برای تهیه نقشه پهنه‌بندی مکانی ویژگی‌های سطحی خاک بر اساس شاخص‌های اعتبارسنجی برای همه ویژگی‌ها مدل (RF) نسبت به مدل‌های درون‌یابی (OK, IDW) از دقت بالاتر و خطای کم‌تری برخوردار بود به‌جز درصد ماسه که روش OK دارای عملکرد بهتری بود (جدول ۳). فلاین و همکاران (۲۰۱۹) با به‌کارگیری مدل‌های ترکیبی رگرسیون

دارای بالاترین عملکرد و مدل‌های رگرسیون کریگینگ، جنگل تصادفی و کویست در رده‌های بعدی قرار گرفتند و در نهایت عنوان نمودند برای نقشه‌برداری در مقیاس مزرعه با تراکم نقاط مشاهداتی بالا، روش کریگینگ بدون متغیرهای کمکی می‌تواند نتایج قابل قبول ارائه نماید (۴۵).

پهنه‌بندی مکانی: نقشه پهنه‌بندی ویژگی‌های خاک بر اساس مناسب‌ترین مدل در شکل ۳ ارائه شده است. نقشه پیش‌بینی درصد کربن آلی (شکل ۳ الف) نشان داد که بیش‌ترین میزان آن در بخش‌های شمال غربی و جنوب غربی به‌صورت کاملاً محدود وجود دارد و بخش غالب منطقه به‌ویژه در مرکز که شدیداً تحت بهره‌برداری زراعت آبی و شخم در جهت شیب است، حداقل مقدار کربن آلی مشاهده گردید. درصد رس و سیلت الگوی پراکنش مشابهی را از خود نشان دادند و روند تغییرات مکانی آن‌ها برخلاف درصد ماسه مشاهده گردید. به‌نحوی که در بخش‌های مرکزی و شرقی منطقه که حداکثر مقدار این دو ویژگی وجود دارد حداقل مقدار درصد ماسه مشاهده می‌شود، در همین راستا در بخش‌های شمالی و تا حدودی در غرب و جنوب غرب منطقه بیش‌ترین مقادیر ماسه ملاحظه گردید (شکل‌های ۳ ب، پ و ث). الگوی تغییرات مکانی کربنات کلسیم معادل (شکل ۳ ت) برخلاف رس و سیلت است به‌طوری که بر روی تیغه‌ها و یال‌ها بیش‌ترین میزان آهک و کم‌ترین مقدار رس و سیلت مشاهده گردید؛ که بر اساس مشاهدات میدانی رخداد فرآیندهای فرسایش سطحی و شیاری به همراه تشدید جریان رواناب در این بخش‌ها (شیب بیش از ۱۵٪) موجب انتقال رس و سیلت از خاک سطحی موجب نمایان شدن آهک زیرسطحی در بخش سطحی و از طرفی به‌دلیل شستشوی کم‌تر آهک در خاک‌های مناطق نیمه‌خشک و خشک است (۵۴).

مکانی را نشان داده است (۱۳). روش کریگینگ معمولی (OK) در درون‌یابی خود اغلب از متوسط وزنی متغیرها بر اساس تئوری متغیر ناحیه‌ای استفاده می‌نماید و نتایج آن وابسته به خودهمبستگی مکانی متغیر هدف است (۳۶). دو روش OK و IDW تقریباً نتایج مشابهی را در پیش‌بینی ویژگی‌های خاک ارائه نمودند. در پژوهشی توسط زراعت‌پیشه و همکاران (۲۰۱۹) گزارش نمودند که روش RF نسبت به سایر مدل‌های مورد استفاده در آن پژوهش عملکرد مناسب‌تری را در خصوص پیش‌بینی مکانی درصد رس و کربن آلی خاک ارائه نموده است (۵۷). در همین راستا مطالعات متعددی توسط پژوهشگران بیان توانمندی بالای شاخص‌های پوشش گیاهی و سایر شاخص‌های مرتبط با آن‌ها در پیش‌بینی ویژگی‌های خاک گزارش نموده‌اند (۱۰، ۳۰، ۳۴ و ۵۱). مبنای انتخاب مناسب‌ترین مدل پیش‌بینی‌کننده رقومی بالاترین مقدار R^2 و کم‌ترین مقدار RMSE برای هر یک از ویژگی‌های مورد بررسی لحاظ گردید. نتایج نشان داد که برای متغیر کربن آلی خاک ($R^2=0/5$ و $RMSE=0/4$)، متغیر کربنات کلسیم معادل ($R^2=0/4$ و $RMSE=0/11/61$)، متغیر رس ($R^2=0/21$ و $RMSE=0/5/65$) و متغیر سیلت ($R^2=0/15$ و $RMSE=0/7/24$) و برای متغیر ماسه با ($R^2=0/14$ و $RMSE=10/26$) حاصل گردید. در نقشه‌برداری رقومی ویژگی‌های خاک مقادیر R^2 بالاتر از ۰/۷ غیرمعمول و مقادیر ۰/۵ و یا کم‌تر رایج است (۳۵). پولادی و همکاران (۲۰۱۹) در نقشه‌برداری رقومی ماده آلی خاک با استفاده از پنج مدل کریگینگ رگرسیون کریگینگ (ترکیب جنگل تصادفی - کریگینگ و کویست - کریگینگ)، جنگل تصادفی و کویست با استفاده از متغیرهای محیطی توپوگرافی، سنجش‌ازدور و سنجش از نزدیک بر اساس آماره‌های R^2 و RMSE گزارش نمودند که روش کریگینگ معمولی

در بخش‌هایی از زمین‌نما که دارای سطوح پایدارتری می‌باشند (دشت دامنه‌ای) به دلیل کاهش شدت جریان‌های سطحی تجمع قابل‌توجهی از رس و سیلت مشاهده می‌گردد.

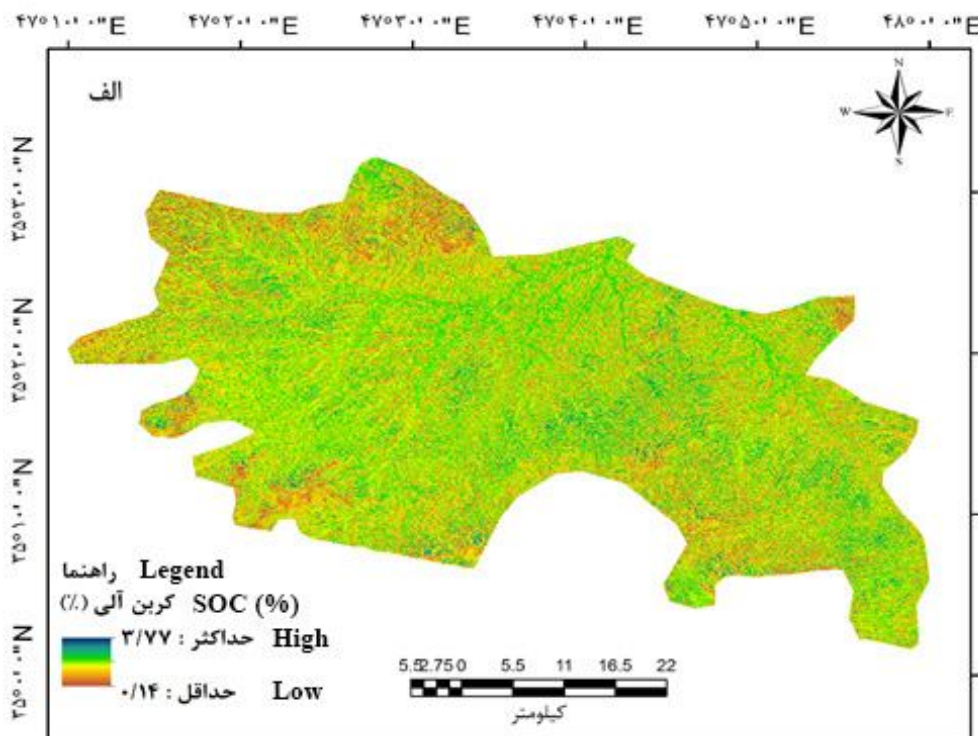
جدول ۳- نتایج اعتبارسنجی مدل‌سازی مکانی ویژگی‌های خاک.

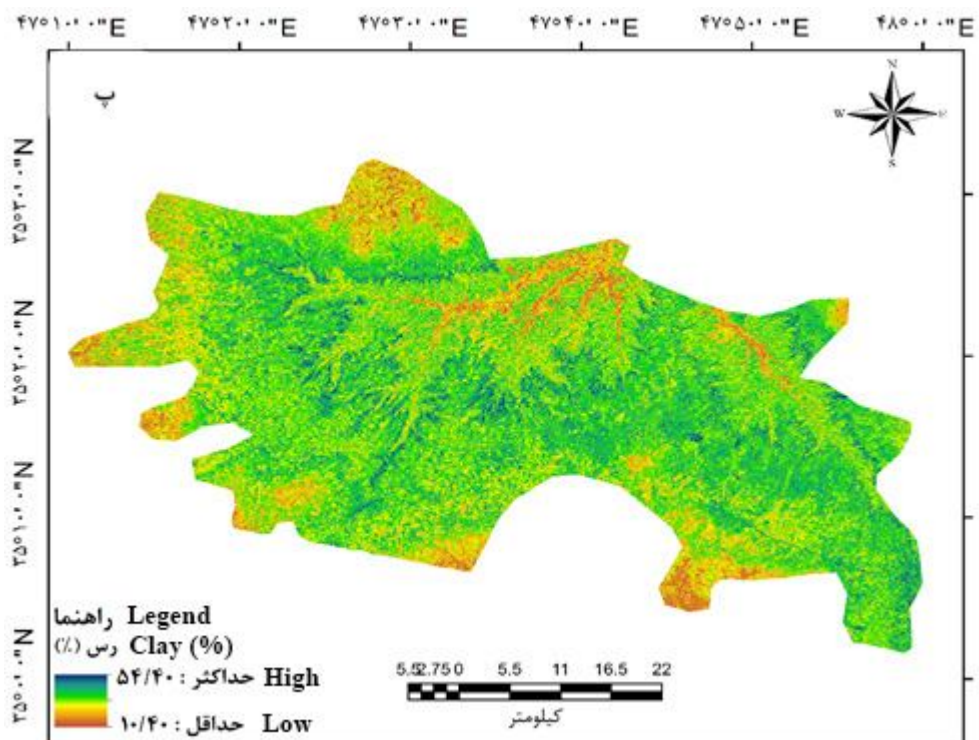
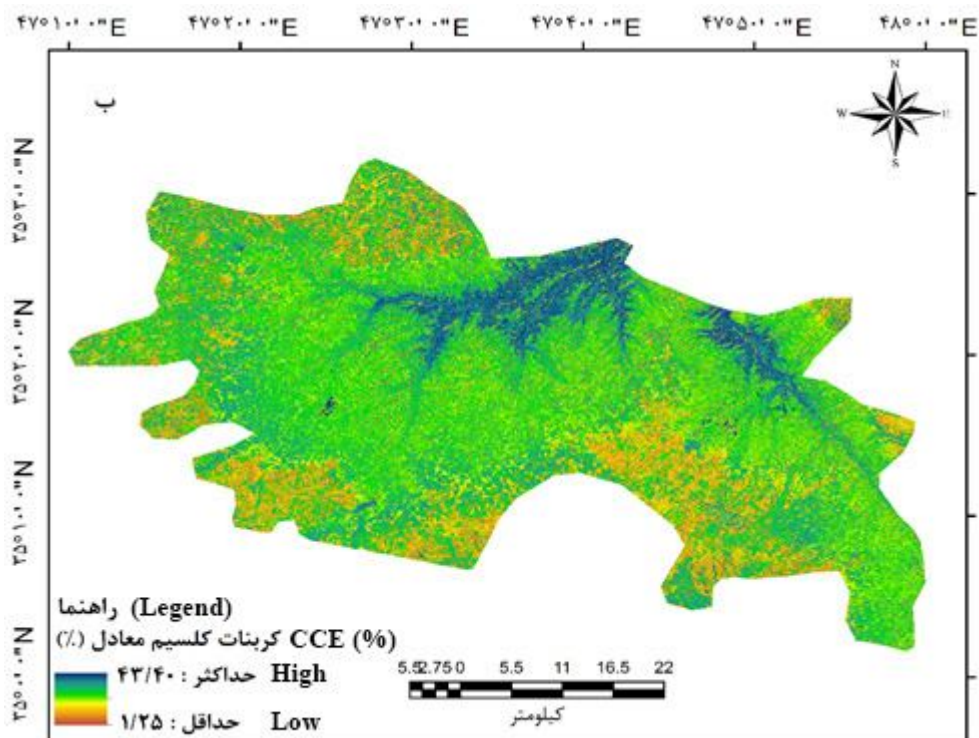
Table 3. Validation results of spatial modeling of soil properties.

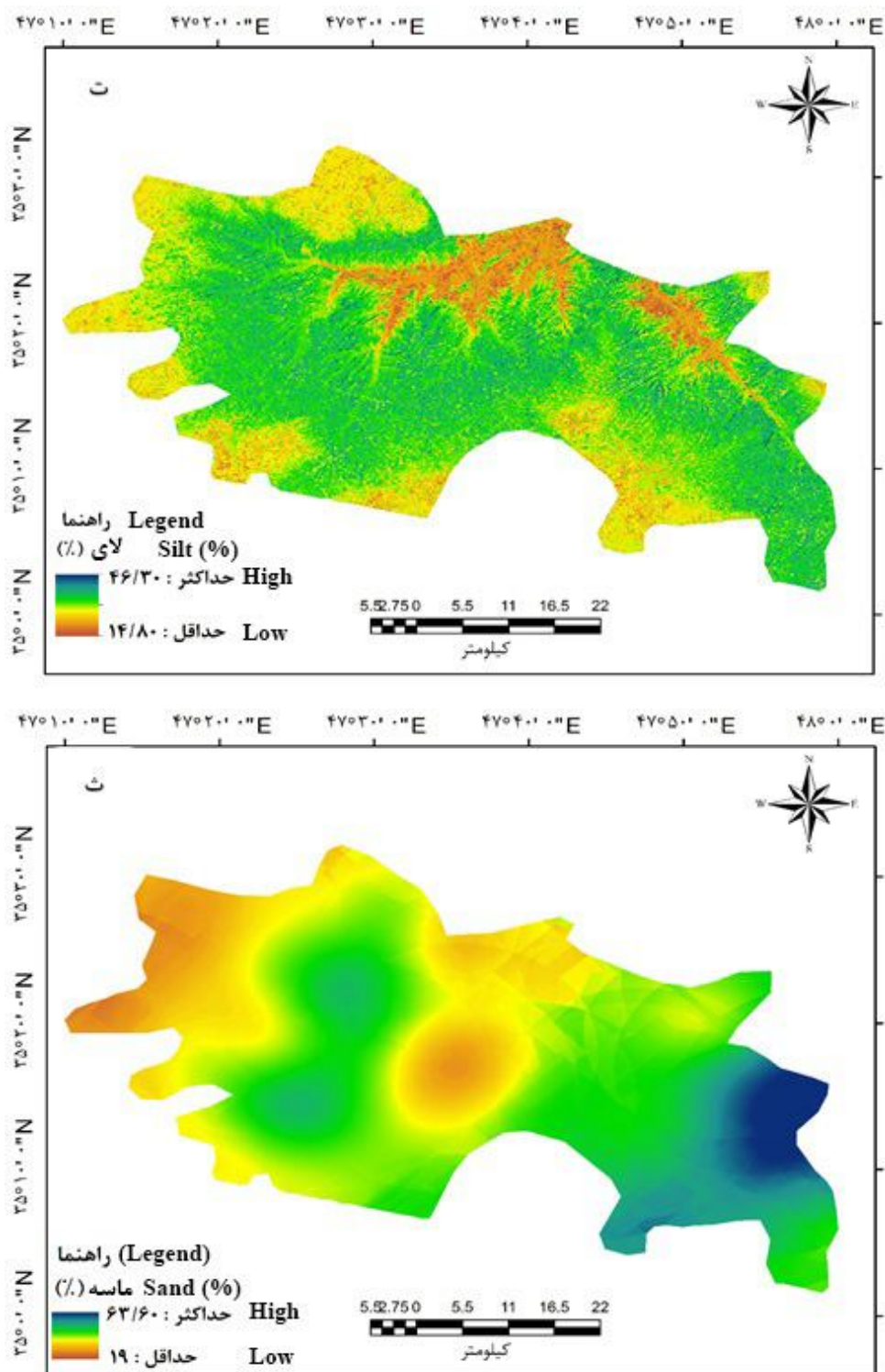
جنگل تصادفی (Random forest)		کریگینگ معمولی (OK)		وزندهی عکس فاصله (IDW)		مدل Model
RMSE (%)	R ²	RMSE (%)	R ²	RMSE (%)	R ²	
0.25	<u>0.5</u>	0.4	0.05	0.4	0.04	کربن آلی خاک (SOC)
10.25	<u>0.4</u>	11.61	0.002	11.85	0.03	کربنات کلسیم معادل (CCE)
5.37	<u>0.21</u>	5.65	0.19	6.51	0.05	رس (Clay)
6.60	<u>0.15</u>	7.24	0.006	6.67	0.11	سیلت (Silt)
13.1	0.11	10.26	<u>0.14</u>	10.30	0.12	ماسه (Sand)

RF, OK, IDW, SOC, CCE, Clay, Silt, Sand به ترتیب جنگل تصادفی، کریگینگ معمولی، وزندهی عکس فاصله، کربن آلی خاک، کربنات کلسیم معادل، رس، سیلت و ماسه. مناسب‌ترین مدل به صورت برجسته نسبت به بقیه نتایج مشخص شده است.

RF, OK, IDW, SOC, CCE, Clay, Silt, Sand in order Random forest, ordinary kriging, Inverse Distance Weighting, Soil Organic Carbon, Calcium Carbonate Equivalent, Clay, Silt, Sand. The most appropriate model is highlighted Compare to the rest of the results.







شکل ۳- پهنه‌بندی مکانی ویژگی‌های خاک بر اساس بهترین مدل پیش‌بینی‌کننده مکانی به ترتیب الف) درصد کربن آلی خاک، ب) درصد کربنات کلسیم معادل، پ) درصد رس، ت) درصد سیلت، ث) درصد ماسه.

Figure 3. Spatial distribution of SOC, CCE, Clay, Silt, and Sand contents using the best prediction model.

کربن آلی، سیلت و سنگریزه بیشتر تحت تأثیر متغیرهای محیطی بازتاب سطحی تصاویر ماهواره‌ای می‌باشند (۱۳). زراعت‌پیشه و همکاران (۲۰۱۹) نیز بیان نمودند که شاخص‌های توصیف‌کننده وضعیت پوشش گیاهی (RVI و SAVI) (۲) نقش مهم‌تری را در مدل‌سازی آهک خاک ایفا نمودند (۵۷). در همین راستا مطالعات متعددی توسط پژوهشگران در بیان توانمندی بالای شاخص‌های پوشش گیاهی و سایر شاخص‌های مرتبط با آن‌ها در پیش‌بینی ویژگی‌های خاک گزارش نموده‌اند (۱۰، ۱۲، ۵۱ و ۳۰). در مقابل در مورد درصد کربن آلی کربنات کلسیم معادل و رس پارامترهای ژئومورفومتری مدل رقومی ارتفاع شاخص موقعیت توپوگرافی و فاصله اقلیدسی از نزدیک‌ترین شبکه آبراهه در درجات بعدی که هر سه پارامترهای وابسته به تغییرات ارتفاع می‌باشند و بر روی درصد شیب، فرسایش و شدت جریان‌های سطحی تأثیر می‌گذارند (۱۹). کیدز و همکاران (۲۰۱۵) در پیش‌بینی درصد رس، پارامترهای ژئومورفومتری از جمله ارتفاع، ارتفاع بالاتر از شبکه کانال، عمق دره، شاخص همواری دره با قدرت تفکیک مکانی بالا را به‌عنوان مهم‌ترین پارامترها در توجیه تغییرپذیری مکانی درصد رس خاک سطحی گزارش نمودند. توپوگرافی نقش اساسی را در تعیین جهت جریان رواناب و شدت انرژی حمل مواد در سیمای اراضی بر عهده دارد و همین دلیل بر روی تکامل خاک‌ها و تغییرپذیری ویژگی‌های خاک تأثیر جدی می‌گذارد (۳۰). در مطالعه‌ای ادھیکاری و همکاران (۲۰۱۸) بیان نمودند مدل رقومی ارتفاع و پارامترهای مشتق شده از آن نقش اساسی را در نقشه‌برداری رقومی ویژگی‌های خاک در منطقه آرکانساس داشتند (۱)؛ بنابراین توپوگرافی از طریق فرآیندهای فرسایش و رسوب و همچنین تابش خورشیدی بر روی توزیع مکانی ویژگی‌های خاک مؤثر است (۶ و ۲۲).

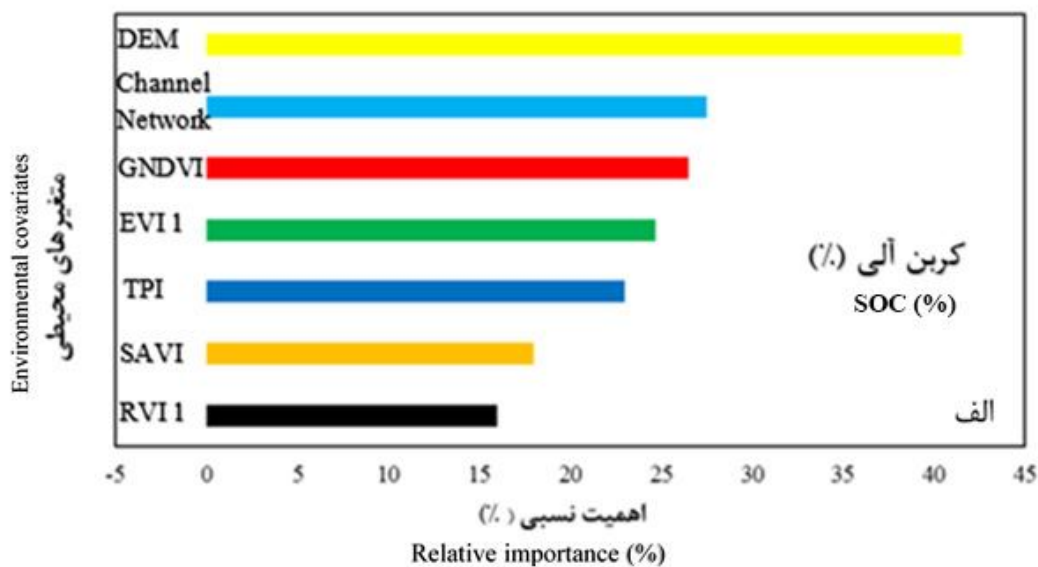
اهمیت نسبی متغیرهای محیطی با مدل جنگل تصادفی: در این مطالعه طبق روش انتخاب متغیر VIF از بین ۳۰ متغیر محیطی تهیه‌شده در نهایت هفت متغیر SAVI، EVI، GNDVI، RVI، DEM، Channel Network و TPI به‌عنوان مهم‌ترین پارامترهای پیش‌بینی‌کننده رقومی که بیش‌ترین واریانس تغییرات فاکتورهای خاک سازی را می‌توانند در ارتباط با پراکنش مکانی خاک‌های منطقه ارائه نمایند، انتخاب گردیدند (جدول ۴). در مدل جنگل تصادفی و سایر مدل‌های یادگیرنده ماشینی زیرمجموعه رگرسیون‌های درختی، محاسبه درصد میانگین خطای هر درخت یکی از پرکاربردترین پارامترها در تعیین اهمیت متغیرهای محیطی مورداستفاده برای مدل‌سازی مکانی ویژگی‌های خاک است (۱۵). اهمیت نسبی متغیرهای مورداستفاده در فرآیند مدل‌سازی بر اساس مدل RF در (شکل ۴ الف) ارائه شده است. متغیرهای Channel، DEM، Network و GNDVI مؤثرترین عوامل در پیش‌بینی کربن آلی خاک، شاخص‌های TPI و EVI، GNDVI و RVI دارای بیش‌ترین نقش را در پیش‌بینی کربنات کلسیم معادل (شکل ۴ ب)، در مورد اجزای ذرات بافت خاک نیز پارامترهای Channel، EVI و Network در مورد درصد رس، در خصوص توزیع مکانی سیلت RVI، Channel Network، SAVI و در نهایت برای پیش‌بینی مکانی ماسه شاخص‌های انعکاس طیف GNDVI، SAVI و RVI بالاترین درجه اهمیت را در فرآیند مدل‌سازی داشتند (شکل ۴ پ، ت و ث). به‌طورکلی در این منطقه شاخص‌های حاصل از نسبت‌گیری بانندی در مدل‌سازی مکانی ویژگی‌های درصد سیلت و ماسه خاک دارای اهمیت بیش‌تری بودند. فلاین و همکاران (۲۰۱۹) گزارش نمودند که تغییرپذیری مکانی دو ویژگی درصد ماسه و رس خاک سطحی نسبت به ویژگی‌های درصد

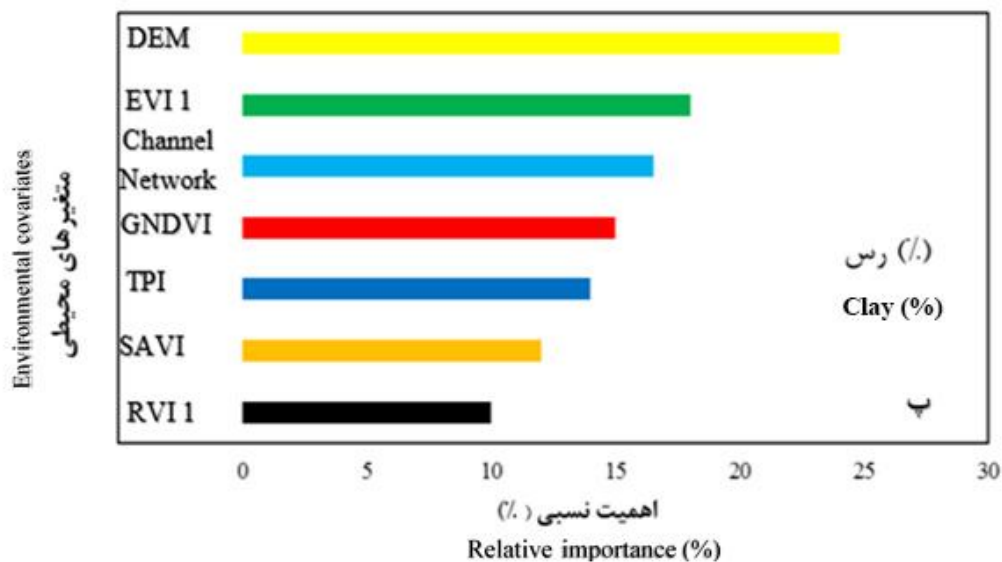
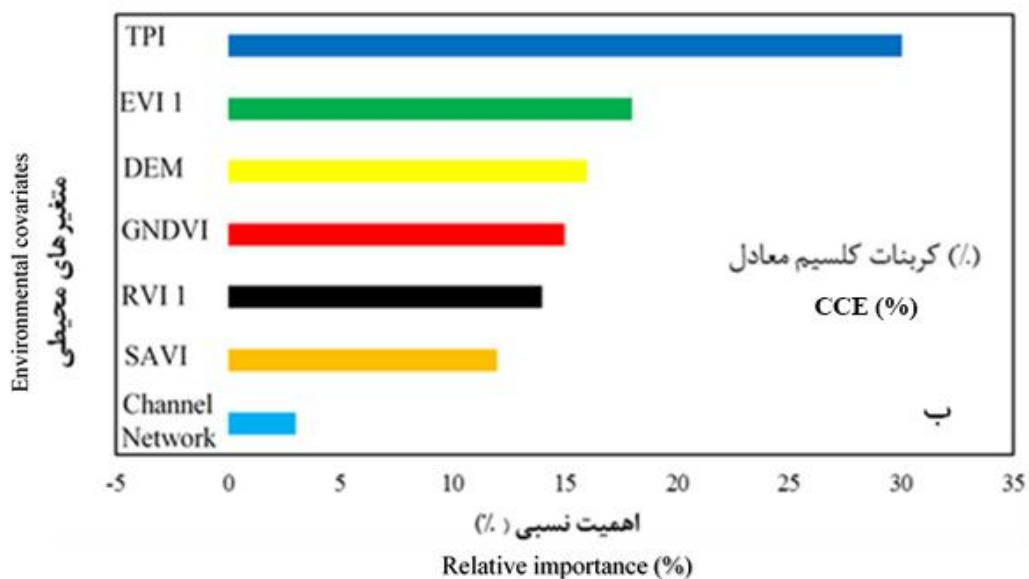
جدول ۴- متغیرهای محیطی مورد استفاده جهت پیش‌بینی ویژگی‌های خاک.

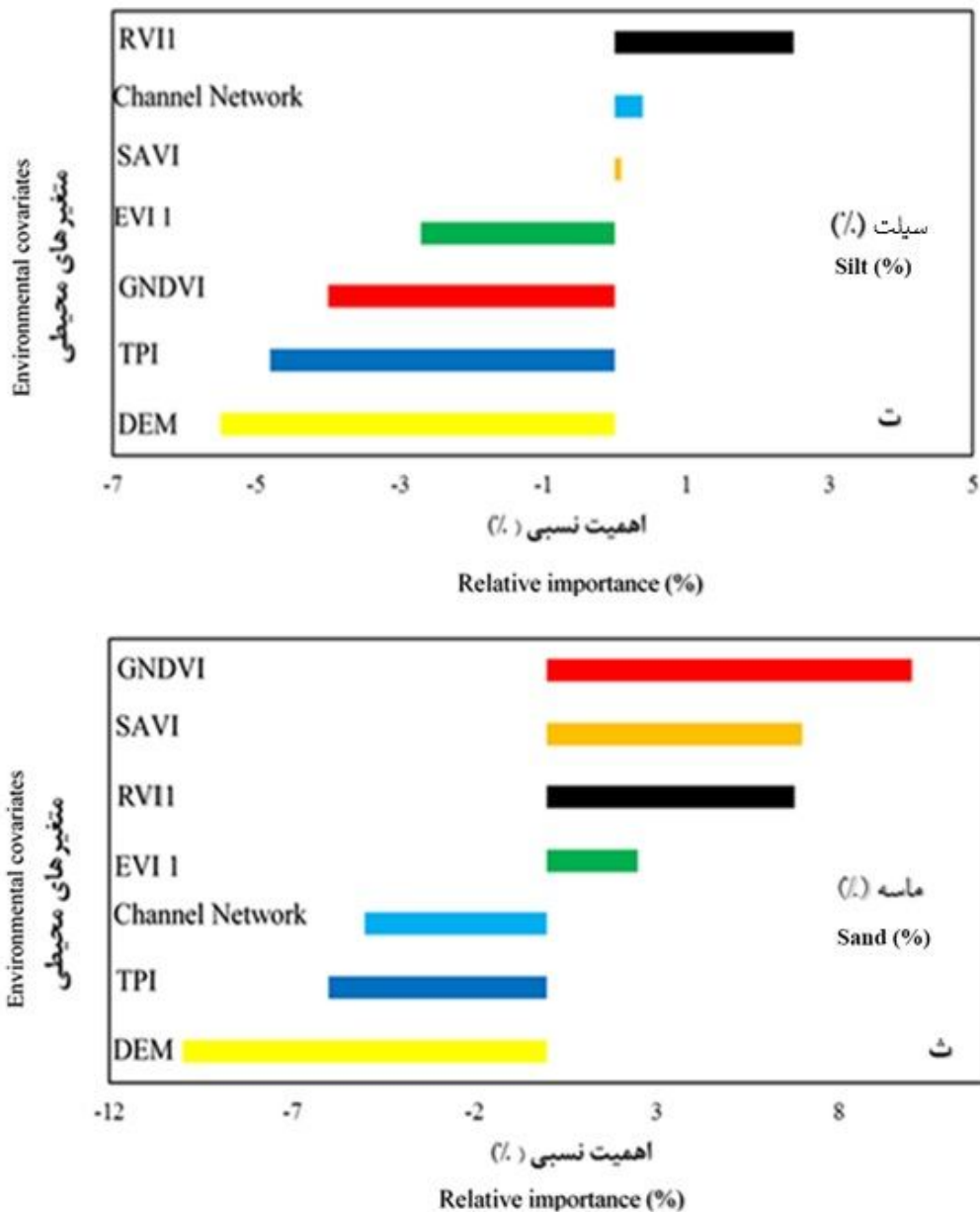
Table 4. Environmental covariates used as predictors soil characteristics.

منبع Reference	فرمول Formula	ماهیت متغیر variable inherent	نماد Symbol	نام متغیر Variable name
(۲۳)	$\frac{(NIR - RED)}{(NIR + RED + L) * (1 + L)}$		SAVI	شاخص پوشش گیاهی تعدیل‌یافته خاک Soil Adjustment Vegetation Index
(۲۴)	$\frac{2.5 * (NIR - RED)}{(NIR + 6RED - 7.5 * B + 1)}$	سنجش از دور Remote Sensing	EVI1	شاخص گیاهی بهبودیافته Enhanced Vegetation Index
(۱۶)	$\frac{NIR - G}{NIR + G}$		GNDVI	سبزی‌نگی شاخص گیاهی تفاضلی نرمال شده Green Normalized Diffusion Vegetation Index
(۲)	$\frac{NIR}{RED}$		RVII	شاخص پوشش گیاهی نسبی Relative Vegetation Index
(۴۴)	Z_s		DEM	مدل رقومی ارتفاع Digital Elevation Model
(۳۲ و ۱۹)	$Z_s - Z_b$	توپوگرافی Topography	Channel Network	فاصله عمودی تا شبکه آبراهه Vertical distance to channel network
(۴۳)	$Z_s - \sum_{n=1}^{n-1} \frac{Z_n}{n}$		TPI	شاخص موقعیت توپوگرافی Topographic Position Index

R, G و B به ترتیب باندهای ۵، ۴، ۳ و ۲ لندست ۸، مقدار ۰/۵ برای تعدیل اثر خاک پس‌زمینه، Z_s = ارتفاع پیکسل مورد نظر، Z_b = ارتفاع نسبت به نزدیک‌ترین شبکه آبراهه، Z_n = ارتفاع نسبت به شبکه آبراهه، n = تعداد کل پیکسل‌های مجاور که در ارزیابی مدنظر قرار می‌گیرند.







شکل ۴- اهمیت نسبی متغیرهای محیطی (%) برای هر ویژگی خاک با استفاده مدل جنگل تصادفی (RF).

Figure 4. Relative importance of environmental covariates (%) for each soil property by RF model.

دارای بالاترین میزان R^2 و کمترین مقدار RMSE برای ویژگی‌های درصد کربن آلی، کربنات کلسیم معادل، رس و سیلت بود و تنها روش کریگینگ معمولی در مورد درصد ماسه نتایج بهتری را ارائه نمود. استفاده از روش شاخص تورم واریانس به خوبی توانست از میان ۳۰ متغیر محیطی موجود، هفت متغیر محیطی بهینه را برای مدل‌سازی مکانی ویژگی‌های

نتیجه‌گیری کلی

پهنه‌بندی مکانی ویژگی‌های سطحی کربن آلی، کربنات کلسیم معادل، رس، سیلت و ماسه که در یک منطقه نیمه‌خشک غرب ایران با استفاده از نقشه‌برداری رقومی خاک موردبررسی قرار گرفت. مدل غیرخطی و ناپارامتریک جنگل تصادفی در مقایسه با مدل‌های درون‌یابی کریگینگ معمولی و وزن‌دهی عکس فاصله

ویژگی‌های مورد مطالعه (کربن آلی، کربنات کلسیم معادل، رس و سیلت) به جز در مورد ماسه ارائه نماید. در پایان برای مطالعات آتی پیشنهاد می‌گردد، از سایر روش‌های نمونه‌برداری مانند ابر مکعب لاتین مشروط یا طبقه‌بندی تصادفی با هدف نمونه‌برداری در داخل واحدهای همگن زمین‌نما به منظور ارتقا صحت و دقت نقشه‌های پیش‌بینی مکانی ویژگی‌های خاک استفاده شود.

خاک انتخاب نماید. از میان هفت متغیر محیطی انتخاب شده جهت مدل‌سازی مکانی برای سه ویژگی کربن آلی، کربنات کلسیم معادل و رس پارامترهای ژئومورفومتری DEM، TPI و Channel Network دارای اهمیت بالاتری بودند ولی در مورد ویژگی‌های سیلت و ماسه پارامترهای انعکاس طیف SAVI و EVI نقش بارزتری را در مدل‌سازی مکانی آن‌ها ایفا نمودند. مدل نیم‌تغییرنمای تجربی نمایی، توانست مناسب‌ترین برازش را از خود همبستگی مکانی

منابع

- Adhikari, K., Owens, P.R., Ashworth, A.J., Sauer, T.J., Libohova, Z., Richter, J.L., and Miller, D.M. 2018. Topographic controls on soil nutrient variations in a silvopasture system. *Agrosystems, Geosciences & Environment*. 1: 1. 1-15.
- Birth, G.S., and McVey, G.R. 1968. Measuring the color of growing turf with a reflectance spectrophotometer. *Agronomy Journal*. 60: 6. 640-643.
- Bostani, A., Salahedin, M., Rahman, M.M., and Namdar Khojasteh, D. 2017. Spatial mapping of soil properties using geostatistical methods in the Ghazvin Plains of Iran. *Modern Applied Science*. 11: 23-37.
- Breiman, L. 2001. Random forests. *Machine learning*. 45: 1. 5-32.
- Breiman, L., and Cutler, A. 2004. *Random Forests*. Department of Statistics, University of Berkeley.
- Brungard, C.W., Boettinger, J.L., Duniway, M.C., Wills, S.A., and Edwards Jr, T.C. 2015. Machine learning for predicting soil classes in three semi-arid landscapes. *Geoderma*. 239-240: 68-83.
- Cambardella, C.A., Moorman, T.B., Parkin, T.B., Karlen, D.L., Novak, J.M., Turco, R.F., and Konopka, A.E. 1994. Field-scale variability of soil properties in central Iowa soils. *Soil Science Society American Journal*. 58: 5. 1501-1511.
- Ceddia, M.B., Vieira, S.R., Villela, A.L.O., Mota, L.D.S., Anjos, L.H.C.D., and Carvalho, D.F.D. 2009. Topography and spatial variability of soil physical properties. *Scientia Agricola*. 66: 3. 338-352.
- Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carre, G., and Munkemuller, T. 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*. 36: 1. 27-46.
- Efron, B. 1979. Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*. 7: 1. 1-26.
- ESRI, 2011. *ArcGIS Desktop: Release 10*. Environmental Systems Research Institute, Redlands, CA.
- Falahatkar, S., Hosseini, S.M., Ayoubi, S., and Salmanmahiny, A. 2016. Predicting soil organic carbon density using auxiliary environmental variables in northern Iran. *Archives of Agronomy and Soil Science*. 62: 3. 375-393.
- Flynn, T., de Clercq, W., Rozanov, A., and Clarke, C. 2019. High-resolution digital soil mapping of multiple soil properties: an alternative to the traditional field survey. *South African Journal of Plant and Soil*. pp. 237-247.
- Gee, G.W., and Bauder, J.W. 1986. Particle-size analysis. *Methods of soil analysis: Part 1- Physical and mineralogical methods. methods of soil analysis*. pp. 383-411.
- Genuer, R., Poggi, J.M., and Tuleau-Malot, C. 2010. Variable selection using random forests. *Pattern Recognition Letters*, 31: 14. 2225-2236.

16. Gitelson, A.A., Kaufman, Y.J., and Merzlyak, M.N. 1996. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote sensing of Environment*. 58: 3. 289-298.
17. Golmohamadi, F., Nabiollahi, K., Taghizadeh-Mehrjardi, R., and Davari, M. 2017. Digital mapping of soil erodibility (Case study: Dehgolan, Kurdistan province). *Journal of Water and Soil Conservation*. 24: 87-103.
18. Grunwald, S. 2009. Multi-criteria characterization of recent digital soil mapping and modeling approaches. *Geoderma*, 152: 195-207.
19. Guo, Z., Adhikari, K., Chellasamy M., Greve, M.B., Owens, P.R., and Greve, M.H. 2019. Selection of terrain attributes and its scale dependency on soil organic carbon prediction. *Geoderma*. 340: 303-312.
20. Hartemink, A.E., McBratney, A., and Mendonca-Santos, M.D.L. 2008. Digital Soil Mapping Technologies for Countries with Sparse Data Infrastructures. *Digital Soil Mapping with Limited Data* Springer. pp. 15-30.
21. Hassani Pak, A.A. 2013. *Geostatistical*. University of Tehran press. (Translated in Persian)
22. Hook, P.B., and Burke, I.C. 2000. Biogeochemistry in a shortgrass landscape: control by topography, soil texture, and microclimate. *Ecology*. 81: 10. 2686-2703.
23. Huete, A.R. 1988. Soil Adjusted Vegetation Index (SAVI). *Remote Sensing of Environment*. 25: 295-309.
24. Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., and Ferreira, L.G. 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*. 83: 195-213.
25. Jafari, M., Asgari, M., Moazami, M., Biniiaz, M., and Tahmores, M. 2008. Spatial variation of some soil properties using geostatistical methods, *Research and Construction in Agriculture and Horticulture*. 80: 177-185. (In Persian)
26. Jiachun, S., Hazin, W., Jianming, X., Jinjun, W., Xingmei, L., Haiping, Z., and Shunlan, J. 2006. Spatial distribution of heavy metal in soil: A case of Changing, China. *Environmental Geology*. 10: 245-264.
27. Johnson, N., Ravnborg, H.M., Westermann, O., and Probst, K. 2002. User participation in watershed management and research. *Water policy*. 3: 6. 507-520.
28. Junior, L.A.Z., Lana, R.M.Q., and Guimaraes, E.C. 2007. Variabilidade espacial do pH, teores de matéria orgânica e micronutrientes em profundidades de amostragem num Latossolo Vermelho sob semeadura direta. *Ciência Rural*. 37: 4. 1000-1007.
29. Keskin, H., Grunwald S., and Harris, W.G. 2019. Digital mapping of soil carbon fractions with machine learning. *Geoderma*. 339: 40-58.
30. Kidd, D., Webb, M., Malone, B., Minasny, B., and McBratney, A. 2015. Eighty-metre resolution 3D soil-attribute maps for Tasmania, Australia. *Soil Research*. 53: 8. 932-955.
31. Liu, J., Shi, B., Jiang, H., Bae, S., and Huang, H. 2009. Improvement of water-stability of clay aggregates admixed with aqueous polymer soil stabilizers. *Catena*. 77: 3. 175-179.
32. Lin, H. 2010. Earth's critical zone and hydrogeology: concepts, characteristics, and advances. *Hydrology and Earth System Sciences*. 14: 25-45.
33. Ma, Y., Minasny, B., and Wu, Ch. 2017. Mapping key soil properties to support agricultural production in Eastern China. *Geoderma Regional*. 10: 144-153.
34. Mahmoudabadi, E., Karimi, A., Haghnia, G.H., and Sepehr, A. 2017. Digital soil mapping using remote sensing indices, terrain attributes, and vegetation features in the rangelands of northeastern Iran. *Environmental Monitoring and Assessment*. 189: 10. 500.
35. Malone, B.P., McBratney, A., Minasny, B., and Laslett, G. 2009. Mapping continuous depth functions of soil carbon storage and available water capacity. *Geoderma*. 154: 138-152.

36. McBratney, A.B., Odeh, I.O., Bishop, T.F., Dunbar, M.S., and Shatar, T.M. 2000. An overview of pedometric techniques for use in soil survey. *Geoderma*. 97: 3-4. 293-327.
37. McBratney, A.B., Mendonça-Santos, M.L., Minasny, B. 2003. On digital soil mapping. *Geoderma*. 117: 3-52.
38. Mehnatkesh, A., Ayoubi, S., Jalalian, A., and Sahrawat, K.L. 2013. Relationships between soil depth and terrain attributes in a semiarid hilly region in western Iran. *Journal of Mountain Science*. 10: 1. 163-172.
39. Minasny, B., and McBratney, A.B. 2016. Digital soil mapping: A brief history and some lessons. *Geoderma*. 264: 301-311.
40. Mousavi, S.R., Sarmadian, F., Dehghani, S., Sadikhani, M.R., and Taati, A. 2017. Evaluating inverse distance weighting and kriging methods in estimation of some physical and chemical properties of soil in Qazvin Plain. *Eurasian Journal of Soil Science*. 6: 4. 327-336.
41. Nelson, R.E. 1982. Carbonate and gypsum. P 181-179. In: A.L. Page (ed.) *Methods of soil analysis*. American Society of Agronomy, Madison.
42. Neter, J., Wasserman, W., and Kutner, M.H. 1989. *Applied linear regression models*. Chicago. 3rd. (ed.) 720: 519. 536.
43. Olaya, V. 2004. A gentle introduction to SAGA GIS. The SAGA User Group eV, Göttingen, Germany. 208p.
44. PALSAR, A. 2016. Japan aerospace exploration agency. Available from: <https://www.asf.alaska.edu/sar-data/palsar/>.
45. Pouladi, N., Moller, A.B., Tabatabai, S., and Greve, M.H. 2019. Mapping soil organic matter contents at field level with Cubist, Random Forest and kriging. *Geoderma*. 342: 85-92.
46. Soil Survey Staff. 2014. *Keys to soil taxonomy*. 12th ed. USDANatural Resources Conservation Service, Washington, DC.
47. Taghizadeh-Mehrjardi, R. 2016. *Modern concepts in Soil Science (Pedometric)*. Ardakan Univ. Press. 311p. (In Persian)
48. 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 USA. pp. 1-9.
49. Veronesi, F., and Schillaci, C. 2019. Comparison between geostatistical and machine learning models as predictors of topsoil organic carbon with a focus on local uncertainty estimation. *Ecological Indicators*. 101: 1032-1044.
50. Liaw, A., and Wiener, M. 2002. Classification and regression by Random forest. *R news*. 2: 3. 18-22.
51. Viscarra Rossel, R.A., Webster, R., and Kidd, D. 2014. Mapping gamma radiation and its uncertainty from weathering products in a Tasmanian landscape with a proximal sensor and random forest kriging. *Earth Surface Processes and Landforms*. 39: 6. 735-748.
52. Walkley, A., and Black, I.A. 1934. An examination of the digestion method for determining soil organic matter and a proposed modification of the chromic acid titration method. *Soil Science*. 37: 29-38.
53. Webster, R., and Oliver, M.A. 2007. *Geostatistics for environmental scientists*: John Wiley and Sons. 330p.
54. Wilford, J., De Caritat, P., and Bui, E. 2015. Modelling the abundance of soil calcium carbonate across Australia using geochemical survey data and environmental predictors. *Geoderma*. 259: 81-92.
55. Wilding, L. 1985. *Spatial variability: its documentation, accommodation and implication to soil surveys*. pp. 166-194.
56. Witten, I.H., Frank, E., and Hall, M.A. 2011. *Data Mining: Practical machine learning tools and techniques*: Morgan Kaufmann. pp. 1-525.
57. Zeraatpisheh, M., Ayoubi, S., Jafari, A., Tajik, S., and Finke, P. 2019. Digital mapping of soil properties using multiple machines learning in a semi-arid region, central Iran. *Geoderma*. 338: 445-452.



Spatial Prediction Some of the Surface Soil Properties Using Interpolation and Machine Learning Models

S.R. Mousavi¹, F. Parsaie², A. Rahmani¹, *M.H. Sedri³ and M. Kohsar Bostani⁴

¹Ph.D. Student of Soil Resource Management, Dept. of Soil Science and Engineering, University of Tehran, ²Ph.D. Student of Soil Resource Management, Dept. of Soil Science and Engineering, University of Shahid Chamran, ³Assistant Prof. of Soil and Water Research Department, Kurdistan Agricultural and Natural Resources Research and Education Center, AREEO, Sanandaj, Iran, ⁴Researcher of Soil and Water Research Department, Kurdistan Agricultural and Natural Resources Research and Education Center, AREEO, Sanandaj, Iran

Received: 10.21.2019; Accepted: 07.12.2020

Abstract

Background and Objectives: Accurate and detailed spatial soil information over the landscape is essential for the precision monitoring of land resources, hydrological applications, land use management. The present study aimed to predict the spatial prediction of SOC, CCE, Clay, Silt, and Sand in the Qorve-Dehgolan region, Kurdistan province.

Materials and Methods: Qorve-Dehgolan region, with mean annual temperature and precipitation of 12 °C and 348 mm (20-year statistical period), has soil moisture and temperature regimes xeric and thermic, respectively. A total of 145 samples were collected from the topsoil (0-30 cm) based on a random sampling pattern. Then, all of the soil samples were transferred to a soil laboratory for physicochemical analysis. Random forest (RF) as a nonparametric model and Ordinary kriging (OK) and inverse distance weighting (IDW) as an interpolation method were used for modeling the soil properties and their spatial autocorrelation. All steps of modeling for RF and interpolation methods (OK and IDW) were performed in RStudio, ArcGIS and, GS⁺ softwares, respectively. A total of 30 environmental covariates, including the Digital Elevation Model (DEM) derivatives in the SAGA GIS 7.3 and Landsat 8 satellite reflective band data in the ERDAS IMAGINE softwares, were developed as environmental variables. All of the environmental covariates were resampled at resolution-30 m. The most appropriate covariates were selected according to the variance inflation factor (VIF). Modeling of soil properties was performed according to 80% and 20% of data sets, respectively for calibration and validation, and two statistics of root mean square error (RMSE) and determination coefficient (R^2) was used to determine the accuracy of the models.

Results: Seven variables including SAVI, EVI1, GNDVI, RVI1, DEM, Channel Network, and TPI were selected from the 30 variables prepared as the most appropriate auxiliary variables based on the variance inflation index. Four remote sensing variables include the soil adjusted vegetation index (SAVI), the green normalized difference vegetation index (GNDVI), the relative vegetation index (RVI) and the enhanced vegetation index (EVI) and three geomorphometric attributes including, digital elevation model (DEM), vertical distance to channel network and the topographic position index (TPI) were the most important parameters. The results of modeling showed that RF model for soil organic carbon variable ($R^2=0.5$ and %RMSE=0.4), calcium carbonate equivalent ($R^2=0.4$ and %RMSE=11.61), clay variable ($R^2=0.21$ and %RMSE=5.65), the Silt variable ($R^2=0.15$ and %RMSE=7.24) and, Ordinary kriging methods for sand variables with ($R^2=0.14$ and %RMSE=10.26) was the most accurate

* Corresponding Author; Email: sedri_mh@yahoo.com

than RF and IDW models. Among the semi-variogram models, the exponential model had the best performance for soil organic carbon, clay, silt, and sand percentage, with the exception of CCE which follows the spherical model. The results of spatial autocorrelation showed that for both variables CCE and Sand had a strong class and, the others had a moderate class. The highest values of the semi-variogram sill were related to the calcium carbonate equivalent and clay, and the lowest values were related to the soil organic carbon and sand contents. These results indicate that, the existence of a random pattern or weak spatial structure in the samples that used to calculate the experimental semi-variogram. Among the seven environmental covariates were used for spatial modeling of top-soil organic carbon, clay and calcium equivalent carbonate, the geomorphometric attributes such as DEM topographic position index are of the most important and NDVI, SAVI and RVI covariates were more important in predicting of sand and silt properties.

Conclusion: Generally, topsoil properties had moderate and strong spatial autocorrelation, but the spatial prediction results were not highly accurate. Therefore, it would be recommended that in future studies other sampling methods like that Conditional Latin hypercube or stratified random and thematic maps such as geomorphology, geology, and soil map units as inputs for spatial modeling toward enhanced modeling performance will be used.

Keywords: Data mining, Digital soil mapping, Random forest, Spatial variability

