

## ارزیابی روش‌های یادگیری ماشین در نقشه‌برداری رقومی کربن آلی خاک‌های زراعی (بخشی از دشت خرم‌آباد)

حمیدرضا متین‌فر<sup>۱\*</sup>، زیبا مقصودی<sup>۱</sup> و سید روح‌اله موسوی<sup>۲</sup> و محبوبه جلالی<sup>۱</sup>

(تاریخ دریافت: ۱۳۹۸/۱۱/۱۶؛ تاریخ پذیرش: ۱۳۹۹/۹/۵)

### چکیده

شناخت توزیع مکانی کربن آلی خاک یکی از ابزارهای کاربردی در تعیین استراتژی‌های مدیریت پایدار اراضی است. طی دو دهه اخیر استفاده از رویکردهای داده‌کاوی در مدل‌سازی مکانی کربن آلی خاک با استفاده از تکنیک‌های یادگیری ماشین به‌طور گسترده‌ای مورد توجه قرار گرفته است. یکی از گام‌های اساسی در کاربرد این روش‌ها، تعیین متغیرهای بهینه پیش‌بینی کننده کربن آلی خاک است. این مطالعه به منظور مدل‌سازی و نقشه‌برداری رقومی کربن آلی خاک سطحی با استفاده از روش‌های یادگیری ماشین و ویژگی‌های خاک شامل درصد سیلت، رس، شن، کربنات کلسیم معادل، میانگین وزنی قطر خاکدانه و اسیدیته انجام پذیرفت. بدین منظور دقت عملکرد مدل‌های جنگل تصادفی، کویست، رگرسیون حداقل مربعات جزئی، رگرسیون خطی چندمتغیره و کریجینگ معمولی برای برآورد میزان کربن آلی خاک سطحی، در ۱۴۱ نمونه از عمق ۰-۳۰ سانتی‌متر در بخشی از اراضی کشاورزی دشت خرم‌آباد با مساحت ۶۸۰ هکتار مورد ارزیابی قرار گرفتند. نتایج آنالیز حساسیت متغیرهای پیش‌ران در مدل‌سازی کربن آلی نشان داد که به ترتیب سه ویژگی درصد سیلت، آهک و میانگین وزنی قطر خاکدانه بیشترین تأثیر را روی تغییرپذیری مکانی کربن آلی خاک داشتند. همچنین مقایسه رویکردهای مختلف تخمین کربن آلی نشان داد که مدل جنگل تصادفی به ترتیب با مقادیر ضریب تبیین ( $R^2$ ) و مجذور میانگین مربعات خطا (RMSE) ۰/۷۵ و ۰/۲۵ درصد بهترین کارایی را نسبت به سایر رویکردهای مورد استفاده در منطقه مطالعاتی ارائه کرد. در مجموع مدل‌های با رویکرد غیرخطی صحت بالاتری نسبت به مدل‌های خطی در مدل‌سازی تغییرات مکانی کربن آلی خاک نشان دادند.

واژه‌های کلیدی: تغییرپذیری مکانی، نقشه‌برداری رقومی خاک، روش‌های مدل‌سازی، پیش‌بینی کربن آلی

۱. گروه علوم و مهندسی خاک، دانشکده کشاورزی، دانشگاه لرستان، خرم‌آباد

۲. گروه علوم و مهندسی خاک، دانشکده مهندسی و فناوری کشاورزی، پردیس کشاورزی و منابع طبیعی، دانشگاه تهران، کرج

\*. مسئول مکاتبات: پست الکترونیکی: [matinfar44@gmail.com](mailto:matinfar44@gmail.com)

## مقدمه

کربن آلی خاک یک جزء اصلی از چرخه جهانی کربن است (۲۲) و به عنوان بزرگ‌ترین سهم از ذخایر کربن آلی زمینی شناخته شده است (۲۳). سهم جهانی کربن آلی خاک برابر با ۱۵۵۰ گیگا تن (Gigatone) است که این مقدار تقریباً دو برابر بیش از میزان کربن جوی فعلی و ۱۵۰ برابر بیشتر از میزان مصرف سالیانه سوخت‌های فسیلی است (۴۰). با وجود این کربن آلی به عنوان یک منبع پتانسیل از کربن جوی است که هر تغییر جزئی در میزان کربن آلی خاک در مناطق وسیع به‌طور قابل توجهی بر چرخه کربن در اکوسیستم‌های جهانی تأثیر می‌گذارد (۳۱). به علت نقش خاک در چرخه جهانی کربن و نقش پر اهمیت پایش ماده آلی خاک برای برآورد کیفیت و مدیریت خاک، تقاضا برای دریافت اطلاعات کربن آلی خاک در حال افزایش است اما دست‌یابی به این اطلاعات با روش‌های مرسوم نقشه‌برداری خاک فعالیتی زمان‌بر و پرهزینه محسوب می‌شود (۸). به‌طور کلی میزان کربن آلی خاک توسط سه عامل اقلیم (دما و میزان رطوبت)، ویژگی‌های فیزیکی و شیمیایی (اجزای بافت شامل درصد ذرات شن، سیلت و رس، ساختمان و کانی‌شناسی) و عوامل بیولوژیکی کنترل می‌شود. در مناطق خشک مقدار ماده آلی توسط ویژگی‌های خاک با توجه به میزان برگشت بقایای گیاهی و درجه حرارت هوا (۲) و در مناطق استوایی و گرم و مرطوب، پوشش گیاهی و ترکیب کانی‌شناسی خاک مهم‌ترین فاکتورهای کنترل‌کننده میزان ماده آلی در خاک‌ها هستند (۳). بنابراین برای دست یافتن به اطلاعات کامل از نحوه تغییرات مکانی کربن آلی در خاک از دانش توزیع مکانی و نقشه‌برداری آن می‌توان بهره گرفت. با توجه به اینکه کربن آلی خاک به عنوان یک منبع مهم و متغیر در اکوسیستم‌های زمینی محسوب می‌شود، آگاهی داشتن از تغییرات مکانی آن اطلاعات مفیدی در مورد وضعیت حاصلخیزی خاک، تغییرات آب و هوایی و اثرات استفاده مداوم از اراضی را ارائه می‌دهد که مورد توجه بهره‌داران و مدیران اراضی است. در روش‌های مرسوم نقشه‌برداری خاک بدون در نظر گرفتن تغییرات مکانی

کربن آلی خاک، میانگین کربن آلی خاک در هر واحد نقشه خاک گزارش می‌شود که این نحوه مطالعه تغییرات مکانی به دلیل عدم ارائه تغییرات پیوستگی مکانی این ویژگی نیازمند بهره‌گیری از روش‌های نوینی است که قادر باشند تغییرات مکانی خصوصیات خاک را در نظر گرفته و موجب صرفه‌جویی در زمان و هزینه مطالعات تهیه نقشه ویژگی‌های خاک شوند (۳۶). نقشه‌برداری رقومی خاک امروزه با استفاده از ابزارها و قابلیت‌های متعددی که دارد، می‌تواند برای دست‌یابی به داده‌هایی با مختصات مکانی مشخص در یک سیستم اطلاعات مکانی خاک مورد استفاده قرار گیرد و امکان تجزیه و تحلیل حجم وسیعی از داده‌ها را فراهم کند (۱۰ و ۳۰). در پژوهش‌های مختلف، گستره گوناگونی از دقت برای برآورد کربن خاک ارائه شده است. زراعت‌پیشه و همکاران (۴۱) در مطالعه‌ای توزیع مکانی ویژگی‌های خاک را با استفاده از متغیرهای محیطی و تجزیه و تحلیل‌های آماری چندمتغیره در منطقه نیمه‌خشک بروجن مورد بررسی قرار دادند. در این مطالعه تعداد ۳۳۴ نمونه خاک از عمق ۳۰-۰ سانتی‌متری برداشت شد، و از ویژگی‌های توپوگرافی مستخرج از مدل رقومی ارتفاع (شاخص رطوبت پستی بلندی، شاخص رطوبت، شاخص همواری کف دره و شیب) و شاخص‌های سنجش از دور (شاخص نرمال شده تفاوت پوشش گیاهی (Normalized Difference Vegetation Index)، شاخص گیاهی نسبی (Ratio Vegetation Index)، شاخص طیفی پوشش گیاهی تعدیل شده (Soil Adjusted Vegetation Index)) حاصل از ماهواره لندست ۷ ETM استفاده شد. نتایج آنها نشان داد خاک‌هایی که مقدار کربن آلی و رس بالاتری دارند در تالاب‌ها، رسوبات بادبزی و تپه‌ها و خاک‌هایی که مقدار کربن و رس کمتری دارند در تپه‌های شیب‌دار، تپه‌های فرسایش یافته و بیرون‌زدگی‌های سنگی قرار دارند. به‌طور کلی نتایج ایشان بیانگر این بود که عوامل محیطی مانند مورفولوژی تپه‌ها و کلیه متغیرهای سنجش از دور تا حد زیادی می‌توانند تغییرپذیری کربن آلی خاک را در منطقه مورد مطالعه توضیح

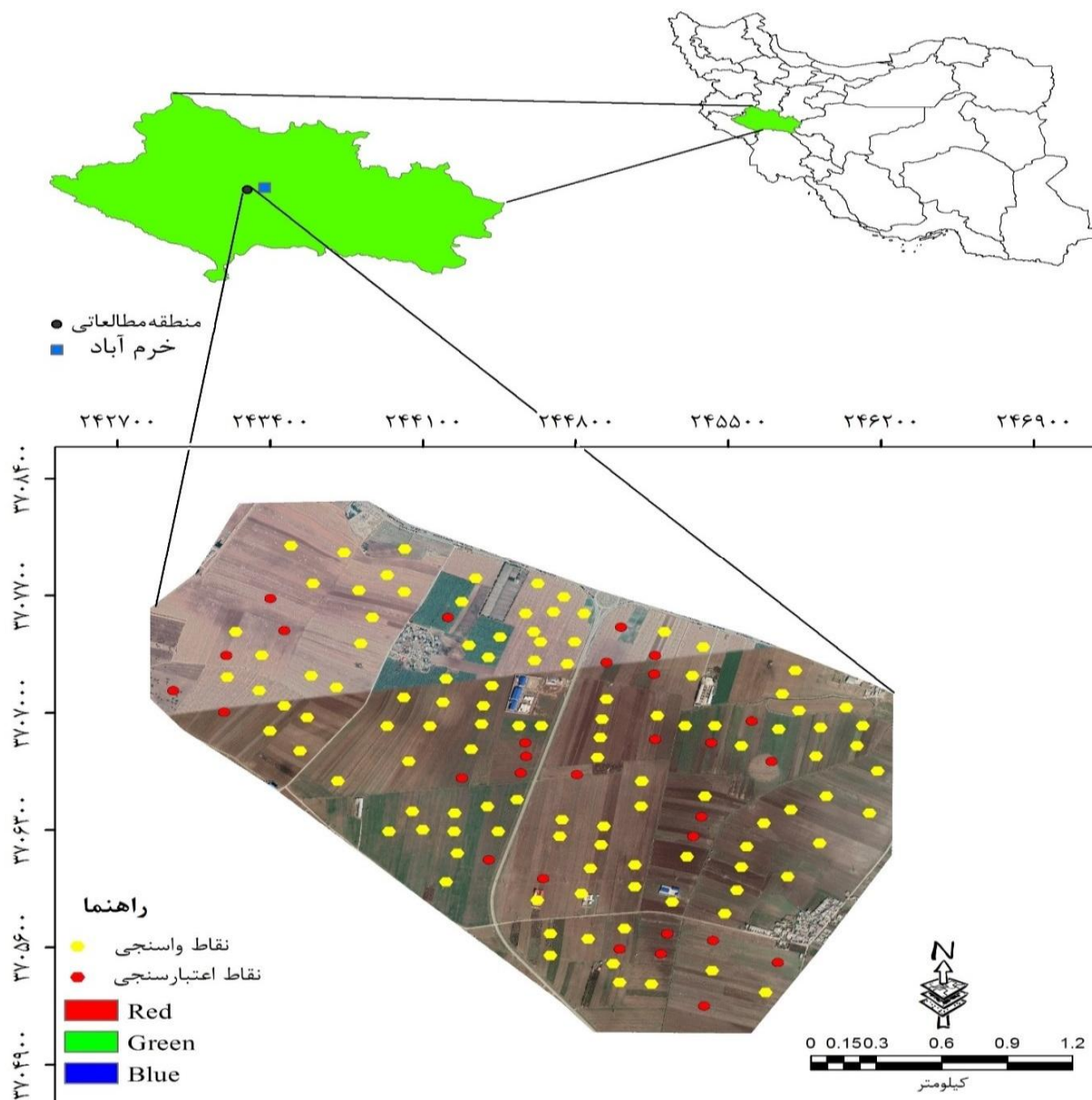
RMSE و MAE به ترتیب برابر ۰/۲۱ و ۰/۱۶ به دست آمد. نتایج نشان داد، نقشه پیش‌بینی شده با واقعیت‌های منطقه مورد مطالعه سازگاری دارد. با توجه به اینکه شناخت عوامل تأثیرگذار بر کربن آلی خاک و درک چگونگی این فرایند بسیار مهم بوده و ارزیابی تغییرات مکانی و ذخایر آن در مقیاس محلی و ملی، گامی مهم در راستای مدیریت پایدار خاک‌ها به ویژه در اراضی کشاورزی است. لذا داشتن اطلاعات کافی از مقدار کربن آلی در خاک و توزیع مکانی آن در مطالعات خاک‌شناسی و زیست‌محیطی اهمیت ویژه‌ای پیدا می‌کند. این پژوهش با هدف بررسی تأثیر برخی خصوصیات خاک بر فرایند مدل‌سازی مکانی کربن آلی خاک سطحی با استفاده از الگوریتم‌های نوین نقشه‌برداری رقومی خاک در بخشی از دشت خرم‌آباد صورت پذیرفته است.

### مواد و روش‌ها

#### معرفی منطقه مورد مطالعه

منطقه مورد مطالعه بخشی از دشت خرم‌آباد با مساحت ۶۸۰ هکتار، حد واسط طول جغرافیایی ۲۴۲۶۰۳ و ۲۴۶۷۰۴ متر و عرض جغرافیایی ۳۷۰۶۸۵۲ و ۳۷۰۷۳۲۹ متر در سیستم مختصات متریک و منطقه ۳۹ شمالی واقع شده است (شکل ۱). متوسط تغییرات ارتفاع ۱۱۳۰ تا ۱۱۷۲ متر از سطح دریا و بر اساس اطلاعات اقلیمی میانگین بارندگی سالیانه آن بر اساس دوره آماری ۳۰ ساله، ۵۱۶ میلی‌متر و حداکثر بارندگی در ماه فروردین ۷۸/۱ میلی‌متر گزارش شده است (ایستگاه هواشناسی استان لرستان- ۱۳۹۷). رژیم رطوبتی و حرارتی خاک‌های منطقه به ترتیب زریک (Xeric) و ترمیک (Thermic) است. از نظر فیزیوگرافی خاک‌های منطقه مورد مطالعه روی واحد دشت‌های دامنه‌ای با حداقل و حداکثر شیب دو و پنج درصد قرار دارند و جهت غالب شیب در منطقه مطالعاتی شمالی- جنوبی است. مواد مادری خاک‌ها بر اساس نقشه زمین‌شناسی منطقه با مقیاس ۱/۱۰۰۰۰۰ (سازمان زمین‌شناسی کشور)، شامل مواد مادری با منشأ آهکی است که مربوط به دوران

دهند. هنگل و همکاران (۱۱)، نقشه‌های جهانی کربن آلی، اسیدینه و اجزای بافت خاک را با بزرگنمایی ۲۵۰ متر و با استفاده از روش‌های رگرسیون خطی (Linear Regression) و الگوریتم یادگیری ماشین درختان تصمیم‌گیری (Random Forest) تصادفی تهیه کردند. نتایج ایشان حاکی از آن بود که روش درختان تصمیم‌گیری تصادفی با ارائه عملکرد بهتر نسبت به روش رگرسیون خطی با میانگین ۷۵-۱۵ درصد کاهش خطا همراه بود. متین‌فر و همکاران (۲۷)، برآورد مقدار ماده آلی خاک را با استفاده از مدل‌های آماری و رگرسیون حداقل مربعات (PLSR) جزئی (Partial Least Square Regression) و رگرسیون مؤلفه اصلی (Principal Component Analysis) انجام دادند. نتایج مدل‌سازی به روش رگرسیون حداقل مربعات جزئی، نشان داد مقادیر  $R^2$ ، RMSE و RPD در مرحله اعتبارسنجی برای پیش‌بینی ماده آلی به ترتیب ۰/۸۵، ۰/۱۴ و ۲/۷۸ بوده است. مدل PLSR برای پیش‌بینی کربن آلی خاک در مناطق خشک و نیمه‌خشک کارایی و دقت بیشتری را نسبت به سایر مدل‌های مورد استفاده ارائه کرد. خاموشی و همکاران (۱۸)، با استفاده از مدل جنگل تصادفی، نقشه رقومی پراکنش کربن آلی را در خاک‌های زراعی دشت قزوین تهیه کردند. نتایج اعتبارسنجی در سه عمق ۲۰-، ۵۰- و ۱۰۰- سانتی‌متری خاک نشان داد که دقت پیش‌بینی مدل‌سازی برای کربن به ترتیب ۰/۷۳، ۰/۷۵ و ۰/۷۵ حاصل شد. ایشان همچنین بیان کردند در صورتی که مدل جنگل تصادفی با متغیرهای محیطی مناسب آموزش داده شود می‌توان آن را به‌عنوان یک الگوریتم یادگیرنده مؤثر در نقشه‌برداری رقومی کربن آلی خاک به‌کار برد. پهلوان‌راد و همکاران (۳۲)، در مطالعه‌ای غلظت کربن آلی خاک را در منطقه زهک در شرق ایران مورد مطالعه قرار دادند. در این مطالعه توزیع کربن آلی خاک با استفاده از نقشه‌برداری رقومی خاک در منطقه‌ای با مساحت ۴۱۰۰۰ هکتار و تعداد ۴۱۷ نمونه خاک که از عمق ۳۰-۰ سانتی‌متری جمع‌آوری شده بود، مورد بررسی قرار گرفت و از رویکرد جنگل تصادفی نیز برای مدل‌سازی استفاده شد. مقدار دو آماره ارزیابی صحت مدل



شکل ۱، موقعیت منطقه مورد مطالعه و موقعیت نقاط نمونه برداری (رنگی در نسخه الکترونیکی)

نقاط با رنگ زرد مربوط به دسته داده‌های مورد استفاده در واسنجی (۸۰٪) و قرمز رنگ برای اعتبارسنجی (۲۰٪)

#### نمونه برداری و تجزیه و تحلیل آزمایشگاهی

عملیات میدانی و نمونه برداری از خاک‌ها شامل انتخاب ۱۴۱ موقعیت نمونه خاک از عمق ۳۰-۵ سانتی متر بر اساس الگوی نمونه برداری تصادفی صورت پذیرفت. نمونه‌ها پس از برداشت ابتدا هواخشک شده و سپس از الک دو میلی متری عبور داده

زمین شناسی کرتاسه تا پلیوسن هستند. رده بندی خاک‌های منطقه بر اساس سامانه رده بندی آمریکایی (۲۰۱۴) و نقشه خاک با مقیاس ۱:۵۰۰۰۰ موجود در دو رده اینسپتی سول و انتی سول قرار دارند. کاربری غالب اراضی منطقه شامل کشاورزی آبی با تیپ‌های بهره‌وری زراعت گندم، ذرت و برنج هستند.

تصادفی و حداقل گره‌های انتهایی است (۳۵). قدرت پیش‌بینی مدل با افزایش قدرت درختان مستقل و کاهش همبستگی بین آنها افزایش می‌یابد (۲۵). مدل‌سازی با روش جنگل تصادفی، ابتدا روی داده‌های آموزشی سپس روی داده‌های اعتبارسنجی اجرا می‌شود. در نهایت مدلی انتخاب می‌شود که دارای کمترین خطای OOB (Out of Bag) باشد. در این مطالعه برای مدل‌سازی جنگل تصادفی از بسته Random Forest (۲۰) در محیط نرم‌افزار RStudio 3.5.0 استفاده شد.

### کو بیست

کو بیست یک الگوریتم پیشرفته رگرسیون درختی است که ساختارهای مختلفی را از داده‌های آموزشی می‌سازد (۱۴). کو بیست یکسری قواعد "اگر - سپس" ایجاد کرده که هر قاعده یک مدل خطی چندمتغیره دارد. قواعد مدل بر اساس روند کاهشی اهمیت، به وسیله الگوریتم کو بیست رتبه‌بندی می‌شود این بدان معنی است که اولین قاعده بیشترین و آخرین آن کمترین مشارکت را در دقت مدل دارا است. تعداد قواعد بر اساس بهترین مدل رگرسیون برازش داده شده، تعیین می‌شود. بسته کو بیست اهمیت متغیرها را به صورت درصد مشارکت بیان و بر اساس متغیرهای مؤثر، متغیر هدف را پیش‌بینی می‌کند. برای کو بیست از بسته Cubist (۱۹) در نرم‌افزار RStudio 3.5.0 استفاده شد.

### رگرسیون خطی چندگانه

با استفاده از این روش می‌توان همزمان به تحلیل و بررسی چندین متغیر مختلف پرداخت. برای به دست آوردن نتایج مطلوب‌تر از طریق رگرسیون خطی چندگانه نمونه‌ها باید دارای فراوانی بالا و همچنین از دقت قابل توجهی برخوردار باشند. زیرا این روش در مقابل اطلاعات نادرست، حساسیت بالایی دارد و ورود چنین داده‌هایی ممکن است منجر به خطاهای بزرگی در نتایج به دست آمده شود. علاوه بر این، برای استفاده از این روش متغیرها باید توزیع نرمال داشته باشند و تغییر آنها از یک رابطه

شدند. ویژگی‌های فیزیکی و شیمیایی منتخب در این تحقیق به ترتیب، توزیع اندازه ذرات خاک (بافت خاک) به روش هیدرومتری، کربنات کلسیم معادل به روش تیتراسیون برگشتی، واکنش خاک (pH) در عصاره گل اشباع، مقدار کربن آلی به روش اکسیداسیون تر و بخش باقی مانده نمونه‌های دست‌خورده برای اندازه‌گیری میانگین وزنی قطر خاکدانه‌ها (MWD) با استفاده از رابطه ۱، محاسبه شد:

$$MWD = \sum_{i=1}^n X_i W_i \quad (1)$$

که در این رابطه  $W_i$ ، جرم خشک خاکدانه‌ها در هر کلاس اندازه برحسب گرم به وزن کل خاک (۵۰ گرم) و  $X_i$ ، میانگین قطر خاکدانه‌های روی الک برحسب میلی‌متر است.

### مدل‌سازی مکانی

در این تحقیق از رویکردهای جنگل تصادفی، کو بیست، رگرسیون خطی چندمتغیره، رگرسیون حداقل مربعات جزئی و کریجینگ معمولی برای مدل‌سازی مکانی کربن آلی خاک استفاده شد.

### جنگل تصادفی

مدل جنگل تصادفی یک روش نوین و قدرتمند در فرایند مدل‌سازی است که پیشرفت قابل ملاحظه‌ای در تکنولوژی داده‌کاوی ارائه داده است. رویکرد جنگل تصادفی مبتنی بر روش‌های جدید ترکیب اطلاعات است که در آن تعداد زیادی درخت تصمیم ایجاد شده است سپس تمام درختان با هم برای پیش‌بینی ترکیب می‌شوند. در این روش برای ایجاد یک درخت رشد، از بهترین متغیرها یا نقاط تقسیم در داخل زیرمجموعه‌های متغیرها استفاده می‌شود که به صورت تصادفی از مجموعه‌ای از متغیرهای ورودی انتخاب شده‌اند و به این شکل خطای کلی مدل کاهش می‌یابد (۴). مدل جنگل تصادفی از سه پارامتر تعریف شده توسط کاربر شامل تعداد متغیرهای مورد استفاده در ساخت هر درخت که قدرت مستقل را بیان می‌کند، تعداد درختان جنگل

روش را نشان می‌دهد:

$$Z_0 = \sum_{i=0}^n W_i Z_i \quad (3)$$

در این رابطه  $Z_0$  برابر با مقادیر برآورد شده،  $W_i$  برابر با وزن و  $Z_i$  برابر با مقادیر نمونه است. وزن‌ها بستگی به درجه همبستگی بین نمونه‌ها و نقاط برآورد شده بستگی دارد و همیشه جمع آنها برابر ۱ است. کریجینگ معمولی یک الگوریتم غیرثابت است که در هر مکان مقدار میانگین را تخمین می‌زند. فرایند اجرای این الگوریتم بیشتر با جستجوی متحرک نقاط همسایه همراه است. درون‌یابی کریجینگ به عنوان روشی بسیار کاربردی تأثیرهای معین بین مشاهدات و متغیرهای کمکی را با استفاده از توابع همبستگی مکانی و تغییرات توابع توضیح می‌دهد (۲۸). برای مدل‌سازی روش کریجینگ از بسته "gstat" در محیط برنامه‌نویسی RStudio 3.5.0 استفاده شد.

#### ارزیابی اعتبار مدل

برای مدل‌سازی کربن آلی خاک، داده‌ها به دو گروه داده‌های واسنجی (۸۰ درصد داده‌ها) و داده‌های اعتبارسنجی (۲۰ درصد داده‌ها) تقسیم شدند. در انتها به منظور ارزیابی دقت در هر یک از روش‌های مدل‌سازی مورد استفاده در برآورد میزان کربن آلی خاک از آماره‌های ضریب تبیین ( $R^2$ )، میانگین مربعات خطا (RMSE) و میانگین خطای استاندارد (MSE) استفاده شد، که معادلات این روش‌ها در روابط ۴، ۵ و ۶ آورده شده است:

اگر انحراف معیار متغیرهای  $X$  و  $Y$  به ترتیب به صورت  $S_x$  و  $S_y$  باشد و کواریانس آنها با علامت  $Cov_{xy}$  نشان داده شود ضریب تبیین از رابطه ۴ قابل محاسبه است:

$$R^2 = \frac{S_{xy}^2}{S_{xx} S_{yy}} \quad (4)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (5)$$

در رابطه ۵،  $n$ : تعداد داده‌ها،  $y_i$ : مقدار اندازه‌گیری شده کربن آلی خاک،  $\hat{y}_i$ : مقدار پیش‌بینی شده کربن آلی خاک است.

خطی پیروی کند. رگرسیون چندگانه در حقیقت ارتباط بین یکسری از متغیرهای پیش‌بینی کننده را با متغیر پاسخ مورد نظر بیان می‌کند. در صورت وجود متغیرهای مستقل  $X_1, X_2, \dots, X_n$  اگر بخواهیم ارتباط خطی بین آنها و متغیر  $Y$  که وابسته به آنهاست ایجاد کنیم رابطه ۲ بین آنها برقرار است:

$$Y = a_0 + a_1 X_1 + a_2 X_2 + \dots + a_n X_n \quad (2)$$

که در این رابطه، از مقادیر  $a_1$  و  $a_2, \dots, a_n$  با عنوان ضرایب رگرسیون یاد می‌شود. این ضرایب، ضرایبی نامشخصی‌اند که در حقیقت مسئول برآورد پارامتر پاسخ هستند (۵).

#### رگرسیون حداقل مربعات جزئی

روش رگرسیون حداقل مربعات جزئی (Partial Least Square Regression)، روش به نسبت جدیدی برای ساخت معادلات رگرسیون است. این روش برای رگرسیون تک‌متغیره و چندمتغیره استفاده می‌شود؛ بنابراین ممکن است چندین متغیر وابسته داشته باشد. برای ایجاد ارتباط بین متغیرهای وابسته و متغیرهای مستقل، کمترین مجذورات جزئی (PLS) متغیرهای تبیینی (مستقل) جدیدی ایجاد می‌کند. برخلاف مدل‌یابی معادلات ساختاری مبتنی بر کواریانس، PLS به جای بازتولید ماتریس کواریانس تجربی، بر بیشترین واریانس تبیین شده متغیرهای وابسته با استفاده از متغیرهای مستقل تمرکز دارد (۳۹). این فن امکان بررسی همزمان روابط متغیرهای پنهان و قابل مشاهده را فراهم می‌آورد و زمانی استفاده می‌شود که حجم نمونه کم و یا توزیع متغیرها نرمال نباشد. برای مدل‌سازی PLSR، مدل نهایی توسط عملگر PLSR (۹) موجود در بسته PLS در نرم‌افزار RStudio 3.5.0 استفاده شد.

#### کریجینگ معمولی

کریجینگ معمولی (Ordinary Kriging) اولین بار توسط مترون در سال ۱۹۷۱ ارائه شد. رایج‌ترین روش درون‌یابی کریجینگ است. در این روش که با استفاده از رابطه ۳ محاسبه می‌شود میانگین ثابت، نامعلوم فرض می‌شود رابطه ۳ متغیرهای این

جدول ۱. آنالیز آماری ویژگی‌های خاک مورد مطالعه

ویژگی خاک (soil character)	واحد (Unit)	انحراف معیار (standard deviation)	حداکثر (maximum)	حداقل (minimum)	میانگین (average)	درصد ضریب تغییرات (Coefficient of variation%)	چولگی (Skewness)	کشیدگی (Kurtosis)
O.C	%	۰/۳۸	۲/۳۴	۰/۰۳	۰/۵۴	۷۰/۳۷	۱/۸۰	۴/۷۴
Clay	%	۷/۸۹	۶۸/۵۶	۲۰/۵۶	۳۷/۸۰	۲۰/۸۷	۰/۸۲	۲/۳۱
Silt	%	۶/۷۲	۴۴/۷۲	۸/۷۲	۲۹/۳۵	۲۲/۸۹	-۰/۴۷	۱/۰۲
Sand	%	۶/۵۴	۴۷/۲۸	۱۰/۰۰	۳۲/۸۴	۱۹/۹۱	-۰/۴۰	۰/۹۲
CaCO <sub>3</sub>	%	۵/۲۲	۲۷/۰۰	۰/۰۶	۷/۷۲	۶۷/۶۱	۰/۹۹	۰/۵۲
MWD	-	۰/۷۳۱	۸/۶۵	۰/۰۴۰	۰/۲۷	۲۷۰/۷۴	۱۱/۱	۱۲۴/۶۳
pH	-	۰/۲۰	۸/۷۲	۷	۷/۸۰	۲/۵۶	۰/۲۱	۴/۴۰

سطحی اکریک هستند که نشان‌دهنده جابه‌جایی آهک و تجمع آن در لایه‌های زیرین خاک است. مجموعه این عوامل باعث تمایز و تفکیک خاک‌ها شده است که به‌طور کلی در دو رده اینسپتی سول و انتی سول قرار دارند.

#### خلاصه آماری داده‌ها

خلاصه آماری داده‌های مشاهداتی برای کربن آلی خاک به‌عنوان متغیر اصلی و سایر ویژگی‌های خاک در جدول ۱ ارائه شده است. میانگین کربن آلی خاک در این منطقه ۰/۵۴ درصد است که نشان‌دهنده پایین بودن میزان کربن آلی خاک است. در عین حال بالا بودن مقدار ضریب تغییرات کربن آلی مؤید تغییرپذیری مکانی زیاد آن در منطقه مورد مطالعه است. دو ویژگی میانگین وزنی قطر خاکدانه‌ها و درصد کربنات کلسیم معادل در کلاس ضریب تغییرات زیاد و ویژگی‌های رس، سیلت، شن و اسیدپت در کلاس ضریب تغییرات کم قرار دارند (۳۸).

#### مقایسه دقت مدل‌های مورد استفاده

نتایج ارزیابی روش‌های جنگل تصادفی، کویست، رگرسیون

$$MSE = \frac{1}{n} \sum_{i=1}^n [Z(x_i) - Z^0(x_i)]^2 \quad (6)$$

در رابطه ۶،  $Z(x_i)$  برابر با مقدار اندازه‌گیری شده متغیر مورد نظر و  $Z^0(x_i)$  برابر با مقدار برآورد شده متغیر مورد نظر است.

#### نتایج و بحث

##### تشریح خاک‌های منطقه مطالعاتی

خاک‌های منطقه مطالعاتی، خاکی‌هایی عمیق به رنگ قهوه‌ای با بافت سنگین (Silt Clay Loam) و دارای ساختمان دانه‌ای (Granular) در افق سطحی هستند که در بعضی از قسمت‌ها با سنگریزه و قلوه‌سنگ که روی لایه‌های به‌رنگ قهوه‌ای تا قهوه‌ای تیره با بافت خیلی سنگین (Clay Loam) و ساختمان مکعبی زاویه‌دار متوسط تا به‌نسبت قوی قرار گرفته است. لایه‌های فوق روی لایه‌هایی به‌رنگ قهوه‌ای مایل به قرمز تا قهوه‌ای متمایل به قهوه‌ای تیره با بافت سنگین تا خیلی سنگین (سیلتی کلی لوم تا سیلتی کلی) و ساختمان مکعبی گوشه‌دار متوسط به‌نسبت قوی که همراه مقدار زیادی آهک است قرار گرفته‌اند. این خاک‌ها دارای افق‌های مشخصه زیرسطحی کمبیک، کلسیک و

جدول ۲. نتایج اعتبارسنجی مدل‌های مورد استفاده برای برآورد کربن آلی سطحی خاک

Model	R <sup>2</sup>	MSE	RMSE
RF	۰/۷۵	۰/۳۸	۰/۲۴
Cubist	۰/۷۰	۰/۳۷	۰/۲۸
PLSR	۰/۱۱	۰/۱۳	۰/۴۲
MLR	۰/۰۱	۰/۳۰	۰/۴۴
OK	۰/۰۰۰۷	۰/۵۱	۰/۵۵

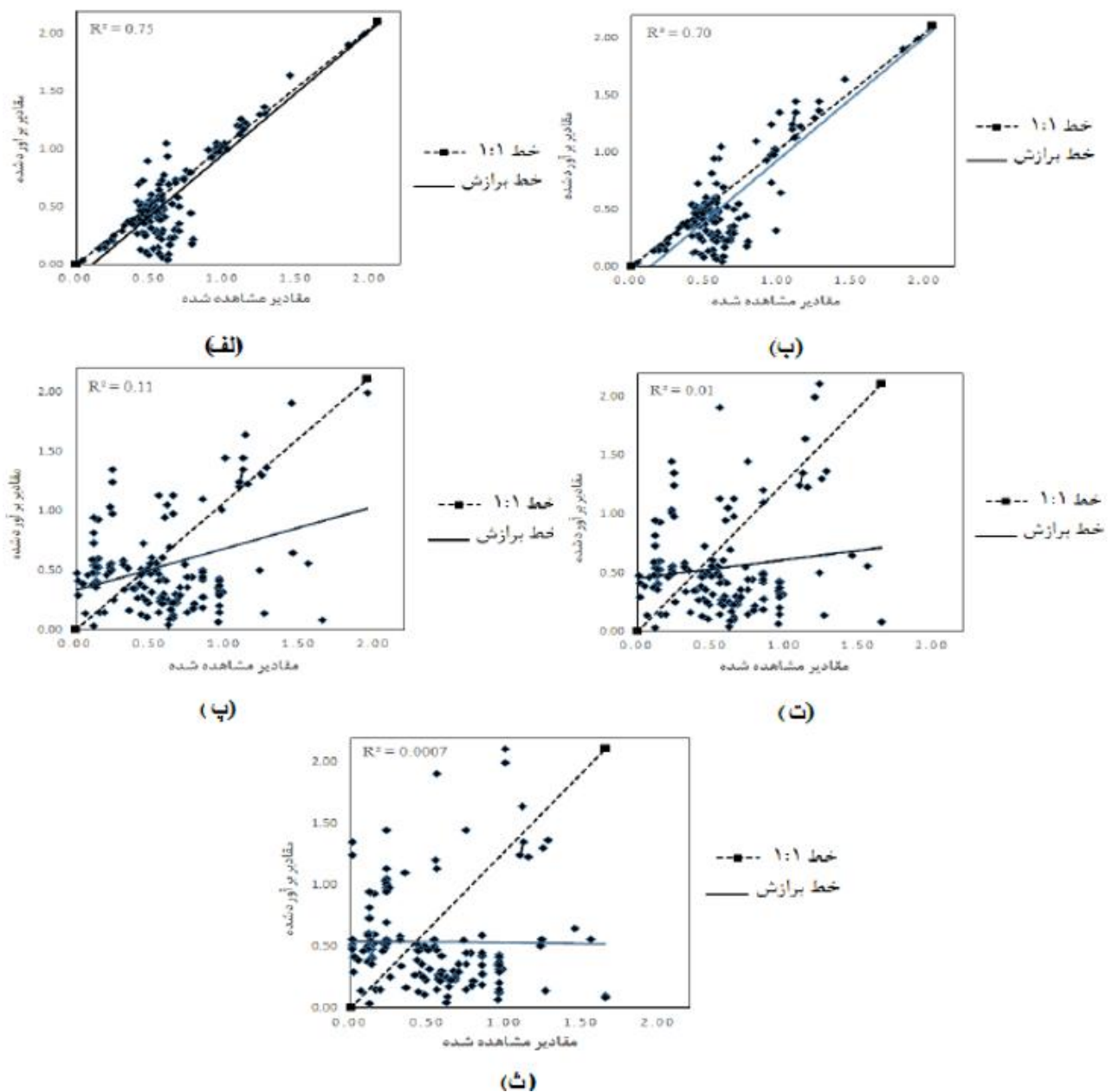
جدول ۳. مساحت و درصد کلاس‌های کربن آلی خاک بر اساس روش RF

Class	Area (ha)	Percentage (%)
۰ - ۰/۵	۴۶/۱۴	۶/۶۸
۰/۵ - ۱	۱۵۹/۰۳	۲۳/۳۵
۱ - ۱/۵	۲۱۸/۰۶	۳۲/۰۶
۱/۵ - ۲	۲۵۶/۷۷	۳۷/۷۳

می‌دهد. با توجه به شناخت از منطقه و آگاهی از تغییرات مقادیر کربن آلی خاک، مدل جنگل تصادفی و کویست نسبت به سایر مدل‌های مورد استفاده دارای بالاترین میزان دقت و حداقل خطا برای تخمین کربن آلی خاک بودند. همان‌طور که نتایج نشان می‌دهد مقدار R<sup>2</sup> به دست آمده از روش جنگل تصادفی و مدل کویست برای تخمین کربن آلی خاک در رده خوب است. راسل و مک برتنی (۲۹) نیز معتقدند که ضریب تبیین  $\geq 0/81$  خیلی خوب،  $0/61 - 0/80$  خوب،  $0/41 - 0/60$  متوسط و  $\leq 40/0$  ضعیف است. همچنین طبق تقسیم‌بندی راسل و مک برتنی (۲۹) نتایج به دست آمده از مدل رگرسیون حداقل مربعات جزئی، رگرسیون خطی چندمتغیره و کریجینگ معمولی کارایی ضعیفی برای پیش‌بینی کربن آلی ذخیره خاک را نشان دادند (جدول ۲). همان‌گونه که پیش‌تر گفته شد با توجه به اینکه ضریب تغییرات کربن آلی خاک در منطقه مورد مطالعه در دامنه تغییرات زیاد قرار می‌گیرد و این نشان‌دهنده غیریکنواختی شدید این ویژگی در منطقه است، لذا

خطی چندمتغیره، رگرسیون حداقل مربعات جزئی و کریجینگ معمولی در تخمین کربن آلی خاک در اراضی مورد مطالعه در جدول ۲ آورده شده است، و برای شناسایی دقیق‌ترین روش، اساس قضاوت کمترین RMSE و بیشترین R<sup>2</sup> مدنظر قرار گرفت. طبق نتایج به دست آمده از جدول ۲، بهترین روش مدل‌سازی برای نقشه‌برداری رقومی کربن آلی خاک، روش جنگل تصادفی با حداقل RMSE برابر ۰/۲۴ و بیشترین R<sup>2</sup> برابر با ۰/۷۵ به دست آمد. نتایج پیش‌بینی مکانی کربن آلی خاک نشان داد که بیشترین مساحت و درصد پراکنش کلاس‌های کربن آلی خاک سطحی مربوط به کلاس شماره ۴ با مساحت ۲۵۶/۷۷ هکتار است که ۳۷/۷۳ درصد از کل منطقه مطالعاتی را شامل می‌شود و کمترین مساحت و درصد پراکنش مربوط به کلاس شماره ۱ با حدود ۴۶/۱۴ هکتار است که تقریباً ۶/۶۸ درصد از کل مساحت منطقه را شامل می‌شود (جدول ۳). شکل ۲ نمودار همبستگی بین مقادیر اندازه‌گیری شده و مقادیر برآورد شده کربن آلی خاک را برای مدل‌های مورد استفاده نشان





شکل ۲. نمودار همبستگی بین مقادیر اندازه‌گیری شده و برآورد شده در مدل‌های مورد استفاده، به ترتیب الف) جنگل تصادفی، ب) کوپست، پ) رگرسیون حداقل مربعات جزئی، ت) رگرسیون خطی چندمتغیره و ث) رگرسیون کریجینگ

در نسبت به روش رگرسیون خطی نشان دادند و بیان داشتند که برتری مدل جنگل تصادفی نسبت به مدل‌های خطی در پیش‌بینی کربن آلی خاک مربوط به روابط غیرخطی است که بین کربن آلی خاک و متغیرهای پیش‌بینی کننده وجود دارد. هنگل و همکاران (۱۳)، در مطالعه خود عملکرد دو روش جنگل تصادفی و رگرسیون خطی را برای نقشه‌برداری

این نتایج دور از انتظار نیست. تحقیقات و پژوهش‌هایی که پیش‌تر روی تغییرات و مقدار کربن آلی خاک‌ها انجام شده است به دلیل تغییرات مکانی بسیار زیاد این متغیر همواره با محدودیت‌هایی روبه‌رو بوده است (۲۶). ژانگ و همکاران (۴۲) در خصوص پیش‌بینی کربن آلی خاک به دو روش جنگل تصادفی و رگرسیون خطی نیز برتری روش جنگل تصادفی را

خاکورزی دانست. با توجه به اینکه کربن آلی خاک در بیشتر عرصه‌های زراعی بسیار کم است تغییرات جزئی در مقدار آن می‌تواند تأثیرات عمیقی بر کارکرد خاک و در نتیجه کیفیت خاک داشته باشد. بر همین اساس مدیریت خاک‌ها در اراضی زراعی نقش بسیار مهمی در توزیع کربن آلی آنها دارد. در همین راستا، پرویزی و همکاران (۳۴)، اثرات مدیریت زراعی و مدیریت خاک را بر میزان کربن آلی خاک در حوزه مرک استان کرمانشاه مورد بررسی قرار دادند. نتایج نشان داد که در کاربری زراعی، الگوی خاکورزی و نظام مدیریت بقایا، عوامل اصلی تغییرات میزان کربن آلی خاک است و با اصلاح سامانه خاکورزی و مدیریت بقایا در مناطق خشک می‌توان در هر هکتار حدود ۳۵ تن کربن آلی در خاک ذخیره کرد. در قسمت‌های شمالی منطقه مورد مطالعه نیز مقدار کربن آلی بیشتری نسبت به نواحی جنوبی وجود دارد. یکی از دلایل این موضوع را می‌توان به مقدار رطوبت بیشتر، پوشش گیاهی مناسب‌تر نسبت به نواحی جنوبی و نیز به‌کارگیری رویکرد کشاورزی حفاظتی اشاره کرد. در همین راستا، بنی و همکاران (۱) در مطالعه خود به این نتایج دست یافتند که در شیب‌های شمالی مقدار کربن آلی بیشتر از قسمت‌های جنوبی است که علت این اختلاف می‌تواند در میزان رطوبت، تراکم پوشش گیاهی و بقایای بازگشتی به خاک باشد.

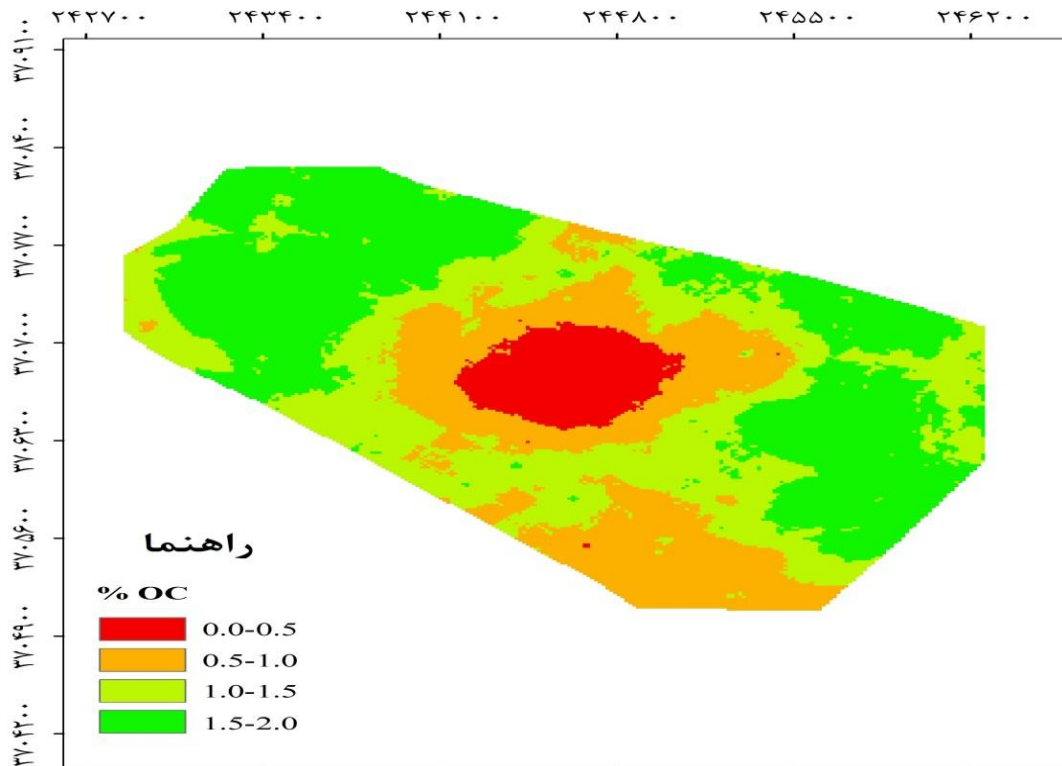
#### اهمیت نسبی ویژگی‌های خاک

مطالعه فاکتورهای مؤثر بر میزان کربن آلی خاک در هر منطقه می‌تواند در پیش‌بینی تغییرات آن تأثیر بسزایی داشته باشد. بررسی مقدار کربن آلی خاک و عوامل مؤثر بر آن در اقلیم‌های خشک و نیمه‌خشک لازمه مدیریت مناسب کربن آلی خاک در مقیاس جهانی است، این مسئله در ایران از اهمیت ویژه‌ای برخوردار است، زیرا بیش از ۸۲ درصد از مساحت کشور در زمره مناطق خشک و نیمه‌خشک جهان قرار گرفته است (۱۶). در تحقیق حاضر ارتباط میان ذخیره کربن آلی خاک با ویژگی‌های خاک مورد بررسی قرار گرفت. شکل ۳ نشان

خصوصیات خاک بررسی کردند، نتایج حاصل از اعتبارسنجی نشان داد که الگوریتم جنگل تصادفی به‌طور مداوم از الگوریتم رگرسیون خطی با کاهش ۷۵-۱۵ درصدی خطای جذر میانگین مربعات (RMSE) در خواص و عمق خاک، بهتر عمل می‌کند. زراعت پیشه و همکاران (۴۱)، برای ویژگی‌های کربن آلی، رس و آهک مقادیر ضرایب تبیین ۰/۵۵، ۰/۱۵ و ۰/۳۰ با ریشه میانگین خطای مربعات ۰/۳۳، ۷/۸۶ و ۹/۵۲ درصد در استفاده از مدل جنگل تصادفی به‌عنوان مدل مناسب‌تر در مقایسه با مدل کوپست، رگرسیون درختی و رگرسیون خطی چندمتغیره گزارش کردند. همچنین بر اساس نتایج به‌دست آمده از روش کریجینگ معمولی شواهد نشان می‌دهد که این روش به‌تنهایی برای برآورد تغییرات مکانی کربن آلی در خاک‌های منطقه مورد مطالعه مناسب نیست و برای توصیف دقیق‌تر این تغییرات بایستی از متغیرهای کمکی نیز بهره گرفت. این مسئله می‌تواند ناشی از تراکم پایین نقاط نمونه‌برداری باشد. به‌عبارتی برای حصول نتایج مطلوب‌تر در روش کریجینگ بایستی تراکم نقاط نمونه‌برداری بالا باشد (۲۴).

#### آنالیز مکانی کربن آلی خاک

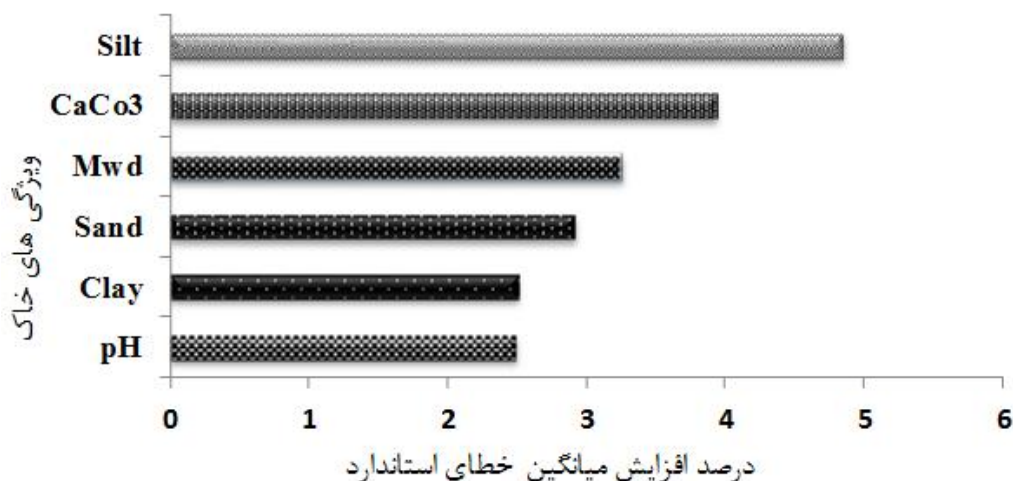
نقشه توزیع مکانی حاصل از برآورد کربن آلی خاک به‌دست آمده از منطقه مورد مطالعه به‌روش جنگل تصادفی در شکل ۳ آورده شده است. در شکل ۳، کلاس شماره ۴ با میزان کربن آلی بین ۲-۱ درصد، بیشترین میزان کربن آلی و کلاس شماره ۱ با مقدار کربن آلی ۰/۵-۰ درصد، پایین‌ترین میزان کربن آلی خاک را به‌خود اختصاص دادند. با توجه به شکل ۳، در کلاس شماره ۱ نقشه توزیع مکانی، میزان کربن آلی خاک بسیار کم است. در اراضی کشاورزی برداشت محصول موجب کاهش ورود بقایای گیاهی تازه به خاک و به‌دنبال آن کاهش ذخایر کربن آلی و مواد غذایی مورد استفاده ریزجانداران خاک می‌شود. همچنین کم بودن مقدار کربن آلی در این محدوده را می‌توان به‌دلیل برداشت کاه و کلش از مزارع، سوزاندن بقایای گیاهی، عدم رعایت تناوب زراعی و مدیریت



شکل ۳. نقشه توزیع مکانی کربن آلی خاک به‌روش جنگل تصادفی

داشتند میزان ذخیره کربن آلی در خاک با میزان رس و سیلت خاک رابطه مثبت بسیار قوی دارد اما اقدامات مدیریتی نامناسب نظیر شخم می‌تواند این رابطه را به‌طور معنی‌داری تحت تأثیر قرار دهد. با این حال نتایج نشان داد که مقدار رس که انتظار می‌رفت در مدل پیش‌بینی کننده جنگل تصادفی، تغییرات کربن را به‌خوبی رصد کند، تأثیر کمتری در پیش‌بینی کربن آلی خاک داشت که می‌تواند نشان‌دهنده پیچیدگی پراکنش این متغیر در خاک‌های منطقه مطالعاتی است. مطالعاتی که روی تخمین مقدار و تغییرات کربن آلی خاک‌ها انجام گرفته است، به‌دلیل تغییرات مکانی بسیار زیاد این متغیر، همواره با محدودیت‌هایی مواجه بوده است (۲۶). در بیان تأثیر میزان آهک بر میزان کربن آلی خاک همان‌طور که نتایج نشان می‌دهد بعد از مقدار سیلت، میزان آهک مهم‌ترین متغیر مؤثر در پیش‌بینی مکانی میزان کربن آلی خاک در نظر گرفته شد. نتایج حاصل از این بررسی با نتایج حاصل از مطالعه پرویزی (۳۳)، که نشان دادند آهک مهم‌ترین عامل تعیین‌کننده مقدار کربن آلی

می‌دهد کمترین مقدار کربن آلی خاک در میانه دشت توزیع شده است، که به‌نظر با وجود شیب کمتر در این ناحیه و دریافت رطوبت بیشتر نسبت به مزارع حاشیه، عملیات مدیریتی و کشت متراکم نقش عمده‌ای در کاهش کربن آلی در این ناحیه داشته است. در این مطالعه نتایج مدل جنگل تصادفی با استفاده از متغیرهای کمکی رس، سیلت، شن، آهک، میانگین وزنی قطر خاکدانه و اسیدیته به‌دست آمد (شکل ۴). در این مدل، متغیر کمکی سیلت، بیشترین تأثیر را روی پیش‌بینی کربن آلی خاک داشت. جانسون و همکاران (۱۵)، در مطالعه خود به این نتیجه رسیدند که بافت خاک دارای بیشترین عامل اثرگذار در برآورد کربن آلی خاک است. کرکیج و همکاران (۳۷)، در مطالعه خود، ارتباط ذخیره کربن آلی با برخی ویژگی‌های خاک در مراتع آذربایجان شرقی را بررسی کردند. نتایج آنها نشان داد ذرات ریز سیلت و رس با ضریب تأثیر به‌ترتیب ۰/۲۲۳ و ۰/۱۷۵ ظرفیت و پتانسیل بالایی در ذخیره کربن آلی خاک را دارا هستند. هاسینک و همکاران (۱۲)، نیز در تأکید این مطلب بیان



شکل ۴. اهمیت نسبی متغیرهای خاک در پیش‌بینی کربن آلی خاک با استفاده از جنگل تصادفی

کربن انتظار می‌رفت در این مطالعه با وارد کردن خصوصیات خاک به تنهایی در مدل‌سازی قادر به تبیین بخشی از تغییرات کربن خاک باشیم. نتایج این تحقیق همچنین مؤید این فرض است که با وارد کردن پارامترهای محیطی (نظیر عوامل توپوگرافیکی، کاربری اراضی، فرسایشی و اقلیمی، داده‌های سنجش از دور) در فرایند مدل‌سازی میزان ضریب تبیین ( $R^2$ ) می‌تواند بهبود یابد به این علت که منظور کردن پارامترهای محیطی در این نوع مطالعات باعث می‌شود تا زمینه ارزیابی دقیق اثرات تغییرات عمده کربن آلی در منطقه فراهم شود. در همین راستا، میناسنی و مک برتنی (۲۹) نقشه رقوم ذخیره کربن آلی خاک را با استفاده از متغیرهای توپوگرافی شامل شیب، جهت شیب و شاخص رطوبت و به‌کارگیری مدل رگرسیون خطی، جنگل تصادفی و شبکه عصبی مصنوعی در دو منطقه اندونزی تهیه کردند و ضرایب تبیین را به ترتیب ۰/۹۹، ۰/۹۵ و ۰/۹۸ گزارش کردند. خورده‌بین و همکاران (۲۱) در پیش‌بینی ذخیره کربن آلی خاک در بخشی از اراضی شهرستان بهبهان از متغیرهای کمکی از جمله اجزای سرزمین، داده‌های حاصل از سنجش از دور و نقشه کاربری اراضی استفاده کردند. نتایج آنها نشان داد که مدل جنگل تصادفی با بالاترین ضریب تبیین ( $R^2 = 0.96$ ) و کمترین مجذور میانگین مربعات خطا ( $RMSE = 2.32$ ) کارایی بهتری در پیش‌بینی

خاک است و بیان کردند که مقدار آهک از طریق اثرگذاری بر بیوماس میکروبی و در نتیجه فرایندهای بیولوژیک و با کاهش روند تجزیه بقایای گیاهی بر چرخه کربن خاک اثر گذاشته و در نهایت کربن آلی بومی خاک را کاهش می‌دهد، همسو بود. همچنین نتایج نشان می‌دهد که میانگین وزنی قطر خاکدانه در بیان تغییرات کربن آلی خاک بعد از آهک نقش مؤثرتری داشته است. با توجه به اینکه منطقه مورد مطالعه اراضی تحت کشت و کار است می‌توان مدیریت اراضی را یک فاکتور مهم در بیان تغییرپذیری بین کربن آلی و پایداری خاکدانه در نظر گرفت. مطالعات بالادنت و همکاران (۶)، نیز در همین راستا بیان کردند که مدیریت اراضی به‌عنوان یک عامل مهم انسانی مؤثر بر تغییرات کربن آلی و پایداری خاکدانه‌ها شناخته می‌شود، همچنین عملیات کشت و کار صرف‌نظر از تأثیر منفی بر مقدار کربن آلی خاک، توزیع اندازه و پایداری خاکدانه‌ها را کاهش می‌دهد. چیلک (۷) نشان داد که کشت و کار موجب کاهش ۶۴ و ۶۲ درصدی میانگین وزنی قطر خاکدانه‌ها در خاک‌های ترکیه شده است. کاوه و همکاران (۱۷) نیز نقش ویژگی‌های خاک به همراه پارامترهای توپوگرافی و اقلیم را بر ذخیره کربن مورد بررسی قرار دادند و نشان دادند که درصد آهک، سیلت و درصد اشباع خاک از مهم‌ترین مشخصه‌های اثرگذار بر ذخیره کربن آلی خاک است. با توجه به اثرگذاری ویژگی‌های خاک بر میزان

ذخیره کربن آلی خاک را دارد.

عملکرد مناسب‌تر مدل جنگل تصادفی و پس از آن مدل کوپست نسبت به سایر مدل‌های مورد استفاده در تخمین کربن آلی خاک سطحی بود. مدل‌های رگرسیون حداقل مربعات جزئی، رگرسیون خطی چندگانه و کریجینگ معمولی به‌خوبی نتوانستند تغییرات کربن آلی خاک سطحی را در منطقه مورد مطالعه توضیح دهند که با توجه به پیچیدگی روابط بین تغییرات کربن آلی خاک سطحی با عوامل کنترل‌کننده محیطی آن و همچنین به دلیل ذات غیرخطی بودن این مدل‌ها، این نتایج دور از انتظار نیست. بنابراین پیشنهاد می‌شود برای تهیه نقشه کربن آلی خاک سطحی در منطقه مورد مطالعه و سایر دشت‌های کشاورزی ایران با شرایط مشابه به بررسی تأثیرات نقش سایر فاکتورهای محیطی پرداخته شود.

## نتیجه‌گیری کلی

هدف از این تحقیق، بررسی تأثیر برخی ویژگی‌های خاک از جمله سیلت، شن، رس، آهک، میانگین وزنی قطر خاکدانه و اسیدیته بر میزان کربن آلی خاک و تعیین مهم‌ترین آنها است. بر همین اساس برآورد کربن آلی خاک سطحی با استفاده از مدل‌های جنگل تصادفی، کوپست، رگرسیون حداقل مربعات جزئی، رگرسیون خطی چندمتغیره و کریجینگ معمولی مورد ارزیابی قرار گرفت. نتایج به‌دست آمده نشان داد که مقدار سیلت، آهک و میانگین وزنی قطر خاکدانه‌ها بیش از سایر ویژگی‌های خاکی مورد بررسی، میزان تغییرات کربن آلی خاک را توجیه کردند. از سوی دیگر نتایج حاصل از مدل‌سازی بیانگر

## منابع مورد استفاده

- Benny, L. A. and P. R. Stephens. 1985. The Feasibility of determining the influence of arable land management on topsoil depth. *Publication Soil Conservation Centre Aokautere* 7: 0111-7971.
- Brahim, N., D. Blavet, T. Gallali and M. Bernoux. 2011. Application of structural equation modeling for assessing relationships between organic carbon and soil properties in semiarid Mediterranean region. *International Journal of Environmental Science & Technology* 8(2): 305-320.
- Bernoux, M., M. da Conceição Santana Carvalho, B. Volkoff and C. C. Cerri. 2002. Brazil's soil carbon stocks. *Soil Science Society of America Journal* 66(3): 888-896.
- Breiman, L. 2001. Random forests. *Machine Learning* 45(1): 5-32.
- Balan, B., S. Mohaghegh and S. Ameri. 1995. State- of- Art- in permeability determination from well log data: Part 1- A comparative study, model development. *Society of Petroleum Engineers* 30978: 17-25.
- Balesdent, J., C. Chenu and M. Balabane. 2000. Relationships of soil organic matter dynamics to physical protection and tillage. *Soil and Tillage* 53: 215-230.
- Celik, I. 2005. Land use effects on organic matter and physical properties of soil in a southern Mediterranean highland of Turkey. *Soil and Tillage* 83: 270-277.
- Cambou, A., R. Cardinael, E. Kouakoua, M. Villeneuve, C. Durand and B. G. Barthès. 2015. Prediction of soil organic carbon stock using visible and near infrared reflectance spectroscopy (VNIRS) in the field. *Geoderma* 261: 151-159.
- Clyde, M. 2017. BAS: Bayesian Adaptive Sampling for Bayesian Model Averaging. R Package Version 1.4.6. Available online at: <https://CRAN.R-project.org/web/packages/BAS>. Accessed 2017.
- Grunwald, S. 2009. Multi-criteria characterization of recent digital soil mapping and modeling approaches. *Geoderma* 152(3): 195-207.
- Hengl, T., G. B. Heuvelink, B. Kempen, J. G. Leenaars, M. G. Walsh, K. D. Shepherd and J. E. Tondoh. 2015. Mapping soil properties of Africa at 250 m resolution: Random forests significantly improve current predictions. *Plos One* 10(6): e0125814.
- Hassink, J. and A. P. Whitmore. 1997. A model of the physical protection of organic matter in soils. *Soil Science Society of America Journal* 61(1): 131-139.
- Hengl, T., J. M. De Jesus, G. B. M. Heuvelink, M. R. Gonzalez, M. Kilibarda, A. Blagotić, W. Shangguan, M. N. Wright, X. Geng, B. Bauer-Marschallinger, M. A. Guevara, R. Vargas, R. A. MacMillan, N. H. Batjes, J. G. B.

- Leenaars, E. Ribeiro, I. Wheeler, S. Mantel and B. Kempen. 2017. SoilGrids250 m: Global gridded soil information based on machine learning. *PLoS One* 12(2): e0169748.
14. Holmes, G., M. Hall and E. Frank. 1999. Generating rule sets from model trees. PP. 1-12. In: N. Foo, (Eds.), *Advanced Topics in Artificial Intelligence, Lecture Notes in Artificial Intelligence*.
15. Wight, J. P., A. J. Ashworth and F. L. Allen. 2016. Organic substrate, clay type, texture, and water influence on NIR carbon measurements. *Geoderma* 261: 36-43.
16. Khoshbakht, K. 2011. Country Report: Iran. Workshop on Climate Change and its Impact on Agriculture. Seoul, Korea. Available at: <http://www.adbi.org>.
17. Kaveh, A., M. H. Mahdian, Y. Parvizi, R. Sokouti Oskouei and M. H. Masihabadi. 2015. Investigating effects of topography, soil and climate factors on soil organic carbon storage in drylands of Kermanshah Province. *Desert Management* 4: 51-65. (In Farsi).
18. Khamoshi, S. E., F. Sarmadian, S. R. Mousavi and A. Rahmani. 2019. Organic soil carbon mapping using random forest model. In: 16<sup>th</sup> Iranian Soil Science Congress, University of Zanjan, Iran, 1-6. (In Farsi).
19. Kuhn, M., S. Weston, C. Keefer, N. Coulter and R. Quinlan. 2008. Rulequest Research Pty Ltd. Rule- and Instance-Based Regression Modeling. Available online at: <https://topepo.github.io/Cubist> 2018. accessed 2018.
20. Kuhn, M. 2015. Caret: Classification and Regression Training. Astrophysics Source Code Library.
21. Khordehbin, S., S. Hojatia, A. Landi and I. Ahmadianfar. 2020. Comparison of different data mining methods in predicting soil organic carbon storage in some lands of Behbahan city. *Iranian Journal of Soil and Water Research* 4(51): 1041-1054. (In Farsi).
22. Lamichhane, S., L. Kumar and B. Wilson. 2019. Digital soil mapping algorithms and covariates for soil organic carbon mapping and their implications: A review. *Geoderma* 352: 395-413.
23. Lal, R. 2004. Soil carbon sequestration impacts on global climate change and food security. *Science* 304(5677): 1623-1627.
24. Liao, K., S. Xu, J. Wu and Q. Zhu. 2013. Spatial estimation of surface soil texture using remote sensing data. *Soil Science and Plant Nutrition* 59(4): 488-500.
25. Ließ, M., B. Glaser and B. Huwe. 2012. Uncertainty in the spatial prediction of soil texture comparison of regression tree and random forest models. *Geoderma* 170: 70-79.
26. Minasny, B., A. B. McBratney, B. P. Malone and I. Wheeler. 2010. Digital mapping of soil carbon. 19th World Congress of Soil Science, Brisbane, Australia.
27. Matinfar, H. R., H. Mahmoodzadeh and A. Fariabi. 2018. Estimation soil organic matter (SOM) content using visible and near infrared spectral data, PLSR and PCR statistical models. *Iranian Remote Sensing & GIS* 10(2): 15-32. (In Farsi).
28. Marchetti, A., C. Piccini, R. Francaviglia and L. Mabit. 2012. Mabit, Spatial distribution of soil organic matter using geostatistics: A key indicator to assess soil degradation status in central Italy. *Pedosphere* 22: 230-242.
29. Minasny, B. and A. B. McBratney. 2016. Digital soil mapping: a brief history and some lessons. *Geoderma* 264: 301-311.
30. Rossel, R. V. and A. B. McBratney. 2008. Diffuse reflectance spectroscopy as a tool for digital soil mapping. In: 165-172. In: Hartemink, A. (Eds.), *Digital Soil Mapping with Limited Data*, Springer.
31. Pahlavan-Rad, M. R. and A. Akbarimoghaddam. 2018. Spatial variability of soil texture fractions and pH in a flood plain (case study from eastern Iran). *Catena* 160: 275-281.
32. Pahlavan-Rad, M. R., K. Dahmardeh and C. Brungard. 2018. Predicting soil organic carbon concentrations in a low relief landscape, eastern Iran. *Geoderma Regional* 15: e00195.
33. Parvizi, Y., M. Gorji, M. H. Mahdian and M. Omid. 2010. Sensitivity analysis for determining priority of factors controlling SOC condition of west of Iran. *International Journal of Environmental and Earth Sciences* 47: 922-926.
34. Parvizi, Y. and M. Gorji. 2014. Influence of rainfed management factors on soil organic carbon content in Kermanshah merc. *Journal of Land Management* 1: 82-89. (In Farsi).
35. Pham, T. G., M. Kappas, C. V. Huynh and H. Nguyen. 2019. Application of ordinary kriging and regression kriging method for soil properties mapping in hilly region of central Vietnam. *International Journal of Geo-Information* 8(3): 147.
36. Somarathna, P. D. S. N., B. Minasny and B. P. Malone. 2017. More data or a better model? Figuring out what matters most for the spatial prediction of soil carbon. *Soil Science Society of America Journal* 81(6): 1413-1426.
37. Sheyday Karkaj, A., E. Sepehri, H. Barani and J. Motamedi. 2017. The relationship between soil organic carbon storage and some soil properties in East Azarbaijan rangelands. *Rangeland Scientific Journal* 2: 125-138. (In Farsi).
38. Wilding, L. P. 1985. Spatial variability: its documentation, accomodation and implication to soil surveys. *Soil Spatial Variability* 1984: 166-194.
39. Wang, J., Z. Li, X. Qin, X. Yang, Z. Gao and Q. Qin. 2014. Hyperspectral predicting model of soil salinity in Tianjin costal area using partial least square regression. *Geoscience and Remote Sensing Symposium* 3251-3254.

40. Yousaf, B., G. Liu, R. Wang, Q. Abbas, M. Imtiaz and R. Liu. 2017. Investigating the biochar effects on C-mineralization and sequestration of carbon in soil compared with conventional amendments using the stable isotope ( $\delta^{13}\text{C}$ ) approach. *Gcb Bioenergy* 29: 1085-1099.
41. Zeraatpisheh, M., S. Ayoubi, A. Jafari, S. Tajik and P. Finke. 2019. Digital mapping of soil properties using multiple machine learning in a semi-arid region, central Iran. *Geoderma* 338: 445-452.
42. Zhang, H., P. Wu, A. Yin, X. Yang, M. Zhang and C. Gao. 2017. Prediction of soil organic carbon in an intensively managed reclamation zone of eastern China: A comparison of multiple linear regressions and the random forest model. *Science of the Total Environment* 592: 704-713.

## Evaluation of Machine Learning Methods in Digital Mapping of Soil Organic Carbon (part of Khorramabad Plain)

H. R. Matinfar<sup>2\*</sup>, Z. Maghsodi<sup>1</sup> and S. R. Mousavi<sup>2</sup> and M. Jalali<sup>1</sup>

(Received: April 27-2020 ; Accepted: November 24-2020)

### Abstract

Knowledge about the spatial distribution of soil organic carbon (SOC) is one of the practical tools in determining sustainable land management strategies. During the last two decades, the utilization of data mining approaches in spatial modeling of SOC using machine learning algorithms have been widely taken into consideration. The essential step in applying these methods is to determine the environmental predictors of SOC optimally. This research was carried out for modeling and digital mapping of surface SOC aided by soil properties i.e., silt, clay, sand, calcium carbonate equivalent percentage, mean weight diameter (MWD) of aggregate, and pH by machine learning methods. In order to evaluate the accuracy of random forest (RF), cubist, partial least squares regression, multivariate linear regression, and ordinary kriging models for predicting surface SOC in 141 selected samples from 0-30 cm in 680 hectares of agricultural land in Khorramabad plain. The sensitivity analysis showed that silt (%), calcium carbonate equivalent, and MWD are the most important driving factors on spatial variability of SOC, respectively. Also, the comparison of different SOC prediction models, demonstrated that the RF model with a coefficient of determination ( $R^2$ ) and root mean square error (RMSE) of 0.75 and 0.25%, respectively, had the best performance rather than other models in the study area. Generally, nonlinear models rather than linear ones showed higher accuracy in modeling the spatial variability of SOC.

**Keywords:** Spatial variability, Digital soil mapping, Modeling approaches, SOC prediction.

---

1. Soil Science and Engineering Department, Lorestan University, Khorramabad, Iran.

2. Science and Soil Engineering Department, Faculty of Agricultural Engineering and Technology, College of Agriculture & Natural Resources, Tehran University, Kraj, Iran.

\*: Corresponding author, Email: matinfar.h@lu.ac.ir