

Analysis of Different Sources of Uncertainty in Digital Mapping of some Soil Properties (Case Study of Ravansar plain, Kermanshah Province)

Shahrokh Fatehi^{1*}, Yahya Parvizi², Mir Naser Navidi³

(Received: July, 2024 Accepted: December, 2024)

Abstract

Digital mapping of sand, silt and clay is used in the design of irrigation systems, land evaluation and soil erosion models. In contrast, the quantification of the spatial uncertainty of the mentioned soil properties provides users with a criterion for the reliability of these maps. For this purpose, in this research, two sources of uncertainty, including the uncertainty of the variation of the available data and the uncertainty of the model, were investigated using the standard deviation criterion. This study was conducted on an area of 57000 hectares in the Ravansar plain of Kermanshah province. 120 observation points and derivatives of DEM and vegetation and parent material indices obtained from Landsat 8 images were used as model inputs. To estimate the uncertainty caused by the variability of the available data, the random forest model was run 50 times based on a random split of the data into two categories: training (25%) and validation (75%). The Quantile regression forest model was used to predict the soil properties and to estimate the model uncertainty. The results of validating the predictions of the random forest model showed that the value of R² was 0.47, 0.38, and 0.27 for sand, silt and clay, respectively, and the RMSE value was determined for sand (8.61), silt (8.25), and clay (8.1). These results show that the predictor variables explain the changes in sand well, but the ability of the random forest model to predict sand is lower than that of the silt, and clay variables. The results also show that the uncertainty of the variations in the available data accounts for about 1/4 of the total uncertainty and the uncertainty of the model accounts for about 3/4 of the total uncertainty. Since the model uncertainty accounts for a larger proportion of the total uncertainty, the machine learning models need to be carefully selected to reduce the uncertainty.

Keywords: Soil Particle Size, Sampling Location, Quantile Regression Forest, Digital Soil Mapping.

Fatehi Sh., Parvizi Y., and Navidi M.N. 2025. Analysis of different sources of uncertainty in digital mapping of some soil properties (case study of Ravansar plain, Kermanshah province). *Applied Soil Research*, 13 (2): 1-16.

1- Research Assistant Prof., Soil and Water Research Department, Kermanshah Agricultural and Natural Resources Research and Education Center, AREEO, Kermanshah, Iran

2- Research Professors of Soil Conservation and Watershed Management Research Institute, AREEO, Tehran

3- Research Associate Professors, Soil and Water Research Institute, Agricultural Research, Education and Extension Organization (AREEO), Karaj, Iran

* Corresponding Author Email: shahrokh.fatehi@gmail.com

تجزیه تحلیل منابع مختلف عدم قطعیت در نقشه‌برداری رقومی برخی ویژگی‌ها - های خاک (مطالعه موردی دشت روانسر، استان کرمانشاه)

شاهرخ فاتحی^{۱*}، یحیی پرویزی^۲، میر ناصر نویدی^۳

(تاریخ دریافت: ۱۴۰۳/۰۵/۰۹ تاریخ پذیرش: ۱۴۰۳/۰۹/۲۱)

چکیده

نقشه رقومی شن، سیلت و رس خاک در طراحی سیستم‌های آبیاری، ارزیابی تناسب اراضی و مدل‌های فرسایش خاک کاربرد دارد. از طرفی کمی‌سازی عدم قطعیت مکانی ویژگی‌های مذکور، معیاری از قابل اعتماد بودن این نقشه‌ها در اختیار کاربران قرار می‌دهد. به همین منظور در این پژوهش دو نوع عدم قطعیت شامل عدم قطعیت تغییرات داده‌های موجود و عدم قطعیت مدل با استفاده از شاخص انحراف معیار مورد بررسی قرار گرفت. این پژوهش در منطقه‌ای به وسعت ۵۷ هزار هکتار در دشت روانسر استان کرمانشاه اجرا شد. از ۱۲۰ نقطه مشاهداتی و مشتقات اولیه و ثانویه مدل رقومی ارتفاع و شاخص‌های پوشش گیاهی و مواد مادری حاصل از پردازش تصاویر لندست ۸ به عنوان ورودی‌های مدل استفاده شد. برای برآورد عدم قطعیت ناشی از تغییرات داده‌های موجود، ۵۰ بار مدل جنگل تصادفی بر اساس تفکیک تصادفی داده‌ها به دو دسته آموزشی (۲۵ درصد) و اعتبارسنجی (۷۵ درصد) اجرا شد. برای تخمین ویژگی‌های خاک و برآورد عدم قطعیت مدل از مدل جنگل رگرسیون چندک استفاده شد. نتایج اعتبارسنجی پیش‌بینی‌های مدل جنگل تصادفی نشان‌دهنده مقدار ضریب تبیین (R^2) ۰/۴۷، ۰/۳۸ و ۰/۲۷ به ترتیب برای شن، سیلت، و رس بود و مقدار ریشه میانگین مربعات خطا (RMSE) برای شن (۸/۶۱)، سیلت (۸/۲۵) و رس (۸/۱) تعیین شد. این نتایج بیان می‌کند که متغیرهای تخمین زنده، تغییرات شن خاک را به خوبی توضیح می‌دهند و اما توانایی مدل جنگل تصادفی برای پیش‌بینی مقدار شن نسبت به متغیرهای سیلت و رس خاک کمتر است. همچنین نتایج نشان داد که عدم قطعیت ناشی از تغییرات داده‌های مشاهداتی، حدود یک چهارم و عدم قطعیت مدل در حدود سه چهارم عدم قطعیت کل را شامل شده‌اند. از آنجا که عدم قطعیت مدل سهم بیشتری از عدم قطعیت کل را در بردارد برای کاهش عدم قطعیت کل بایستی در انتخاب مدل‌های یادگیری ماشین دقت لازم صورت گیرد.

واژه‌های کلیدی: اندازه ذرات خاک، مکان نمونه‌برداری، جنگل رگرسیون چندک، نقشه‌برداری رقومی خاک

فاتحی ش.، پرویزی ی.، نویدی م.ن. ۱۴۰۴. تجزیه تحلیل منابع مختلف عدم قطعیت در نقشه‌برداری رقومی برخی ویژگی‌های خاک (مطالعه موردی دشت روانسر، استان کرمانشاه). تحقیقات کاربردی خاک. جلد ۱۳، شماره ۲. صفحه: ۱-۱۶.

۱- استادیار پژوهش، بخش تحقیقات خاک و آب، مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی استان کرمانشاه، سازمان تحقیقات، آموزش و ترویج کشاورزی، کرمانشاه، ایران (مکاتبه کننده)

۲- استادیار پژوهش، پژوهشکده حفاظت خاک و آبخیزداری، سازمان تحقیقات، آموزش و ترویج کشاورزی، تهران، ایران

۳- دانشیار پژوهش، موسسه تحقیقات خاک و آب، سازمان تحقیقات، آموزش و ترویج کشاورزی، کرج، ایران

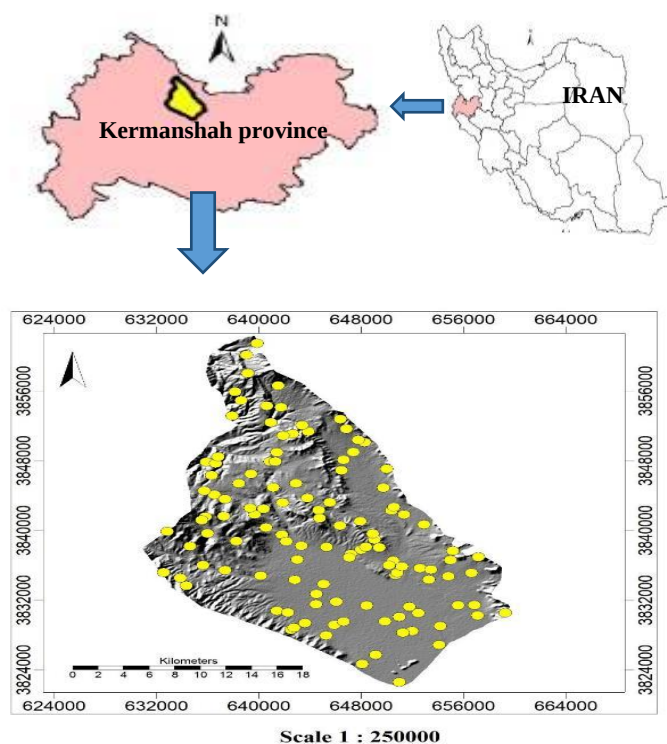
*پست الکترونیک: shahrokh.fatehi@gmail.com

مقدمه

در نقشه‌برداری رقومی خاک، پیش‌بینی توزیع مکانی ویژگی‌های خاک در مکان‌های فاقد اندازه‌گیری بر اساس رابطه داده‌های مشاهداتی و متغیرهای کمکی محیطی با استفاده از سیستم‌های استنباطی مکانی و غیر مکانی خاک صورت می‌گیرد (Lagacherie *et al*, 2007). از آنجا که تغییرات مکانی خاک پیچیده است و نمی‌توان آن‌ها را کاملاً مدل‌سازی نمود بنابراین به خطای پیش‌بینی منجر می‌شود. این پیش‌بینی‌ها با مقادیر واقعی متفاوت بوده و از این رو در مکان‌های پیش‌بینی در مورد مقدار واقعی قطعیتی وجود ندارد. مطالعات مختلف نشان داده است نقشه‌برداری رقومی ویژگی‌های خاک در معرض منابع مختلفی از عدم قطعیت قرار دارد (Heuvelink, 2014; Bishop *et al*, 2006; Behrens *et al*, 2008). علاوه بر عدم قطعیت در پارامترهای مدل و ساختار مدل، ممکن است عدم قطعیت در هنگام اندازه‌گیری ویژگی خاک در آزمایشگاه، موقعیت داده‌های مکانی خاک و در متغیرهای کمکی وجود داشته باشد. این منابع عدم قطعیت بر نتیجه تخمین‌های مدل نقشه‌برداری رقومی خاک تأثیر خواهند گذاشت. همانطور که بیان گردید نقشه‌برداری رقومی خاک بر روابط آماری بین ویژگی‌های خاک لندازه‌گیری شده و متغیرهای کمکی محیطی در مکان‌های نمونه‌برداری متکی است. اگر در مکان‌های اخذ نمونه خطا وجود داشته باشد، یعنی مشاهدات خاک و متغیرهای کمکی دقیقاً بر هم منطبق نباشند و این امر باعث ایجاد خطا در روابط استنباط شده می‌شود (Cressie & Kornak, 2003). منبع دیگر خطا در روابط استنباط شده ناشی از عدم قطعیت مکانی در متغیرهای کمکی پیش‌بینی کننده مورد استفاده برای نقشه‌برداری رقومی خاک است که اغلب دارای منشأ، مقیاس، قدرت تفکیک مکانی متفاوتی هستند. مطالعات متعددی تأثیر عدم قطعیت پیش‌بینی کننده‌های محیطی (به عنوان مثال عدم قطعیت در داده‌های ارتفاع) را در مدل‌های خاک-زمین نما بررسی کرده‌اند (Heuvelink, 2014; Bishop *et al*, 2006; Behrens *et al*, 2008; Gomez, 2015). بر نیاز به تحلیل عدم قطعیت در تخمین ویژگی‌های خاک، به ویژه در زمینه داده‌های فراطیفی تأکید کرد. پوگیو (Poggio, 2016) یک رویکرد مدل‌سازی

مکانی مبتنی بر آمار بیزی برای کمی‌سازی عدم قطعیت در نقشه‌های ویژگی‌های خاک ارائه کرد و آن را به عنوان جایگزین مناسبی برای روش‌های موجود معرفی نمود. این پژوهش در مجموع بر نیاز به درک جامع و در نظر گرفتن عدم قطعیت در نقشه‌برداری رقومی ویژگی‌های خاک تأکید داشت. استفاده از جنگل رگرسیون چندک^۱ در نقشه‌برداری رقومی خاک در چندین مطالعه مورد بررسی قرار گرفته است. جنگل رگرسیون چندک یک ابزار قدرتمند در نقشه‌برداری رقومی خاک برای تخمین عدم قطعیت است. تیریس (Theres, 2022) از جنگل رگرسیون چندک برای تخمین خصوصیات خاک در هند استفاده کرد و در مقایسه با روش‌های سنتی بررسی خاک به دقت بالاتری دست یافت. وانگ (Wang, 2022) جنگل رگرسیون چندک را با رویکرد مبتنی بر شباهت ادغام کرد که منجر به دقت کلی بالاتر در نقشه‌برداری خاک شد. این مطالعات در مجموع پتانسیل جنگل رگرسیون چندک را در بهبود دقت و کارایی نقشه‌برداری رقومی خاک نشان می‌دهند. سورت و همکاران (Saurette *et al*, 2022) جنگل رگرسیون چندک را در یک روش نقشه‌برداری رقومی خاک سه بعدی گنجانده و اثربخشی آن را در تفسیر روابط غیرخطی خاک-زمین نما و کاهش عدم قطعیت پیش‌بینی، نشان داد. نیکو (Nikou, 2022) دریافت که جنگل رگرسیون چندک از سایر روش‌ها در تخمین مواد آلی خاک بهتر عمل می‌کند و در عین حال تخمین خوبی از عدم قطعیت ارائه می‌دهد. کسرای (Kasraei, 2021) جنگل رگرسیون چندک را با تکنیک‌های یادگیری ماشین برای تولید تخمین‌های عدم قطعیت قابل اعتماد و پایدار ادغام کرد. زامتاری (Szatmári, 2022) جنگل رگرسیون چندک را با روش‌های دیگر مقایسه کرد و دریافت که یکی از بهترین‌ها در تعیین کمیت عدم قطعیت است. این مطالعات در مجموع اثربخشی جنگل رگرسیون چندک را در تخمین عدم قطعیت در نقشه‌برداری رقومی خاک نشان می‌دهند. عدم قطعیت کل در نقشه‌برداری رقومی خاک به چهار مؤلفه عدم قطعیت در اندازه‌گیری داده‌های خاک، عدم قطعیت در متغیرهای کمکی خاک، عدم قطعیت در مدل و عدم قطعیت در تغییرات داده‌های مشاهداتی موجود قابل تقسیم است. رویکردهایی که در منابع علمی را برای برآورد کمی عدم قطعیت تخمین ویژگی‌های خاک از طریق

¹ Quantile Regression Forest



شکل ۱- موقعیت نقاط مشاهداتی (۱۲۰ نمونه) و ناحیه مورد مطالعه در ایران و استان کرمانشاه

Figure 1. Location of soil samples (n=120) and the studied area in Kermanshah province and Iran

جمع آوری داده‌ها و متغیرهای کمکی

در پژوهش حاضر با استفاده از بسته نرم‌افزاری cLHS در محیط R و متغیرهای کمکی (جدول ۱)، موقعیت جغرافیایی ۱۲۰ نقطه مشاهداتی تعیین گردید (شکل ۱). در گام بعدی نمونه‌برداری خاک تا عمق ۳۰ سانتیمتری خاک سطحی صورت گرفت و سپس در آزمایشگاه درصد اندازه ذرات شن، سیلت و رس خاک نمونه‌ها با روش هیدرومتری به دست آمد (Bouyoucos, 1962). تجزیه و تحلیل آماری اولیه روی تمام داده‌ها با استفاده از نرم‌افزار R صورت گرفت.

۲۸ متغیر کمکی شامل ۲۰ متغیر کمکی مربوط عامل خاکساز توپوگرافی، ۲ متغیر کمکی مرتبط با پوشش گیاهی به همراه ۶ متغیر کمکی در ارتباط با فاکتور مواد مادری به عنوان متغیرهای پیش‌بینی کننده مورد استفاده قرار گرفت (جدول ۱). شاخص‌های توپوگرافی با استفاده از نرم‌افزار ساگا نسخه ۹ از مدل رقومی ارتفاع SRTM با قدرت تفکیک ۳۰ متر تهیه گردیدند. شاخص‌های توپوگرافی مانند تابش مستقیم^۳، تابش پخشیده^۴، نسبت

مدل‌های مختلف به کار رفته است عبارت است از ۱- واریانس پیش‌بینی کریجینگ جهانی^۱، ۲- بوت استرپینگ^۲، ۳- کمی سازی عدم قطعیت تجربی از طریق تقسیم‌بندی داده‌ها و اعتبارسنجی متقابل و ۴- کمی سازی عدم قطعیت تجربی از طریق خوشه‌بندی فازی و اعتبارسنجی متقابل. در همه این روش‌ها مجموع سه مؤلفه عدم قطعیت ساختار مدل، عدم قطعیت پارامتر مدل و عدم قطعیت ورودی مدل به عنوان عدم قطعیت کل برآورد می‌شود. بنابراین، در بیشتر پژوهش‌های گذشته به ندرت به بررسی مؤلفه‌های عدم قطعیت در نقشه‌برداری رقومی خاک به طور مجزا پرداخته شده است. پژوهش حاضر به هدف تخمین مؤلفه عدم قطعیت ناشی از تغییرات داده‌های موجود ذرات رس، سیلت و شن خاک با استفاده از مدل‌های جنگل تصادفی و عدم قطعیت مدل در پیش‌بینی ویژگی‌های مذکور با استفاده از جنگل رگرسیون چندک به صورت منفرد و بررسی سهم هر مؤلفه در عدم قطعیت کل در منطقه روانسر استان کرمانشاه اجرا شد. لازم به ذکر است در این پژوهش فرض بر این است عدم قطعیت ناشی از متغیرهای کمکی و اندازه‌گیری آزمایشگاهی داده‌های خاک ناچیز باشد.

مواد و روش‌ها

منطقه مورد مطالعه

این پژوهش در محدوده شهرستان‌های روانسر و جوانرود در ناحیه‌ای به وسعت حدود ۵۷/۸۵ هزار هکتار در ۶۰ کیلومتری شمال غربی استان کرمانشاه، در بین مختصات جغرافیایی ۴۶° ۲۶/۸' تا ۴۶° ۴۲/۹' طول شرقی و ۳۴° ۳۳/۳' تا ۳۴° ۵۱/۶' عرض شمالی اجرا شد (شکل ۱). این محدوده در برگیرنده انواع کاربری‌های اراضی شامل جنگل، مرتع، اراضی دیم و آبی می‌باشد. سازندهای زمین‌شناسی آن عمدتاً آهکی و در برخی از قسمت‌های منطقه مورد مطالعه مارن، کنگلومرا و ماسه‌سنگ دیده می‌شود. خاک‌های آن عمدتاً در چهار رده خاک انسیپتی سولز، ورتی سولز، انتی سولز و مالی سولز قرار می‌گیرند.

³ Direct Insolation
⁴ Diffuse Insolation

¹ Universal kriging prediction variance
² Bootstrapping

شاخص INR در فرمول اسکورپن به عنوان شاخص نسبت آهن نرمال شده^{۲۳} است که بیانگر کانی‌های آهن در مواد مادری می‌باشد. CHNR به عنوان شاخص نسبت رس نرمال شده^{۲۴} در فرمول اسکورپن، نماینده رس و کانی‌های دارای هیدروکسیل خاک است. شاخص RONR در فرمول اسکورپن به عنوان شاخص نسبت رخنمون سنگی نرمال شده^{۲۵}، نشان‌دهنده سنگ‌های رسوبی در مقابل سنگ‌های آذرین می‌باشد. شاخص FNR در فرمول اسکورپن به عنوان شاخصی برای آهن دو ظرفیتی نرمال شده^{۲۶} تعریف شده و نشان‌دهنده کانی‌های دارای آهن دو ظرفیتی است (Soil Science Division Staff, 2017) (جدول ۱).

انتخاب ویژگی^{۲۷}

انتخاب ویژگی، فرآیند شناسایی و انتخاب زیرمجموعه‌ای از متغیرهای تخمین زنده به عنوان ورودی‌های مدل‌های یادگیری ماشین است که بیشترین ارتباط را با متغیر هدف دارند. دلایل مختلفی برای انتخاب زیرمجموعه‌ای از متغیرهای کمکی برای کالیبره کردن مدل وجود دارد. برخی از آنها عبارتند از: (۱) کالیبره کردن سریع‌تر مدل یادگیری ماشین، (۲) کاهش پیچیدگی، (۳) افزایش دقت پیش‌بینی، (۴) جلوگیری از هم خطی چندگانه^{۲۸} و (۵) جلوگیری از بیش برآورد مدل‌های یادگیری ماشین (Wadoux et al, 2022). استفاده از رگرسیون چند متغیره گام به گام یکی از روش‌های کاهش متغیر است که رسیدن به اهداف فوق به ویژه مورد چهارم را ممکن می‌سازد. بنابراین، در این پژوهش از روش رگرسیون چند متغیره گام به گام برای انتخاب متغیرهای کمکی با رابطه معنی‌دار با متغیرهای هدف استفاده شد. همچنین فاکتور تورم واریانس^{۲۹} برای حذف متغیرهای کمکی دارای همبستگی باهم (هم خطی^{۳۰}) با استفاده از تابع vif() در بسته نرم‌افزاری car مورد استفاده قرار گرفت و در نهایت بهترین

تابش مستقیم به پخشیده^۱ و گرمایش ناهمسانگرد روزانه^۲، ارتفاع^۳، شیب^۴، جهت شیب^۵، تحدب^۶، بافت^۷ و موقعیت شیب میانی^۸، شاخص موقعیت توپوگرافیک^۹، شاخص همگرایی^{۱۰}، شاخص همواری درجه با درجه تفکیک بالا^{۱۱}، شاخص همواری خط‌الراس با درجه تفکیک بالا^{۱۲}، شاخص ناهمواری سرزمین^{۱۳} به منظور نشان دادن مورفومتری در مقیاس محلی انتخاب شده‌اند. عمق دره^{۱۴}، فاصله عمودی تا شبکه آبراهه^{۱۵} مربوط به وضعیت شبکه آبراهه‌ها و عدد ناهمواری میلیتون^{۱۶}، شاخص خیزی توپوگرافی^{۱۷}، شاخص قدرت جریان^{۱۸} مربوط به مشخصات هیدرولوژیکی منطقه مورد مطالعه است. همه‌ی این متغیرهای کمکی مربوط فاکتور توپوگرافی اسکورپن می‌باشند (Conrad et al, 2015).

برای به دست آوردن شاخص‌های پوشش گیاهی و مواد مادری از تصاویر ماهواره لندست ۸ تاریخ خرداد و تیر ۱۳۹۸ استفاده شد. کلیه تصحیحات رادیومتری، اتمسفری و توپوگرافی بر روی این تصاویر با استفاده از بسته نرم‌افزاری RStoolbox در محیط R انجام گرفت. شاخص پوشش گیاهی نرمال شده^{۱۹} (Rouse et al, 1973) و شاخص پوشش گیاهی تصدیق شده^{۲۰} یا SAVI در فرمول اسکورپن به عنوان نماینده عامل پوشش گیاهی استفاده شد (Jensen, 2005). مواد مادری خاک به عنوان یکی از فاکتورهای اسکورپن نقش مهمی در تشکیل اجزا بافت خاک دارد به همین منظور از شاخص‌های تصاویر ماهواره‌ای که احتمال می‌رفت نماینده نحوه تشکیل، مقدار و توزیع شن، سیلت و رس خاک باشند به عنوان متغیر کمکی در نظر گرفته شد. مانند شاخص اندازه ذرات^{۲۱} نشان‌دهنده نسبت اندازه ذرات نرمال شده است (Xiao et al., 2006). شاخص نسبت کربنات کلسیم نرمال شده^{۲۲} شاخص برای برآورد کربنات کلسیم در مواد مادری خاک به کار می‌رود.

¹⁶ Melton Ruggedness Number

¹⁷ Topographic Wetness Index

¹⁸ Stream Power Index

¹⁹ Normalized difference vegetation index

²⁰ Soil Adjusted Vegetation Index

²¹ Grain Size Index

²² Carbonate Normalized Ratio

²³ Iron Normalized Ratio

²⁴ Clay(hydroxyls) Normalized Ratio

²⁵ Rock Outcrop Normalized Ratio

²⁶ Ferrous Normalized Ratio

²⁷ Feature selection

²⁸ multicollinearity

²⁹ variance inflation factor

³⁰ collinearity

¹ Direct to Diffuse Ratio

² Diurnal Anisotropic Heating

³ Elevation

⁴ Slope

⁵ Aspect

⁶ Convexity

⁷ Texture

⁸ Mid Slope Position

⁹ Topographic Position Index

¹⁰ Convergence Index

¹¹ Multiresolution valley bottom flatness(MRVBF) index

¹² Multiresolution ridge top flatness (MRRTF) index

¹³ Terrain ruggedness Index

¹⁴ Valley Depth

¹⁵ Vertical Distance to Channel Network

مدل با متغیرهای کمکی استخراج گردید و از این متغیرهای کمکی نهایی به عنوان متغیرهای کمکی مهم در مدل‌های جنگل تصادفی و جنگل رگرسیون چندک برای پیش‌بینی اندازه ذرات شن، سیلت و رس در ناحیه مورد مطالعه استفاده شد.

نقشه‌برداری رقومی خاک

در این پژوهش از مدل جنگل تصادفی و جنگل رگرسیون چندک استفاده شد. الگوریتم جنگل تصادفی با موفقیت در مطالعات مختلف برای شبیه‌سازی توزیع مکانی خصوصیات خاک به کار گرفته شده است. روش‌های یادگیری ماشین، به ویژه آن‌هایی که مبتنی بر تکرار نمونه‌گیری مجدد و برازش مدل تقویت‌شده هستند، می‌توانند برآوردهای تجربی عدم قطعیت را ارائه دهند که می‌توان به رایج‌ترین آن الگوریتم جنگل رگرسیون چندک اشاره کرد.

جنگل تصادفی

مدل جنگل تصادفی، توسط بریمن (Breiman, 2001) برای بهبود عملکرد درخت رگرسیون با استفاده از تکنیک‌های تجمیع باز نمونه‌گیری‌ها¹ پیشنهاد شد. این مدل امکان استفاده از مجموعه‌ای از درختان رگرسیون و در نتیجه دستیابی به تخمین‌های قوی‌تر با خطای داخلی کم‌تر را فراهم می‌کند. دو ویژگی مهم جنگل‌های تصادفی شامل تصادفی بودن نمونه‌گیری و تصادفی بودن انتخاب متغیرهای پیش‌بینی‌کننده است. اولین ویژگی به این معنی است که هر درخت در یک مدل جنگل تصادفی از یک زیرمجموعه نمونه تصادفی ساخته می‌شود. زیرمجموعه با باز نمونه‌گیری و جایگزینی آن با مجموعه نمونه اصلی ایجاد می‌گردد. نمونه‌ها می‌تواند تکراری از زیرمجموعه‌های مختلف یا همان زیرمجموعه باشد که نشان می‌دهد برخی از نمونه‌ها چندین بار در یک درخت استفاده خواهند شد. در مورد ویژگی دوم می‌توان بیان داشت که تنها یک زیرمجموعه تصادفی از تمام متغیرهای کمکی، هنگام تقسیم هر گره در هر درخت انتخاب می‌شود. این دو ویژگی مدل جنگل تصادفی می‌تواند تنوع درخت‌های تصمیم را در جنگل‌های تصادفی افزایش دهند و در نتیجه عملکرد مدل را بهبود بخشند. به طور کلی، الگوریتم جنگل‌های تصادفی شامل چهار مرحله اصلی زیر است:

یک زیرمجموعه از همه نمونه‌ها را به عنوان مجموعه آموزشی با روش نمونه‌گیری مجدد (بوت استرپ) انتخاب می‌شود، (۲) یک درخت تصمیم با زیرمجموعه‌ای از نمونه‌های انتخابی تولید و سپس (۳) دو مرحله بالا K بار تکرار می‌شود تا K درخت‌های تصمیم ساخته شود و در نهایت (۴) مقدار متغیر هدف را با مدل جنگل‌های تصادفی آموزش‌دیده (K درخت)، پیش‌بینی می‌شود و نتیجه نهایی از نتایج تخمین‌های درختان مختلف محاسبه می‌شود. در این پژوهش برای برآورد عدم قطعیت تغییرات داده‌های موجود از نتایج اجرای ۵۰ بار مدل جنگل تصادفی بر اساس تفکیک تصادفی داده به دو دسته آموزشی (۲۵ درصد) و اعتبارسنجی (۷۵ درصد) و بسته نرم‌افزاری Randomforest استفاده شد.

جنگل رگرسیون چندک

برای تخمین ویژگی‌های خاک و برآورد عدم قطعیت مدل در منطقه مورد مطالعه از مدل رگرسیون چندک استفاده شد. مینش‌اوسین (Meinshausen, 2006) برای اصلاح خروجی‌های الگوریتم جنگل تصادفی با اجازه دادن به تخمین فواصل تخمین متغیرهای هدف، مدل رگرسیون چندک را پیشنهاد داد. QRF توسعه‌یافته مدل جنگل تصادفی است و مزیت QRF نسبت به مدل جنگل تصادفی در ارتباط با گره‌ها در هر درخت است، مدل جنگل تصادفی فقط میانگین مشاهداتی را که در این گره قرار می‌گیرند نگه می‌دارد و تمام اطلاعات دیگر را نادیده می‌گیرد در حالی که QRF ارزش تمام مشاهدات را در این گره نگه می‌دارد و توزیع شرطی را بر اساس این اطلاعات ارزیابی می‌کند. در این پژوهش برای تخمین عدم قطعیت مدل، از مدل جنگل تصادفی چندک و به همین منظور از بسته نرم‌افزاری quantregForest در محیط R استفاده شد.

اعتبارسنجی

در این مطالعه، برای ارزیابی مدل‌ها از روش تفکیک داده‌ها به دو دسته آموزشی و اعتبارسنجی استفاده شد. خطای مدل عبارت است از اختلاف بین داده‌های واقعی و داده‌های پیش‌بینی شده از اجرای مدل. هرچه این اختلاف کمتر باشد نشان‌دهنده صحت بالای مدل مورد نظر است. خطای مدل را به شکل‌های مختلفی می‌توان نشان داد.

¹ Bootstrap

خطای مطلق و ریشه میانگین مربعات خطا کمتر باشد
مدل از صحت و دقت بالایی برخوردار خواهد بود.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n [z_{xi}^* - z_{xi}]^2}{n}} \quad (1)$$

مرسوم‌ترین آن‌ها ریشه میانگین مربعات خطای تخمین
(Root Mean Squared Error)، میانگین خطای مطلق
(Mean Absolute Error) می‌باشد. معیار دیگر برای
تخمین دقت مدل، ضریب تبیین R^2 است هر چه مقدار
ضریب تبیین به یک نزدیک باشد و هرچه مقدار میانگین

جدول ۱- متغیرهای کمکی محیطی حاصل از مدل رقومی ارتفاع و تصاویر ماهواره لندست ۸

Table 1. Environmental auxiliary variables obtained from the digital elevation model and Landsat 8 satellite images

Scorpan factor	Environmental covariates	min	max	mean	mathematical equation
Relief	Diffuse Insolation	286.8	319.5	311.1	—
	Direct Insolation	1393	2064	1838	—
	Direct to Diffuse Ratio	4.71	6.88	5.91	—
	Diurnal Anisotropic Heating	-0.28	0.25	0.003	—
	Aspect	0.025	6.18	2.99	—
	Convexity	28	64.29	45.85	—
	Convergence Index	-56.83	57.87	-1.28	—
	Elevation	1318	1850	1448	—
	Melton Ruggedness Number	0	6.49	1.19	—
	Mid Slope Positon	0	0.96	0.67	—
	MRVBF	0	5.99	1.92	—
	MRRTF	0	4.95	0.2	—
	Slope	0	0.36	0.09	—
	Stream Power Index	0.06	2997.50	61.18	—
	Terrain ruggedness Index	0.06	7.2	1.8	—
	Texture	0	18.36	2.81	—
	Topographic Position Index	-2.5	8.6	0.24	—
	TopographicWettnes Index	4.44	16.95	8.56	—
Parent material	Valley Depth	3.6	542	78.91	—
	Vertical Distance to Channel Network	0	183.18	26.48	—
	Grain Size Index	-0.1	0.17	0.11	$GSI = \frac{RED - BLUE}{RED + GREEN + BLUE} *$
	Carbonate Normalized Ratio	-1.24	0.12	-0.7	$CaNR = \frac{RED - GREEN}{RED + GREEN} *$
	Iron Normalized Ratio	-0.1	0.21	0	$INR = \frac{RED - SWRI(a)}{RED + SWRI(a)} *$
vegetation	Clay(hydroxyls) Normalized Ratio	0.11	0.34	0.2	$CINR = \frac{SWRI(a) - SWRI(b)}{SWRI(a) + SWRI(b)} *$
	Rock Outcrop Normalized Ratio	0.2	0.38	0.3	$RONR = \frac{SWRI(a) - GREEN}{SWRI(a) + GREEN} *$
	Ferrous Normalized Ratio	-0.41	0.07	-0.4	$FNR = \frac{SWRI(a) - NIR}{SWRI(a) + NIR} *$
	Normalized differencevegetation index	0.15	0.51	0.25	$NDVI = \frac{NIR - RED}{NIR + RED} *$
	Soil Adjusted Vegetation Index	0.23	0.76	0.37	$SAVI = \frac{NIR - RED}{NIR + RED + L} (1 + L) **$

* در تمامی روابط فوق RED، GREEN، BLUE، NIR، SWRI(a) و SWRI(b) به ترتیب باندهای آبی، سبز، قرمز، مادون قرمز و باندهای مادون قرمز کوتاه (باند ۶ و ۷) در ماهواره لندست ۸ می‌باشد.

** L فاکتور تصحیح می‌باشد که مقدار صفر برای پوشش گیاهی متراکم، ۱ برای پوشش گیاهی تنک و ۰/۵ برای پوشش گیاهی متوسط است.

$$MAE = \frac{1}{n} \sum_{i=1}^n |z_{xi}^* - z_{xi}| \quad (2)$$

در این معادله $R^2 = \text{ضریب تبیین}$ ، $y_i = \text{متغیر پیش‌بینی}$ شده، $x_i = \text{متغیر اندازه‌گیری شده}$ ، $\bar{y} = \text{میانگین متغیر پیش‌بینی}$ شده، $\bar{x} = \text{میانگین متغیر اندازه‌گیری شده}$ است.

جمع مقادیر نقشه انحراف معیار مدل و نقشه حساسیت به تغییرات داده‌های موجود به دست آمد.

نتایج و بحث

خلاصه آماری اجزای جافت خاک در جدول ۲ ارائه شده است. مقدار شن بین ۱/۷ تا ۶۲/۲ درصد، مقدار سیلت ۱۵/۷ تا ۶۴/۲ درصد، مقدار رس بین ۱۸ تا ۶۱ درصد، متغیر است. دامنه وسیع تغییرات این ویژگی‌های خاک ناشی از تنوع عوامل خاکسازي شامل مواد مادری و توپوگرافی و پوشش گیاهی می‌باشد. کمترین ضریب تغییرات مربوط به رس خاک (۱۸/۷۲٪) و بالاترین مقدار آن مربوط به شن خاک (۷۱/۲۹٪) است. به نظر می‌رسد بالا بودن ضریب تغییرات مقدار شن خاک در منطقه مورد مطالعه با پراکنش و تنوع سازندهای زمین‌شناسی (ماسه‌سنگ، کنگلومرا و سنگ آهک) در ناحیه مورد مطالعه در ارتباط باشد. ضریب تغییرات کمتر به مفهوم کم بودن پراکندگی و نشانگر همگن بودن جامعه است. از این نظر مقدار شن در خاک از نظر آماری همگن نبوده که این خود بر نرمال نبودن توزیع داده‌های آن تأثیر گذاشته است. این در حالی است که سیلت با p-value بیش از ۰/۰۵ از توزیع نرمال پیروی می‌کند هیستوگرام مقدار شن خاک دارای چولگی زیاد به سمت راست بوده (۱/۳۶) اما رس کمی چولگی (۰/۷-) به سمت چپ دارد که نشان می‌دهد توزیع آن به توزیع نرمال نزدیک‌تر است (شکل ۲).

در معادلات فوق $RMSE = \text{ریشه میانگین مربعات خطای تخمین}$ ، $z_{x_i}^* = \text{متغیر پیش‌بینی شده}$ ، $z_{x_i} = \text{متغیر اندازه‌گیری شده}$ است.

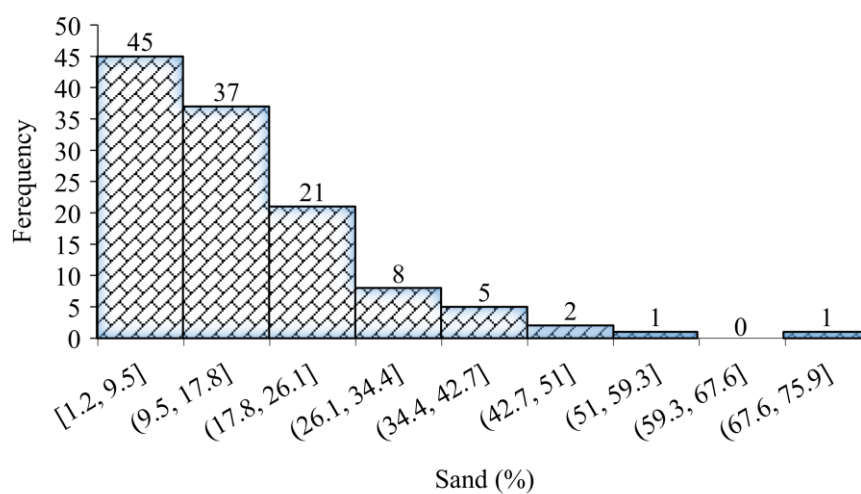
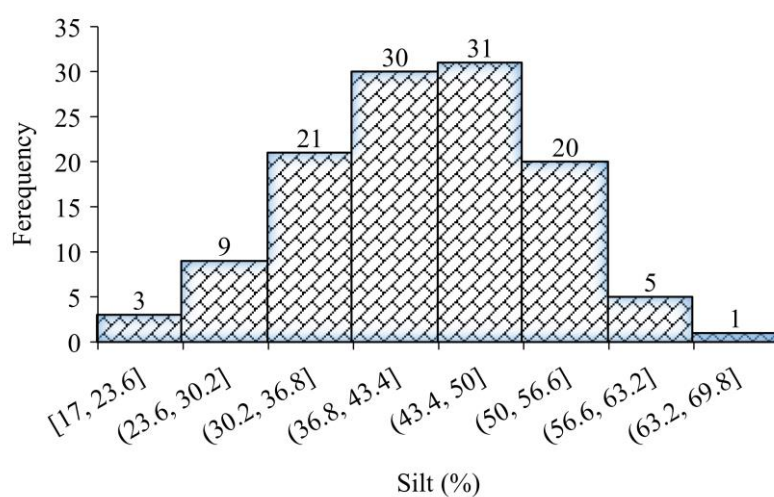
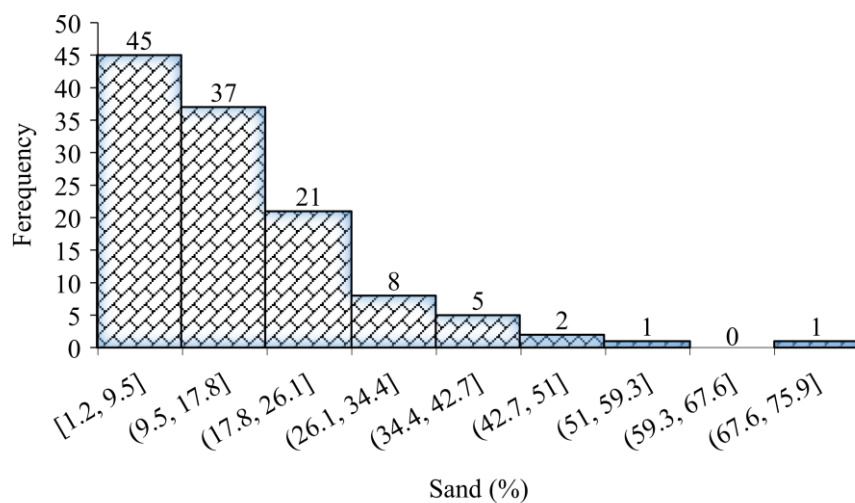
$$R^2 = \frac{\sum_{i=1}^n (x_i y_i - \bar{x} \bar{y})}{(\sum_{i=1}^n x_i^2 - \bar{x}^2)(\sum_{i=1}^n y_i^2 - \bar{y}^2)} \quad (3)$$

تخمین عدم قطعیت در تغییرات داده‌های موجود یا حساسیت مدل به داده‌های موجود

برای تجزیه و تحلیل حساسیت مدل به داده‌های موجود، ۵۰ بار داده‌های مشاهداتی به طور تصادفی به زیرمجموعه‌های آموزشی (۲۵٪ کل داده‌ها یعنی ۳۰ داده مشاهداتی) و آزمایشی (۷۵٪ کل داده‌ها یعنی ۹۰ داده مشاهداتی) تقسیم شدند. سپس اجرای مدل و تولید نقشه تخمین برای متغیرهای شن، سیلت و رس خاک در هر تفکیک داده انجام شد، در اینجا، انحراف معیار مقادیر تخمین زده شده در سطح پیکسل، حساسیت مدل به تغییرات داده‌های موجود را نشان می‌دهد. عدم قطعیت مدل با استفاده مدل جنگل رگرسیون چندک و همان mtry و ntree به دست آمده از اجرای مدل جنگل تصادفی و کل داده‌های مشاهداتی با محاسبه انحراف معیار به دست آمد. استفاده از انحراف معیار در نقشه‌برداری رقومی خاک برای تعیین عدم قطعیت در چندین مطالعه استفاده شده است. تیکسیرا (Teixeira, 2018) و استامپ (Stumpf, 2017) هر دو کاربرد نقشه‌های انحراف معیار را برای شناسایی مناطق با عدم قطعیت بالا بررسی کردند. تیکسیرا (Teixeira, 2018) به طور خاص از این نقشه‌ها برای بهبود واحدهای نقشه‌برداری خاک استفاده کرد. در پژوهش حاضر از انحراف معیار به عنوان شاخص عدم قطعیت استفاده شد. نقشه عدم قطعیت کل نیز از حاصل

جدول ۲- آماره‌های شن، سیلت و رس
Table 2. Statistics of sand, silt and clay

Soil properties	Unit	Min	Max	Range	Median	Mean	Varince	Standard Deviation	CV	Skewness	Kurtosis	Shapiro-Wilk Test P-value
sand	%	1.69	62.21	60.52	13.78	16.86	144.4	12.02	71.29	1.36	1.75	0
silt	%	15.74	64.15	48.41	38.69	38.68	88.5	9.41	24.32	-0.11	-0.32	0.78
clay	%	18.03	60.78	42.75	44.75	44.47	69.28	8.32	18.72	-0.7	0.66	0.004



شکل ۲- هیستوگرام شن، سیلت و رس
Figure 2. Histogram of sand, silt and clay

انتخاب متغیرهای کمکی

روش رگرسیون گام به گام دوطرفه برای تعیین متغیرهای کمکی بدون هم خطی و دارای رابطه معنی دار با مقادیر شن، سیلت و رس خاک به کار رفت و مهم ترین متغیرهای کمکی انتخابی در جدول ۳ ارائه شده است. برای شن خاک مهم ترین متغیرهای کمکی به ترتیب فاصله عمودی تا شبکه آبراهه ها، بافت سرزمین، MRRTF، شاخص قدرت جریان، شاخص خیزی توپوگرافی و گرمایش ناهمسانگرد روزانه بودند که در سطح ۰/۰۵ و کمتر معنی دار هستند. در رابطه خطی گام به گام متغیرهای کمکی با سیلت خاک، به ترتیب شاخص پوشش گیاهی تصدیق شده و شاخص موقعیت توپوگرافی، تابش پخشیده و عمق دره، در سطح ۰/۰۵ و کمتر، رابطه معنی دار نشان دادند. رس خاک با متغیرهای انتخابی در سطح ۰/۰۱ درصد رابطه معنی دار نشان داد که مهمترین آنها فاصله عمودی تا شبکه آبراهه ها، شاخص موقعیت توپوگرافی و بافت سرزمین بودند. دو متغیر کمکی

عمق دره و فاصله عمودی تا شبکه آبراهه ها متغیر کمکی انتخابی مشترک برای هر سه ویژگی خاک مذکور است که اهمیت ویژگی های هیدرولوژیکی منطقه و عامل توپوگرافی در پراکنش بافت خاک و گروه های اندازه ای آن را نشان می دهد. در تأیید نتایج مذکور، خسروانی و همکاران (Khosravani et al, 2004) نشان دادند که عمق دره مهمترین متغیر کمکی برای پیش بینی شن، سیلت و رس خاک می باشد. رامچران و همکاران (Ramcharan et al, 2018) و آکپا و همکاران (Akpa et al, 2014) گزارش کردند که مدل رقومی ارتفاع و مشتقات آن مهمترین متغیرها ورودی برای نقشه برداری رقومی اندازه ذرات خاک هستند. به طور کلی توپوگرافی به عنوان عامل کنترل کننده غالب برای خصوصیات خاک شناخته شده است که بر توزیع و تجمع آب و انرژی و انتقال رسوب تأثیر می گذارد (MacMillan et al, 2004).

جدول ۳- متغیرهای کمکی انتخاب شده به عنوان ورودی به مدل های جنگل تصادفی و جنگل رگرسیون چندک برای برآورد شن، سیلت

و رس

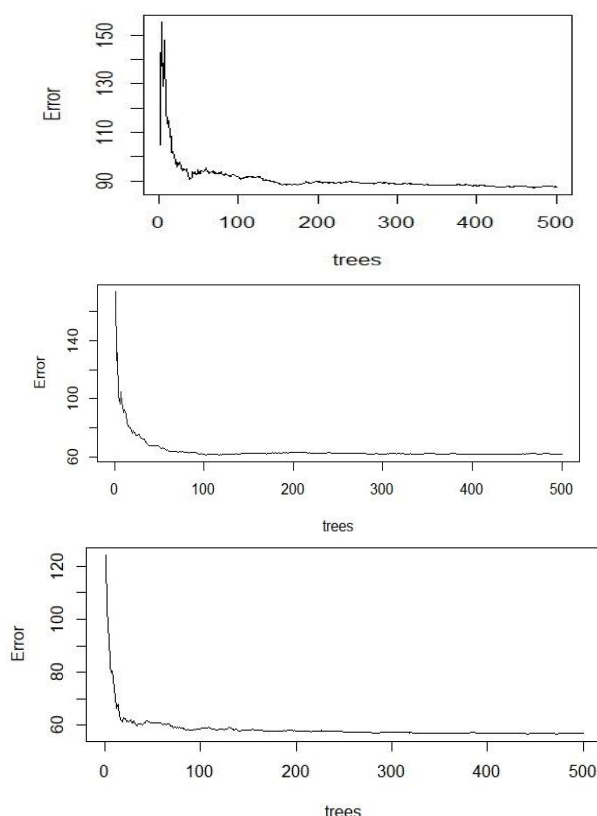
Table 2- Selected Covariates as input to the random forest and quantile regression forest models to estimate the sand, silt and clay

Soil propertise	Importance covariates	Adjusted R ²
Sand	Savi , Diffuse Insolation**, Diurnal Anisotropic Heating. , Mid.Slope Positon , MRRTF*, Stream Power Index. , Terrain Ruggedness Index .TRI. , Texture***, Topographic Wetness Index. , Valley Depth*, Vertical Distance to Channel_Network***	0.57
Silt	savi., Cnr, Diffuse Insolation***, Diurnal Anisotropic Heating, MRRTF, Topographic Position Index*, Valley Depth***, Vertical Distance to Channel Network	0.41
Clay	Gsi*, Direct Insolation*, Elevation*, Melton Ruggedness Number*, Texture***, Topographic Position Index***, Valley Depth**, Vertical Distance to Channel Network***, Ronr*	0.41

علامت معنی دار بودن: "***" = سطح ۰، "*" = سطح ۰/۰۱، "." = سطح ۰/۰۵، "بدون علامت" = سطح ۱ درصد

جدول ۴ ارائه شده است. مقدار R² به ترتیب برای شن، سیلت، و رس ۰/۴۷، ۰/۳۸ و ۰/۲۷ به دست آمده است اما مقدار RMSE و MAE به ترتیب برای شن، سیلت و رس کمتر شده و سیر نزولی دارد. ضریب تبیین نسبتاً بالا برای شن نشان داد که متغیرهای تخمین زننده، تغییرات متغیر شن را به خوبی توضیح می دهند و اما مقادیر بالای RMSE و MAE برای شن بیان نگر کارایی کم مدل جنگل تصادفی در تخمین مقدار شن نسبت به متغیرهای سیلت و رس خاک است.

در گام اول به منظور تعیین تعداد درخت مناسب، مدل جنگل تصادفی با استفاده از کل داده های مشاهداتی و متغیرهای انتخابی اجرا شد. شکل ۳ رابطه بین تعداد درخت و خطا را برای هر سه ویژگی خاک نشان می دهد که برای هر سه متغیر گروه های اندازه ای بافت خاک تا اندازه ۵۰۰ درخت کاهش خطا ثابت شده و این نشان می دهد ۵۰۰ درخت، تعداد درخت مناسب برای اجرای مدل می باشد. خروجی نهایی اجرای مدل بیانگر مقدار mtry برابر با ۲ بود. از طرفی، نتیجه اعتبارسنجی متقابل برای کل داده ها در



شکل ۳- از بالا به پایین به ترتیب نمودار تعداد درخت در برابر خطا در اجرای مدل جنگل تصادفی با کل داده‌ها برای شن، سیلت، رس
Figure3. From top to bottom, the graph of the number of trees versus the error in running the random forest model with all the data for sand, silt, clay, respectively.

جدول ۴- نتایج اعتبارسنجی دو جانبه مدل جنگل تصادفی بر اساس کل داده‌های مشاهداتی

Table 4. results of cross validation of random forest model based on all observational data

Soil propertise	Validation measures		
	RMSE	MAE	R ²
Sand	8.7	6.51	0.47
Silt	7.65	6.14	0.38
Clay	7.48	6.02	0.28

حساسیت مدل یا عدم قطعیت در تغییرات داده‌های موجود

بر اساس تعداد درخت معادل با ۵۰۰ و مقدار mtry برابر با ۲، مدل جنگل تصادفی برای ۲۵ درصد داده‌ها برای هر سه متغیر شن، سیلت و رس خاک با ۵۰ تکرار اجرا شد. نتایج اعتبارسنجی مستقل برای ۷۵ درصد داده‌ها به صورت میانگین در جدول ۳ ارائه شده است. برای شن مقادیر شاخص‌های اعتبارسنجی دو جانبه (جدول ۴) و اعتبارسنجی با داده‌های مستقل (جدول ۵) نزدیک هم بود اما برای سیلت و رس مقادیر این شاخص‌های RMSE و MAE افزایش و مقدار R² کاهش داشته است. به نظر می‌رسد به دلیل استفاده از تعداد کمتری داده مشاهداتی (۳۰ نمونه) برای واسنجی مدل، کارایی مدل برای پیش‌بینی سیلت و رس خاک کم شده است. اما همچنان متغیرهای

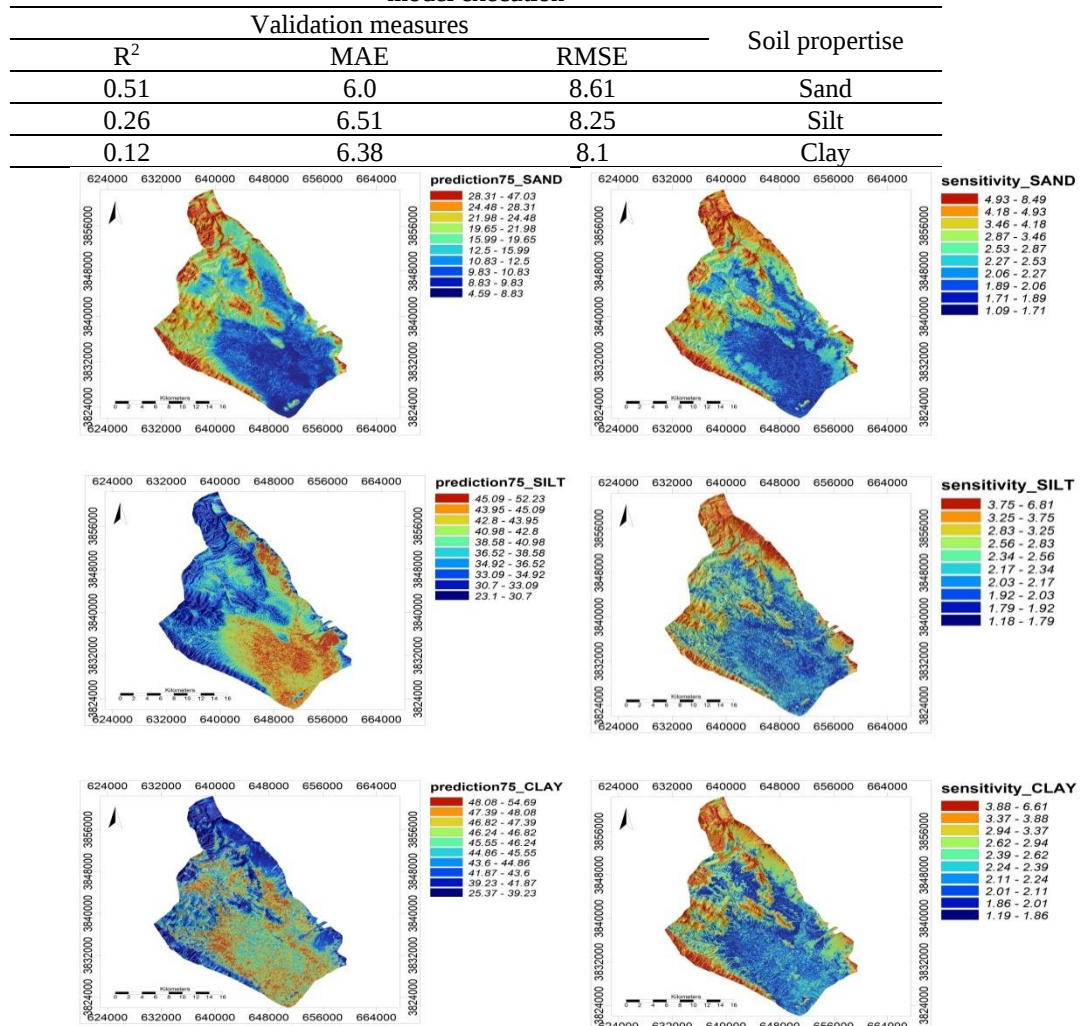
تخمین زننده توان کمتری برای توضیح تغییرات این دو ویژگی خاک در منطقه نسبت به شن دارند. نقشه انحراف معیار بر اساس ۵۰ مدل تحقق یافته و ۲۵ درصد داده‌ها به عنوان نقشه حساسیت مدل به داده‌های موجود، در شکل نشان داده شده است. برای هر سه ویژگی مقدار انحراف معیار در اراضی کوهستانی و تپه ماهوری بالاتر از اراضی با شیب کم و دشت‌ها می‌باشد که بیانگر عدم قطعیت یا حساسیت بیشتر مدل به تغییرات داده در این مناطق می‌باشد. به نظر می‌رسد دلیل آن تعداد نقاط مشاهداتی کمتر در اراضی کوهستانی است از طرفی دامنه انحراف معیار در نقشه شن خاک از ۱ تا ۸/۵ متغیر بوده که نسبت به مقدار دامنه انحراف استاندارد در رس (۱/۲ تا ۶/۶) و سیلت (۱/۲

تغییرات داده‌های موجود برای هر سه ویژگی شن، سیلت و رس خاک حدود ۲۵ درصد از کل عدم قطعیت را در بر می‌گیرد. شکل ۴- سمت چپ، نقشه پیش بینی شن، سیلت و رس خاک بر اساس ۷۵٪ کل داده‌ها را نشان می‌دهد. رنگ قهوه‌ای بر روی این نقشه‌ها بیان می‌کند مقدار شن نسبت به سیلت و رس خاک در اراضی مرتفع بیشتر از دشت‌ها است.

تا ۶/۸ بیشتر بوده که نشان می‌دهد تعداد نمونه مشاهداتی کمتری در بخشی از مواد مادری منطقه که منشأ شن خاک-ها هستند، اخذ شده است (شکل ۴- سمت راست). در اراضی مسطح و با شیب کم که اراضی کشاورزی (دیم و آبی) در آنها غالب است مقدار انحراف معیار کمتر (عدم قطعیت کمتر) بوده که دلیل آن می‌تواند ناشی از تعداد نقاط مشاهداتی بیشتر و یکنواختی مواد مادری آبرفتی در دشت-ها باشد. جدول ۶ نشان می‌دهد که عدم قطعیت ناشی از

جدول ۵- نتایج اعتبارسنجی مدل جنگل تصادفی بر اساس ۷۵ درصد داده‌های مشاهداتی و ۵۰ بار اجرای مدل

Table 5. validation statistics of the random forest model based on 75% of observational data and 50 times of model execution



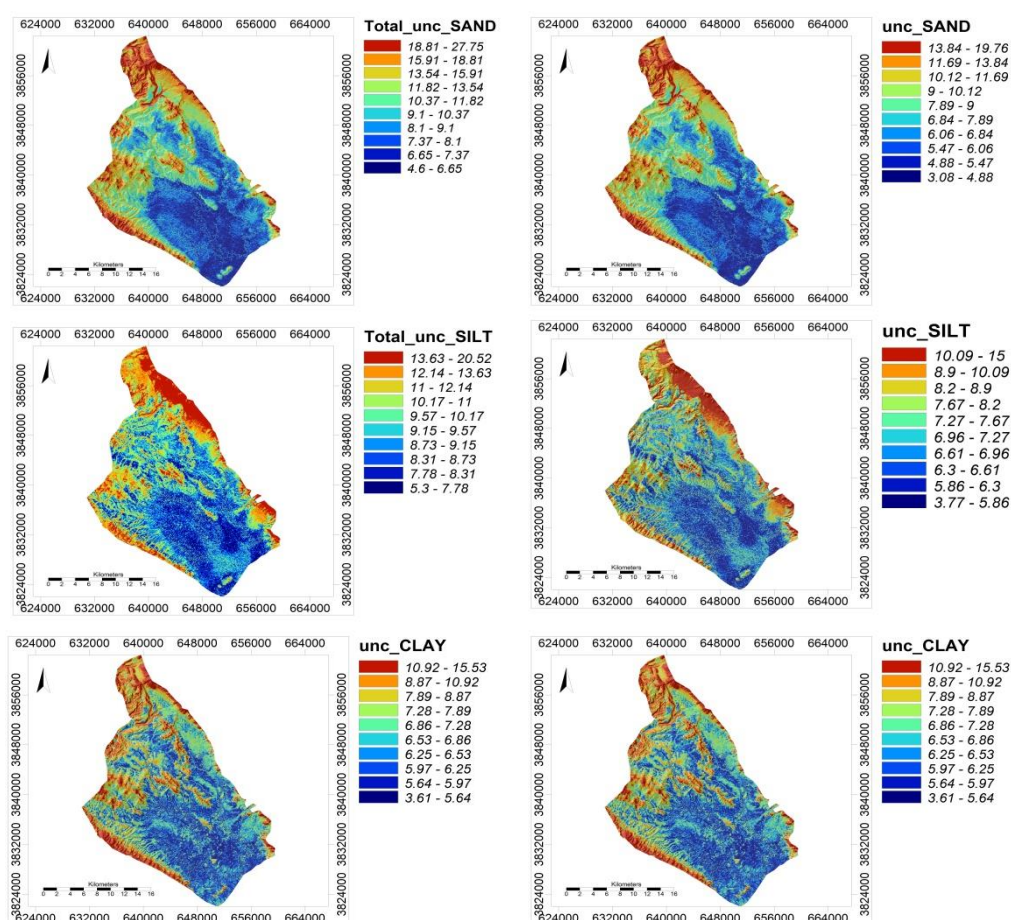
شکل ۴- سمت راست، از بالا به پایین نقشه حساسیت مدل به تغییرات داده‌های شن، سیلت و رس خاک حاصل از ۲۵٪ داده‌ها و اجرای ۵۰ بار اجرای مدل جنگل تصادفی، سمت چپ از بالا به پایین نقشه تخمین داده‌های شن، سیلت و رس خاک با استفاده از مدل جنگل تصادفی بر اساس ۷۵٪ کل داده‌ها

Figure 4. Right - from top to bottom, the sensitivity map of the model to the variations in sand, silt and clay soil data obtained from 25% of the data and 50 times of running the random forest model, Left- from top to bottom, predictive maps of soil sand, silt and clay data using random forest model based on 75% of the all data

نقشه عدم قطعیت مدل

در شکل ۵- سمت راست، عدم قطعیت مدل یا نقشه انحراف معیار مدل برای هر سه ویژگی شن، سیلت و رس ارائه شده- است. در این نقشه‌ها مشخص است که دامنه تغییرات انحراف معیار مدل در رس و سیلت مشابه بوده (۳/۵ تا ۱۵) اما در مورد شن از ۳ تا ۱۹/۸ متغیر است. در اینجا شن، عدم قطعیت بیشتری در برخی نواحی نقشه را نشان می- دهد. مقایسه عدم قطعیت ناشی از تغییرات داده‌های موجود با عدم قطعیت مدل نشان داد که عدم قطعیت مدل هر سه ویژگی شن، سیلت و رس خاک نسبت بیشتری از عدم قطعیت کل (حدود ۷۵٪) را تشکیل می‌دهد (جدول ۶) و در نتیجه سهم بیشتری در افزایش عدم قطعیت کل را دارد

(شکل ۵- سمت چپ). به نظر می‌رسد در بالا بودن دامنه عدم قطعیت شن خاک علاوه بر تعداد نمونه کم، نقش توزیع ناکافی نمونه‌های خاک در فضای مشخصه متغیرهای کمکی خاک را نباید نادیده گرفت. رنگ قهوه‌ای در تمام نقشه‌ها نشان‌دهنده عدم قطعیت بالا است که با اراضی کوهستانی و تپه ماهوری منطقه تطابق دارد (شکل ۵) و توزیع ناکافی نمونه‌های خاک در این سیمای اراضی مشهود است (شکل ۱). از طرفی نرمال نبودن داده‌های شن خاک نشان می‌دهد که از خاک‌های دارای شن زیاد نمونه اندکی اخذ شده است (شکل ۲) و به نظر می‌رسد این هم یکی از دلایل افزایش عدم قطعیت مدل است.



شکل ۵- سمت راست- از بالا به پایین، نقشه عدم قطعیت مدل در تخمین شن، سیلت و رس با استفاده از جنگل رگرسیون چندک، سمت چپ- از بالا به پایین عدم قطعیت کل در تخمین شن، سیلت و رس

Figure 5. Right- from top to bottom, the model uncertainty map in the estimation of sand, silt and clay using the quantile regression forest, Left- from top to bottom, the total uncertainty of sand, silt and clay estimation

دهد. اما در کمی‌سازی عدم قطعیت، موقعیت پیش‌بینی‌های غیرواقعی نشان داده می‌شود که با عدم قطعیت بزرگ مشخص می‌شود. بنابراین، با وجود این که ضریب تبیین

وادوکس و همکاران (Wadoux et al, 2020) بیان کردند در حالی که نتایج اعتبارسنجی متقابل ممکن است توافق قوی بین ویژگی خاک پیش‌بینی شده و اندازه‌گیری شده را نشان

نشان‌دهنده اهمیت تهیه نقشه عدم قطعیت در مطالعات نقشه‌برداری رقومی خاک می‌باشد که بر اساس آن می‌توان به راهکارهایی برای کاهش عدم قطعیت نقشه‌ها دست یافت.

شن ($R^2=0/51$) از سیلت و رس خاک ($R^2=0/26$) و ($R^2=0/12$) بالاتر است اما دامنه عدم قطعیت شن، هم بالاتر از دامنه عدم قطعیت این دو ویژگی خاک است که

جدول ۶- آماره‌ی درصد عدم قطعیت تغییرات داده‌های موجود و درصد عدم قطعیت مدل نسبت به عدم قطعیت کل
Table 6. The Statistics of percentage of uncertainty of variations in the available data and the percentage of uncertainty of the model related to the total uncertainty

Ratio of uncertainty of data variations and model uncertainty to total uncertainty	Min	Max	Mean
sensitivity_SAND/Total_unc_SAND*100	15.8	39.3	25.7
unc_SAND/Total_unc_SAND*100	60.7	84.7	74.3
sensitivity_SILT/Total_unc_SILT*100	13.8	39.6	24.8
unc_SILT/Total_unc_SILT*100	60.4	86.2	75.2
sensitivity_CLAY/Total_unc_CLAY*100	11.7	40.7	26.1
unc_CLAY/Total_unc_CLAY*100	59.3	88.3	73.9

نتیجه‌گیری کلی

خاک تخمینی از توزیع ویژگی‌های خاک در منطقه مورد مطالعه را نشان می‌دهند اما این تخمین‌ها عنصری از عدم قطعیت را همراه خود دارند که به طور مساوی در منطقه تحت پوشش نقشه‌برداری رقومی خاک توزیع نشده است. بنابراین از نتایج عدم قطعیت در انتخاب طرح نمونه‌برداری مناسب می‌توان استفاده نمود. به عبارتی بهتر، محاسبه عدم قطعیت نقشه‌های رقومی خاک می‌تواند اطلاعاتی برای بهبود کیفیت نقشه‌برداری رقومی خاک با بهینه‌سازی طرح نمونه‌برداری در اختیار تولیدکنندگان نقشه‌های رقومی خاک قرار دهد.

نتایج این پژوهش نشان‌دهنده کارایی کم مدل جنگل تصادفی برای تخمین شن خاک با توجه به ریشه میانگین مربعات خطای بالای آن نسبت به سیلت و رس بود. نتایج بررسی دو منبع مختلف عدم قطعیت در تهیه نقشه رقومی اجزا بافت خاک شامل شن، سیلت و رس نشان داد که عدم قطعیت ناشی از تغییرات داده‌های مشاهداتی، در یک چهارم عدم قطعیت کل و عدم قطعیت مدل در حدود سه چهارم عدم قطعیت کل سهم دارد. با توجه به وزن بیشتر عدم قطعیت مدل در این پژوهش، به کاربری مدل‌های با پیچیدگی بیشتر و پارامتر کمتر توأم با استفاده از داده‌های مشاهداتی کافی می‌تواند در کاهش عدم قطعیت نقشه‌های رقومی تولید شده موثر باشد. به طور کلی، نقشه‌های رقومی

References

- Akpa, S. I. C., Odeh, I. O. A., Bishop, T. F. A. and Hartemink, A. E. 2014. Digital mapping of soil particle-size fractions for Nigeria. *Soil Science Society of America Journal*, 78: 1953–1966.
- Behrens, T., Schmidt, K., and Scholten, T. 2008. An approach to removing uncertainties in nominal environmental covariates and soil class maps. In: Hartemink A., McBratney A. and Mendoca-Santos, M.L. (Eds.), *Digital Soil Mapping with Limited Data*. Springer, pp. 213–224.
- Bishop T.F.A., Minasny B. and McBratney A.B. 2006. Uncertainty analysis for soil-terrain models. *International Journal of Geographical Information Science*, 20(2): 117–134.
- Bouyoucos G.J. (1962). Hydrometer method improved for making particle size analyses of soils. *Agronomy Journal*, 54(5): 464-465
- Breiman L. 2001. Random forests. *Machine Learning*, 45(1): 5-32
- Cressie N. and Kornak J. 2003. Spatial statistics in the presence of location error with an application to remote sensing of the environment. *Statistical Science*, 18(4): 436–456.
- Conrad O., Bechtel B., Bock M., Dietrich H., Fischer E., Gerlitz L., Wehberg J., Wichmann V. and Böhner, J. 2015. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. *Geoscientific Model Development Discussions*, (8): 2271-2312.
- Gomez C., Drost, A.P.A. and Roger, J.M. 2015. Analysis of the uncertainties affecting predictions of clay contents from VNIR/SWIR hyperspectral data. *Remote Sensing of Environment*, 156: 58-70.
- Heuvelink G.B.M. 2014. Uncertainty quantification of Global Soil Map products. In: Global Soil Map Basis of the global spatial soil information system. Arrouays D., McKenzie N.J., Hempel J., Richer de-Forges A.C. and McBratney A.B., (eds). 335-340.
- Heuvelink G.B.M. 1998. Error propagation in environmental modelling with GIS. Taylor and Francis, London. 144 pp.
- Jensen, J.R. 2005. Introductory digital image processing: A remote sensing perspective, 3rd edition. Pearson Prentice Hall. 296-300, 301-321, 315-316.
- Khosravani P., Baghernejad M., Moosavi A.A. and FallahShams S.R. 2024. Digital Mapping of Soil Texture Particles with Machine Learning Models and Environmental Covariates. *Journal of Water and Soil*, 37(6): 923-942. (In Persian)
- Lagacherie P., McBratney A.B. and Volz, M. 2007. Digital soil mapping: An introductory perspective. Elsevier, Amsterdam.
- Macmillan R.A. Jones, R.K., & McNabb, D.2004. Defining a hierarchy of spatial entities for environmental analysis and modeling using digital elevation models (DEMs). *Computers, Environment and Urban Systems*, 28:175-200.
- Meinshausen N. 2006. Quantile regression forests. *Journal of Machine Learning Research*, 7: 983-999.
- Nikou M. and Tziachris, P. 2022. Prediction and Uncertainty Capabilities of Quantile Regression Forests in Estimating Spatial Distribution of Soil Organic Matter. *ISPRS International Journal of Geo-Information*, 11: 130.
- Poggio L., Gimona A. and Mark B. 2016. Bayesian spatial modelling of soil properties and their uncertainty: The example of soil organic matter in Scotland using R-INLA. *Geoderma*, 277: 69 - 82.
- Ramcharan A., Hengl T., Nauman T., Brungard C., Waltman S., Wills S. and Thompson J. 2017. Soil Property and Class Maps of the Conterminous US at 100-meter Spatial Resolution based on a

- Compilation of National Soil Point Observations and Machine Learning. *Soil Science Society of America Journal*, 82(1):186-201
- Rouse J.W., Hass R.H.J., Schell A. and Deering D.W. 1973. Monitoring vegetation systems in the Great Plains with ERTS. 3th ERTS Symposium. 10-14 Dec. Washington, DC., USA
- Szatmári G. and Pásztor L. 2019. Comparison of various uncertainty modelling approaches based on geostatistics and machine learning algorithms. *Geoderma*, 337: 1329-1340.
- Saurette D., Zhang Y., Ji W., Huq Easher T., Li H., Shi Z., Adamchuk V. and Biswas A. 2020. Three-dimensional digital soil mapping of multiple soil properties at a field-scale using regression kriging. *Geoderma*, 366: 42-53.
- Soil Science Division Staff. 2017. Soil survey manual. C. Ditzler, K. Scheffe, and H.C. Monger (Eds.). USDA Handbook 18. Government Printing Office, Washington, D.C.
- Stumpf F., Schmidt K., Goebes P., Behrens T., Schönbrodt-Stitt S., Wadoux A., Xiang W. and Scholten T. 2017. Uncertainty-guided sampling to improve digital soil maps. *Catena*, 153: 30–38.
- Teixeira D., Marques J., Silva S.D., Vasconcelos V., de Carvalho J.O., Martins E. and Pereira G. 2018. Mapping units based on spatial uncertainty of magnetic susceptibility and clay content. *Catena*, 164.
- Theres L. and Rs S. 2022. Prediction of Soil Properties Using Quantile Regression Forest Machine Learning Algorithm – A Case Study of Salem and Rasipuram Block, Tamil Nadu, India. *International Journal of Environment and Climate Change*, 2530-2553.
- Wadoux A.M.J.C., Minasny B. and McBratney A.B. 2020. Machine learning for digital soil mapping: Applications, challenges and suggested solutions. *Earth-Science*, 210: 33-59.
- Wang D. Wang P., Cong W., & Wang P. 2022. Calibrating probabilistic predictions of quantile regression forests with conformal predictive systems. *Pattern Recognition Letters*, 156: 2-3
- Xiao J., Shen Y., Tateishi R. and Bayaer W. 2006. Development of topsoil grain size index for monitoring desertification in arid land using remote sensing. *International Journal of Remote Sensing*, 27: 2411–2422.