



دانشگاه گواران، منابع طبیعی

نشریه پژوهش‌های حفاظت آب و خاک
جلد بیست و چهارم، شماره اول، ۱۳۹۶
<http://jwsc.gau.ac.ir>

کاربرد روش درختان تصمیم‌گیری تصادفی در پیش‌بینی کلاس‌های خاک در اراضی با پستی و بلندی کم (مطالعه موردی: شهرستان هیرمند)

خلیل‌اله میرک‌زهی^۱، * علی شهیریاری^۱، محمدرضا پهلوان‌راد^۲ و ابوالفضل بامری^۱

^۱ گروه علوم و مهندسی خاک، دانشکده آب و خاک، دانشگاه زابل، زابل، ایران، ^۲ بخش تحقیقات خاک و آب، مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی سیستان، سازمان تحقیقات، آموزش و ترویج کشاورزی، زابل، ایران
تاریخ دریافت: ۹۵/۱۰/۱؛ تاریخ پذیرش: ۹۶/۲/۲۶

چکیده

سابقه و هدف: شناسایی و نقشه‌برداری خاک، به‌عنوان روشی برای تعیین الگوی پراکنش خاک، توصیف و نمایش آن به شکل قابل فهم و تفسیر برای کاربران مختلف، پایه و اساس اطلاعات خاک برای مدل‌سازی‌های محیطی می‌باشد. نقشه‌برداری رقوم‌های خاک شامل ایجاد ارتباط بین کلاس‌ها یا خصوصیات خاک با فاکتورهای محیطی دخیل در تشکیل و تکامل خاک با استفاده از مدل‌های ریاضی و آماری است که می‌تواند نقشه‌های خاک با دقت و صحت معین در زمان کم‌تر ایجاد نماید و باعث کاهش هزینه‌های پروژه‌های شناسایی و تهیه نقشه‌های خاک گردد. این پژوهش جهت تهیه نقشه کلاس‌های گروه‌های بزرگ و زیرگروه‌های خاک با استفاده از تکنیک درختان تصمیم‌گیری تصادفی در اراضی شهرستان هیرمند در دشت سیستان انجام گرفت.

مواد و روش‌ها: در این مطالعه ۱۰۸ پروفیل خاک در سطحی حدود ۶۰۰۰۰ هکتار از اراضی شهرستان هیرمند حفر گردید. ۱۶ متغیر محیطی شامل خصوصیات زمین، شاخص شوری و شاخص پوشش گیاهی به‌عنوان تخمین‌گر برای تهیه نقشه خاک، مورد استفاده قرار گرفتند. پس از تعیین گروه‌های بزرگ و زیرگروه‌های خاک، نقشه کلاس‌های خاک با استفاده از روش درختان تصمیم‌گیری تصادفی (RF) تهیه شد. شایان ذکر است که ۸۰ درصد داده در آموزش مدل و ۲۰ درصد برای اعتبارسنجی مستقل استفاده شدند.

یافته‌ها: نتایج مطالعات خاکشناسی نشان داده که خاک‌های تشکیل‌شده در دشت سیستان تکامل زیادی نداشتند و عمدتاً در رده‌های انتی‌سول و اریدی‌سول قرار دارند. بیش‌ترین تعداد خاکرخ در گروه‌های بزرگ مربوط به Typic Torrifluvents و بیش‌ترین تعداد خاکرخ در زیرگروه‌های بزرگ مربوط به Typic Torrifluvents بود. همچنین نتایج روش RF نشان داد که کم‌ترین مقدار خطای تخمین نمونه‌های خارج از سبد در گروه‌های بزرگ و زیرگروه‌های خاک به ترتیب ۴۳/۵۳ و ۵۰/۵۹ بود. نتایج اعتبارسنجی مستقل نشان داد که بهترین دقت به‌دست آمده برای گروه‌های بزرگ و زیرگروه‌های بزرگ خاک به ترتیب ۴۸ و ۵۳ درصد بود. بین متغیرهای مختلف محیطی عمق دره، شاخص همگرایی، شبکه کانال‌ها و شوری در گروه‌های بزرگ خاک و عمق دره، ارتفاع و سطح حوزه در زیرگروه‌های خاک دارای بیش‌ترین اهمیت در تخمین کلاس‌های خاک بودند.

* مسئول مکاتبه: shahriari.ali@uoz.ac.ir

نتیجه‌گیری: نتایج نشان داد که در مناطق خشک مورد مطالعه با پستی و بلندی کم، خاک‌ها عمدتاً جوان هستند و همچنین در این مناطق تنوع خاک کم است. در چنین مناطقی روش نقشه‌برداری رقومی و تکنیک درختان تصمیم‌گیری تصادفی می‌تواند برای پیش‌بینی کلاس‌های خاک و تهیه نقشه‌های خاک بسیار مفید بوده و مورد استفاده قرار گیرد.

واژه‌های کلیدی: نقشه‌برداری رقومی خاک، تکنیک درختان تصمیم‌گیری تصادفی، دقت نقشه، مناطق خشک، دشت سیستان

مقدمه

نقشه خاک از مهم‌ترین منابع اطلاعاتی است که برای ارزیابی اراضی و آمایش سرزمین به کار برده می‌شود. بسیاری از اراضی کشور نقشه‌برداری نشده‌اند و مناطقی هم که دارای نقشه هستند، مدت زمان زیادی از تهیه آن‌ها می‌گذرد (۲۹). امروزه نقشه‌برداری خاک با روش‌های مرسوم در کشورهای توسعه‌یافته تقریباً به اتمام رسیده و اطلاعات پایه به صورت رقومی جهت استفاده و پردازش با روش‌های نوین در دسترس است (۲۳). روش‌های مرسوم نقشه‌برداری وقت‌گیر بوده و هزینه بالایی نیاز دارند، به همین منظور یافتن روش‌های دیگر نقشه‌برداری که زمان و هزینه کم‌تری داشته و از دقت قابل‌قبولی نیز برخوردار باشند، ضروری به نظر می‌رسد. استفاده از روش‌های پدومتری در تخمین توزیع مکانی داده‌های اندازه‌گیری شده خاک به دلیل در نظر گرفتن پیوستگی مکانی داده‌ها نسبت به روش‌های معمولی آماری دارای دقت بیشتری است (۱۳). نقشه‌برداری رقومی خاک (DSM)^۱ به‌عنوان زیرمجموعه‌ای از روش‌های پدومتری می‌تواند جایگزین روش‌های معمولی نقشه‌برداری گردد و همچنین در هزینه و زمان صرفه‌جویی نماید (۱۲). نقشه‌برداری رقومی خاک بیانگر مجموعه‌ای از محاسبات رایانه‌ای برای پیش‌بینی پراکنش خاک‌ها در سیمای سرزمین می‌باشد که همگام با پیشرفت‌های به‌وجود آمده در پردازش اطلاعات، تکامل یافته است

و با استفاده از مدل‌های مختلف، اقدام به ساده‌سازی پیچیدگی‌های موجود در سامانه طبیعی خاک می‌نماید (۲۴). نقشه‌برداری رقومی خاک شامل روش‌ها و مدل‌هایی است که بین توزیع خاک (کلاس‌ها یا ویژگی‌های خاک) و داده‌هایی که به آسانی و با قیمت ارزان از طریق روش‌های سنجش‌ازدور، تصاویر و عکس‌های ماهواره‌ای و داده‌های ژئومورفومتری به‌دست و تحت عنوان متغیرهای کمکی محیطی^۲ نامیده می‌شوند ارتباط برقرار می‌کند. اساس نقشه‌برداری رقومی معادله تشکیل خاک (۲۰) و معادله کمی شده آن (۲۳) می‌باشد. روش‌های فراوان و مختلفی از مدل‌های ماشین یادگیری در این روش نقشه‌برداری خاک استفاده می‌شوند. این‌ها شامل رگرسیون لجستیک (۱۶، ۱۹) توابع پدوترانسفر^۳، زمین آمار، مدل‌های آماری تجربی، طبقه‌بندی فازی، تصمیم‌گیری درختی^۴ درختان تصمیم‌گیری تصادفی (۲، ۲۷، ۳۱، ۳۵) و شبکه‌های عصبی مصنوعی هستند (۳، ۱۸، ۲۵). تکنیک ماشین‌های یادگیری همچنین برای پیش‌بینی عمق کلاس‌های خاک (۵)، در قسمت بیولوژیکی کلاس‌های پوسته خاک (۷)، کلاس‌های زهکشی خاک (۱۱، ۲۲) و افق‌های مشخصه خاک (۱۹) به کار می‌روند. هدف از ماشین‌های یادگیری پیدا کردن یک عملکرد تقریباً مفید است که زمینه پیش‌بینی ارتباط بین متغیرهای ورودی و نتایج موردنظر را نشان دهد

2- Environmental covariates
3- Pedotransfer
4- Decision tree

1- Digital Soil Mapping

مورد استفاده قرار می‌گیرند و بنابراین نیازی به اعتبارسنجی مستقل در این روش نمی‌باشد. تکنیک RF همچنین اهمیت متغیرها در مدل‌سازی را نیز تعیین می‌کند. برونگارد و همکاران (۲۰۱۵) به پیش‌بینی کلاس‌های خاک با استفاده از ماشین‌های یادگیری پرداختند، هدف از این مطالعه مقایسه مدل‌های ماشین‌های یادگیری متعدد و مجموعه‌ای از متغیرهای کمکی در پیش‌بینی کلاس‌های خاک در سه حوزه متمایز جغرافیایی در قسمت نیمه‌خشک غرب ایالت متحده آمریکا (جنوب نیومکزیکو، جنوب غربی یوتا و شمال شرقی وایومینگ) از هر سه حوزه به‌عنوان مطالعات نقشه‌برداری خاک استفاده شد. مدل‌های ماشین یادگیری به سه گروه بر اساس پیچیدگی مدل تقسیم شدند: ساده، متوسط و پیچیده. به‌طور کلی، مدل‌های پیچیده به‌طور مداوم دقیق‌تر از مدل‌های ساده و یا نسبتاً پیچیده بود و به این نتیجه رسیدند که RF به‌عنوان دقیق‌ترین مدل در بین مناطق مورد مطالعه بود و مدل‌های پیچیده با استفاده از متغیرهای کمکی انتخاب با حذف ویژگی‌های بازگشتی منجر به دقیق‌ترین پیش‌بینی می‌شود (۹). زمانی که توزیع فراوانی مشاهدات خاک تقریباً برابر باشد و چند کلاس وجود داشته باشد، مدل‌های ماشین‌های یادگیری دقیق‌ترین مدل‌ها هستند. برونگارد (۲۰۰۹) RF را برای پیش‌بینی مکانی کلاس‌ها و خصوصیات خاک استفاده کرد، وی دریافت که RF یک روش مفید برای پیش‌بینی کلاس‌های خاک، خصوصیات خاک و توزیع پوشش گیاهی است (۸). لایب و همکاران (۲۰۱۲) RF و RT^۵ را برای پیش‌بینی توزیع بافت خاک استفاده کرده و مشاهده کردند که روش RF دارای دقت بالاتری نسبت به RT بود (۲۱). برونگارد (۲۰۰۹) RF را برای پیش‌بینی مکانی کلاس‌ها و خصوصیات خاک استفاده کرد؛ وی

(۱۴). ژیونگ و همکاران (۲۰۱۲) از حذف ویژگی‌های بازگشتی^۱ برای شناسایی عوامل مهم برای نقشه‌برداری کربن آلی خاک در فلوریدا استفاده کردند (۳۹). پهلوان‌راد و همکاران (۲۰۱۴)، از نقشه‌برداری رقومی برای بروزرسانی نقشه‌های خاک استفاده کردند (۲۷). نتایج آنان نشان داد نقشه بروز شده سری خاک ۱۳/۴ درصد دقیق‌تر از نقشه سری خاک معمولی است. جعفری و همکاران (۲۰۱۲) روش رگرسیون دوتایی^۲ و رگرسیون چندمتغیره^۳ را با ۱۲۶ خاکرخ در سطح ۹۰۰۰۰ هکتار، برای پیش‌بینی توزیع گروه‌های بزرگ در منطقه زرنند کرمان مورد استفاده قرار دادند (۱۹). تقی‌زاده مهرجردی و همکاران (۲۰۱۴) روش تصمیم‌گیری درختی را برای پیش‌بینی مکانی زیرگروه‌های بزرگ خاک در اردکان یزد به‌کار بردند (۳۶). ور و همکاران (۲۰۱۵) از روش‌های مختلف تهیه نقشه رقومی خاک برای نقشه‌برداری ذخیره کربن آلی خاک، استفاده نمودند. نتایج آن‌ها دقت بالای نقشه‌های تهیه شده با این روش را نشان داد و تأکیدی بر امکان استفاده از روش‌های مختلف نقشه‌برداری رقومی در تهیه نقشه ویژگی‌های خاک بود (۳۷). یکی از این مدل‌های جدید مورد استفاده در نقشه‌برداری رقومی خاک، روش تکنیک درختان تصمیم‌گیری تصادفی (RF) است. مدل درختان تصمیم‌گیری تصادفی مدل توسعه‌یافته از مدل طبقه‌بندی و رگرسیون درختی (CART)^۴ می‌باشد. در روش RF به‌جای رشد یک درخت هزاران درخت طبقه‌بندی تولید می‌شود (۶). در این روش یک سوم داده‌های اصلی در ایجاد هر درخت استفاده نمی‌شوند که تحت عنوان نمونه‌های بیرون از سبد (OOB) نامیده می‌شوند. چون این نمونه‌ها در آموزش درخت استفاده نشده‌اند در تست

- 1- Recursive feature elimination
- 2- Binary logistic regression
- 3- Multinomial logistic regression
- 4- Classification and regression trees

5- Regression tree

اراضی شهرستان هیرمند در دشت سیستان در حد فاصل طول‌های جغرافیایی $61^{\circ} 45'$ تا $61^{\circ} 50'$ شرقی و عرض‌های جغرافیایی $30^{\circ} 58'$ تا $31^{\circ} 18'$ شمالی واقع شده است (شکل ۱). رژیم رطوبتی منطقه مورد مطالعه اریدیک و رژیم دمایی آن هایپرترمیک و مواد مادری آن آبرفتی و جزء دشت‌های سیلابی محسوب می‌شود. اراضی دشت سیستان متأثر از رسوبات رودخانه هیرمند و انشعابات آن می‌باشد و از لحاظ تکامل خاکرخی جزء خاک‌های جوان به‌شمار می‌رود. منطقه تقریباً مسطح و با شیب یک تا دو درصد می‌باشد. شیب عمومی منطقه از سمت جنوب به سمت شمال منطقه مطالعاتی که دریاچه خشکیده هامون قرار دارد، می‌باشد (شکل ۱ سمت راست). عمده اختلاف ارتفاع به‌دلیل وجود تپه‌های شنی موجود در منطقه مطالعاتی می‌باشد. چندین متغیر محیطی که شامل خصوصیات زمین (40) و شاخص پوشش گیاهی و شوری بودند در این مطالعه استفاده شد. تعاریف متغیرهای محیطی در ویلسون و گالانت (2000) و هنگل و رویتر (2008) یافت می‌شود ($38, 15$). خصوصیات زمین شامل شیب، طول شیب، موقعیت نسبی شیب، جهت شیب، اختلاف جزئی شیب، گرادیان شیب پایینی، سطح ویژه حوزه، شبکه کانال‌ها، گودشدگی، شاخص همگرایی، انحنای مقطعی، ارتفاع، انحنای طولی، شاخص خیزی، عمق دره و فاصله عمودی تا شبکه کانال از نقشه DEM (شکل ۱، سمت راست) با بزرگنمایی 30 متر با استفاده از نرم‌افزارهای ArcGIS 10.3 و SAGA 2.1.4_Win استخراج گردیدند. جدول ۱ خصوصیات زمین و فاکتورهای خاکسازي را نشان می‌دهد.

دریافت که RF یک روش مفید برای پیش‌بینی کلاس‌های خاک، ویژگی‌های خاک و توزیع پوشش گیاهی می‌باشد (۸). بارتولد و همکاران (2013) با استفاده از RF توزیع تیپ‌های مختلف خاک با روش رده‌بندی WRB^۱ را در سطحی حدود 3600 کیلومترمربع با 145 خاکرخ در چین مورد ارزیابی قرار دادند و مشاهده کردند که در منطقه مورد مطالعه آن‌ها اقلیم و کاربری بیش‌ترین تأثیر را بر توزیع تیپ خاک داشته است. در مطالعه آن‌ها مقدار خطای پیش‌بینی مدل برای سطح اول رده‌بندی روش WRB برابر $51/6$ درصد بود (۲). پهلوان‌راد و همکاران (2016) نقشه‌برداری رقومی خاک را در 85000 هکتار از اراضی استان گلستان انجام دادند و دو مدل تکنیک تصمیم‌گیری تصادفی (RF) و رگرسیون لاجیستیک را برای گروه‌های بزرگ، زیرگروه‌های خاک و سری خاک به‌کار بردند. نتایج نشان داد تکنیک درختان تصمیم‌گیری تصادفی دقت بهتری از رگرسیون لاجیستیک دارد (۲۸).

تاکنون مطالعه‌ای در مورد استفاده از روش نقشه‌برداری رقومی خاک و تکنیک RF برای پیش‌بینی کلاس‌های خاک در اراضی مسطح و دشت‌های سیلابی در ایران انجام نشده است. از این‌رو پژوهش حاضر با هدف ارزیابی دقت تکنیک RF در تخمین نقشه کلاس‌های خاک در سطح گروه‌های بزرگ و زیرگروه‌های خاک در اراضی مسطح، در اراضی شهرستان هیرمند در دشت سیستان انجام گرفت.

مواد و روش‌ها

محدوده مورد مطالعه دارای مساحت حدود 60000 هکتار و در استان سیستان و بلوچستان، در

$$NDSI = \frac{Red - NIR}{Red + NIR} \quad (2)$$

در این معادلات NIR و Red مقادیر انعکاس امواج مادون قرمز نزدیک و قرمز هستند.

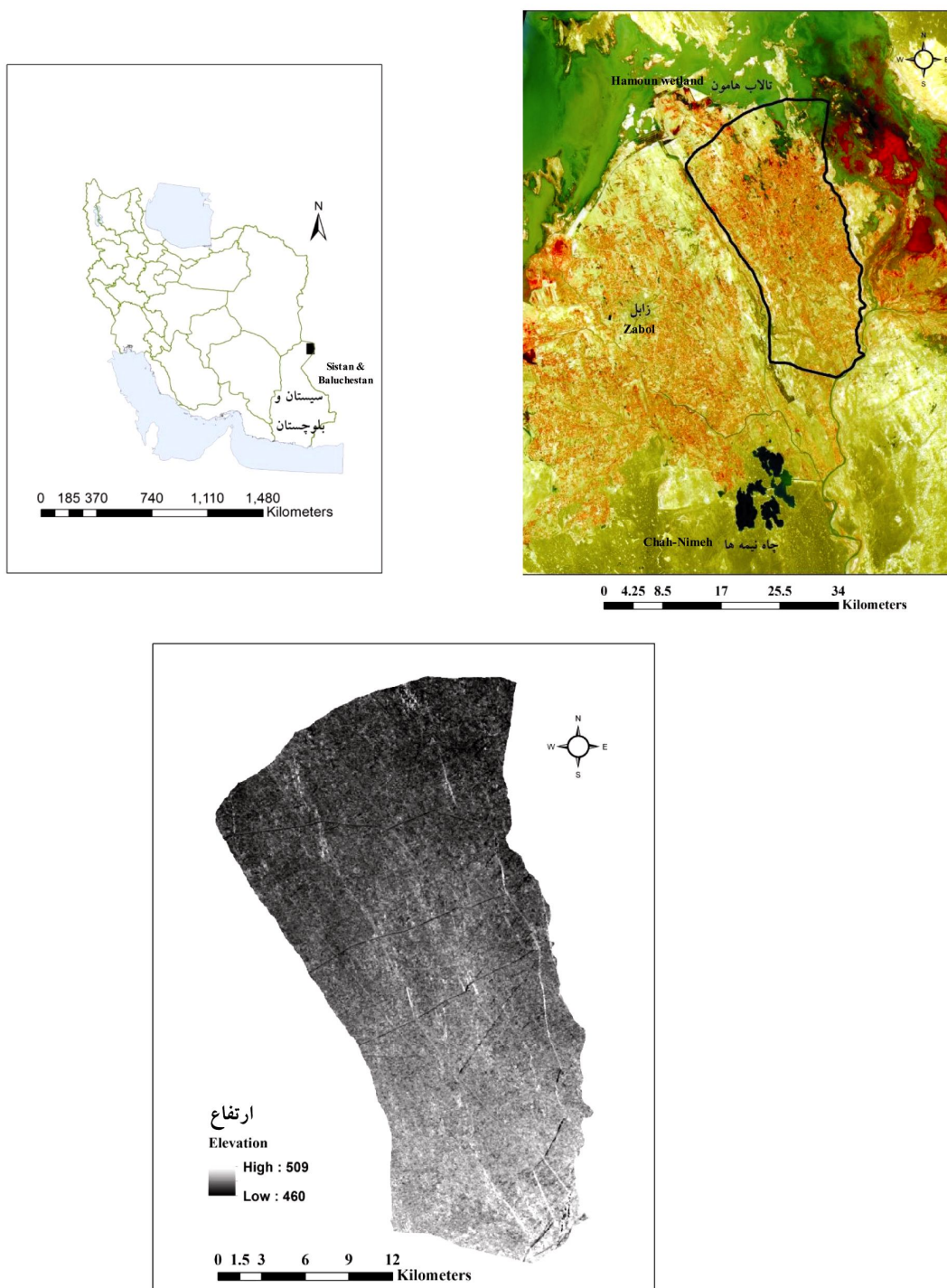
رابطه‌های ۱ و ۲ روش محاسبه NDVI (شاخص پوشش گیاهی) و NDSI (شاخص شوری) را نشان می‌دهند. که با استفاده از تصاویر ماهواره‌ای IRS (۹ فروردین ۱۳۸۵) با استفاده از GIS 10.3 استخراج شدند.

$$NDVI = \frac{NIR - Red}{NIR + Red} \quad (1)$$

جدول ۱- خصوصیات زمین و فاکتورهای خاکساز.

Table 1. Land properties and pedogenic factors.

فاکتورهای خاکساز (Pedogenic factors)	خصوصیات زمین (Land properties)
	شیب (Slope)
	طول شیب (Slope length)
	موقعیت نسبی شیب (Relative slope position)
	جهت شیب (Aspect)
	اختلاف جزئی شیب (Analytical hillshading)
	گرادیان شیب پایینی (Down slope distance gradient)
	سطح ویژه حوزه (Catchment Area)
	سطح پایه شبکه کانال‌ها (Channel network base level)
پستی و بلندی (Relief)	گودشدگی (Closed depression)
	شاخص همگرایی (Convergence Index)
	انحنای مقطعی (Cross-sectional curvature)
	ارتفاع (Elevation)
	انحنای طولی (Longitudinal curvature)
	شاخص خیسگی (Topographic wetness index)
	عمق دره (Valley depth)
	فاصله عمودی به شبکه کانال (Vertical distance to channel network)
موجودات زنده (Organisms)	پوشش گیاهی (Vegetation)



شکل ۱- موقعیت منطقه مورد مطالعه و مدل رقومی ارتفاع آن.

Figure 1. Location of studied area and DEM.

درست متغیرها با مقادیری که به‌طور تصادفی برای هر درخت تولید شده است جایگزین می‌شود و اثر این تغییر را روی طبقه‌بندی اندازه‌گیری می‌کند. اگر این جایگزینی اثری روی خطای اندازه‌گیری نداشته باشد اهمیت آن متغیر کم است و اگر مقدار خطای اندازه‌گیری افزایش یابد آن متغیر مهم می‌باشد (۶).

نتایج و بحث

نتایج نشان داد که خاک‌های تشکیل شده در منطقه مطالعاتی تکامل بالایی نداشتند. با توجه به این‌که مواد مادری در دشت سیستان از مواد آبرفتی رودخانه هیرمند می‌باشند و همچنین به‌دلیل شرایط اقلیمی (رژیم دمای هایپرترمیک و رژیم رطوبتی اریدیک) و فعالیت بادهای معروف به بادهای ۱۲۰ روزه که باعث ایجاد فرایند فرسایش بادی شدید در منطقه است، بنابراین امکان تکامل این خاک‌ها وجود نداشته و خاک‌ها در رده‌های انتی‌سولز و اریدی‌سولز قرار گرفتند. بیول و همکاران (۲۰۱۱) نیز معتقدند که انتی‌سول‌ها مهم‌ترین خاک‌های همراه با اریدی‌سول‌ها هستند (۱۰).

افق‌های شناسایی سطحی بیش‌تر افق اکریک بودند و افق‌های شناسایی زیر سطحی کمبیک و افق سالیک (که در سطح نیز مشاهده می‌شد) مشاهده شد. فرآیندهای خاکسازي در پیشرفته‌ترین حالت به تشکیل ساختمان و یا تغییر رنگ (افق کمبیک) و یا شور شدن خاک (افق سالیک) منتج شده است.

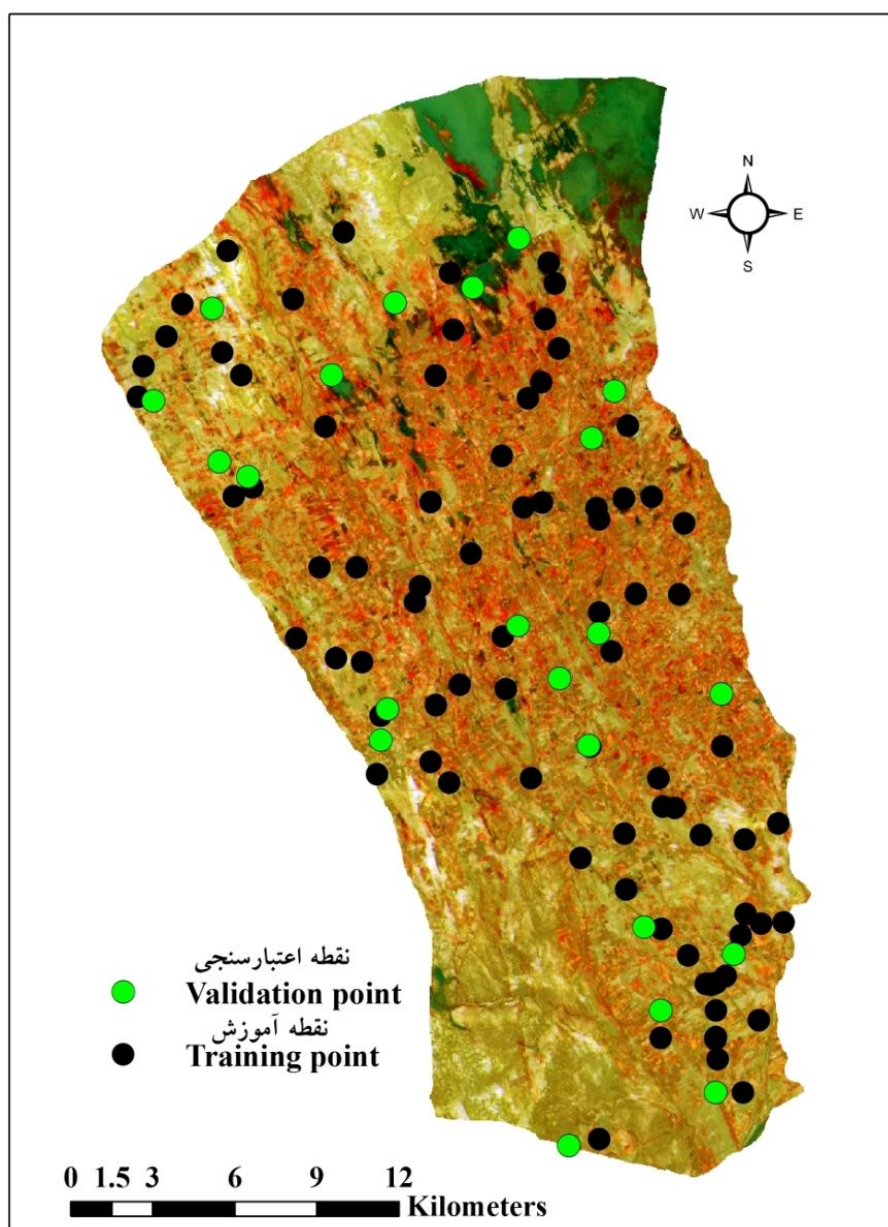
در این مطالعه ۱۰۸ خاک‌رخ به روش نمونه‌برداری تصادفی نظارت‌شده در سری‌های مختلف خاک حفر گردید و خاک‌رخ‌ها رو طوری تقسیم کردیم که کل منطقه مورد مطالعه را دربرگیرند و از این خاک‌رخ‌ها ۸۰ درصد برای آموزش و ۲۰ درصد برای اعتبارسنجی مستقل انتخاب شدند که ۸۵ خاک‌رخ برای آموزش و ۲۳ خاک‌رخ برای اعتبارسنجی مستقل استفاده شد (شکل ۲).

جدول ۲ داده‌های آموزش و اعتبارسنجی برای کلاس‌های مختلف رده‌بندی خاک را نشان می‌دهد. خاک‌رخ‌ها بر اساس راهنمای تشریح خاک‌رخ و نمونه‌برداری صحرائی خاک (۲۰۱۲) تشریح و از افق‌های خاک نمونه‌برداری انجام گرفت (۲۶). سپس خاک‌ها هوا خشک شده و از الک دو میلی‌متری عبور داده شدند و آزمایش‌های فیزیکی و شیمیایی روتین بر روی نمونه‌ها با استفاده از روش‌های استاندارد و متداول انجام گرفت و خاک‌رخ‌ها بر اساس سیستم تاکسونومی (۲۰۱۴) تا سطح گروه بزرگ و زیرگروه رده‌بندی شدند. نتایج رده‌بندی در جدول ۳ ارائه شده است (۳۴).

لایه‌های متغیرهای محیطی با پیکسل‌های ۳۰ متر مربعی و کلاس‌های خاک وارد نرم‌افزار R 3.0.1 و با بسته نرم‌افزاری Random Forest مدل درختان تصمیم‌گیری تصادفی اجرا گردید. اهمیت متغیرهای محیطی و تخمین کلاس‌های خاک در این پژوهش نیز با استفاده از این مدل تعیین شد. RF به دو روش میانگین کاهش دقت^۱ و میانگین کاهش ضریب جینی^۲ اهمیت متغیرها را نشان می‌دهد. در روش میانگین کاهش دقت که در اکثر مطالعات استفاده شده، مقادیر

1- Mean decrease in accuracy

2- Mean decrease in Gini coefficient



شکل ۲- محل نقاط نمونه‌برداری شده در تصویر کاذب منطقه مطالعاتی.

Figure 2. Location of sampling points in the false image of the study area.

جدول ۲- داده‌های آموزش و اعتبارسنجی برای کلاس‌های مختلف.

Table 2. Training and validation data for various classes.

اعتبارسنجی Validation	آموزش Training	زیرگروه Subgroup	اعتبارسنجی مستقل Validation	آموزش Training	گروه بزرگ Great Group
11	42	Typic Torrifluvents	11	42	Torrifluvents
7	24	Fluentic Haplocambids	7	29	Haplocambids
1	4	Sodic Haplocambids	5	14	Haplosalids
5	14	Typic Haplosalids			

جدول ۳- رده‌بندی خاک خاکرخ‌های مورد مطالعه.

Table 3. Soil classification of studied profiles.

تعداد	زیرگروه	تعداد	گروه بزرگ	تعداد	زیررده	تعداد	رده
N	Subgroup	n	Great Group	N	Suborder	n	Order
53	Typic Torrifuvents	53	Torrifuvents	53	Fluvents	53	Entisol
31	Fluventic Haplocambids	36	Haplocambids	36	Cambids	55	Aridisol
19	Typic Haplosalids	19	Haplosalids	19	Salids		
5	Sodic Haplocambids						

دقیق‌تری انجام شود و این شرایط تفسیر نتایج و تهیه نقشه‌ها را پیچیده می‌کند (۱۰).

اجرای مدل: نتایج اجرای مدل نشان داد که کم‌ترین مقدار خطای تخمین نمونه‌های OOB (نمونه‌های خارج از سبد) در گروه‌های بزرگ ۴۳/۵۳ درصد که با تعداد ۸۰۰ درخت و ۶ متغیر در هر درخت به دست آمد. کم‌ترین مقدار خطای تخمین نمونه‌های OOB (نمونه‌های خارج از سبد) برای زیرگروه‌های بزرگ خاک، ۵۰/۵۹ درصد بود که با تعداد ۳۰۰ درخت و ۷ متغیر در هر درخت به دست آمد. بارتولد و همکاران (۲۰۱۳) به مقدار خطای تخمین OOB ۵۱ درصد در پیش‌بینی کلاس‌های خاک در سطح اول روش WRB دست یافتند (۲). استام و همکاران (۲۰۱۰) با ۶۷۲ مشاهده خاک‌رخ در سطحی معادل ۴۷۰۰۰ هکتار، مقدار خطای تخمین OOB به دست آمده در سطح فامیل خاک را برای مدل RF، ۵۵/۲ درصد گزارش کردند (۳۵). راکر و همکاران (۲۰۱۰) در پیش‌بینی کلاس خاک با روش RF مقدار خطای ۵۱ درصد را گزارش کردند (۳۲). برون‌گارد (۲۰۰۹) با ۳۰۰ نمونه خاکرخ خاک در سطح ۳۰۰۰۰ هکتار در سطح زیرگروه بزرگ خاک مقدار خطای تخمین ۶۲ درصد را به دست آوردند (۸). پهلوان‌راد و همکاران (۲۰۱۴) در مطالعه خود در بخشی از اراضی استان گلستان از

شاتز و اندرسون (۲۰۰۵) معتقدند که تجمع نمک‌های طبیعی سدیم و منیزیم در سطح خاک در مناطق خشک تحت‌تأثیر فرایند شور شدن^۱ باعث ایجاد افق‌های سالیک شده و عمدتاً در زیررده Salids رده‌بندی می‌شوند (۳۳).

جدول ۲ کلاس‌های خاک و تعداد خاکرخ‌ها را در گروه‌های بزرگ و زیرگروه‌های بزرگ در منطقه مورد مطالعه نشان می‌دهد. بر اساس نتایج جدول بیش‌ترین تعداد خاکرخ در گروه‌های بزرگ مربوط به Torrifuvents، و بیش‌ترین تعداد خاکرخ در زیرگروه‌های بزرگ مربوط به Typic Torrifuvents می‌باشد که بیش از نصف خاکرخ‌ها را تشکیل می‌دهند. در مناطقی که دارای رژیم اریدیک بوده و خاک‌ها دارای افق‌های شناسایی زیرسطحی هستند، عمدتاً خاک‌ها در رده اریدی‌سولز قرار می‌گیرند و همچنین خاک‌های تشکیل شده بر روی لندفرم‌های آبرفتی جوان مانند دشت سیلابی، عمدتاً در رده انتی‌سولز رده‌بندی می‌شوند (۳۳). بیول و همکاران (۲۰۱۱) بیان کردند که به‌دلیل حضور انقطاع‌های سنگی غیرقابل پیش‌بینی که در اثر سیلاب‌های مختلف و با شدت‌های متفاوت در زیر رده Fluvents به‌وجود می‌آیند، در زمان نقشه‌برداری خاک در این خاک‌ها باید بازدیدهای و عملیات صحرائی بیش‌تر و

1- Salinization

منطقه مورد مطالعه به ترتیب در شکل‌های ۳ و ۴ نشان داده شده است. همان‌طور که در شکل‌ها نشان داده شده است بیش‌ترین سطح اراضی در گروه‌های بزرگ مربوط به Torrifluents و در زیرگروه‌های بزرگ مربوط به Typic Torrifluents است و پس از آن Haplocambids و Fluentic Haplocambids قرار دارد. علت فلوونت بودن بیش‌تر کلاس‌ها به‌خاطر نزدیکی بودن به رودخانه مرزی پریان مشترک می‌باشد. خاک‌های این منطقه تحت تأثیر رسوبات این رودخانه قرار دارد و منجر به تشکیل خاک‌های فلوونت با بافت متوسط گردیده است. در مناطق شمالی منطقه، به‌علت نزدیکی به دریاچه هیرمند و رسوب بیش‌تر ذرات رس، افق کمبیک در این مناطق تشکیل شده است. در قسمتی از اراضی افق سالیک دیده می‌شود که به‌دلیل نحوه رسوبگذاری رودخانه و وجود لایه‌های با نفوذپذیری کم که آبشویی نمک را دچار محدودیت می‌کنند، رها شدن اراضی به‌دلیل خشکسالی ۱۸ ساله اخیر و تبخیر زیاد در منطقه می‌باشد.

تکنیک RF جهت تهیه نقشه کلاس‌های خاک در سطح زیرگروه بزرگ خاک استفاده کردند و مقدار خطای ۵۲/۵۳ درصد را به‌دست آوردند و بیان نمودند که روش RF می‌تواند یک روش قابل اعتماد و با دقت مناسب باشد که با تعداد نمونه کم تخمین قابل قبول ارائه می‌دهد (۳۰).

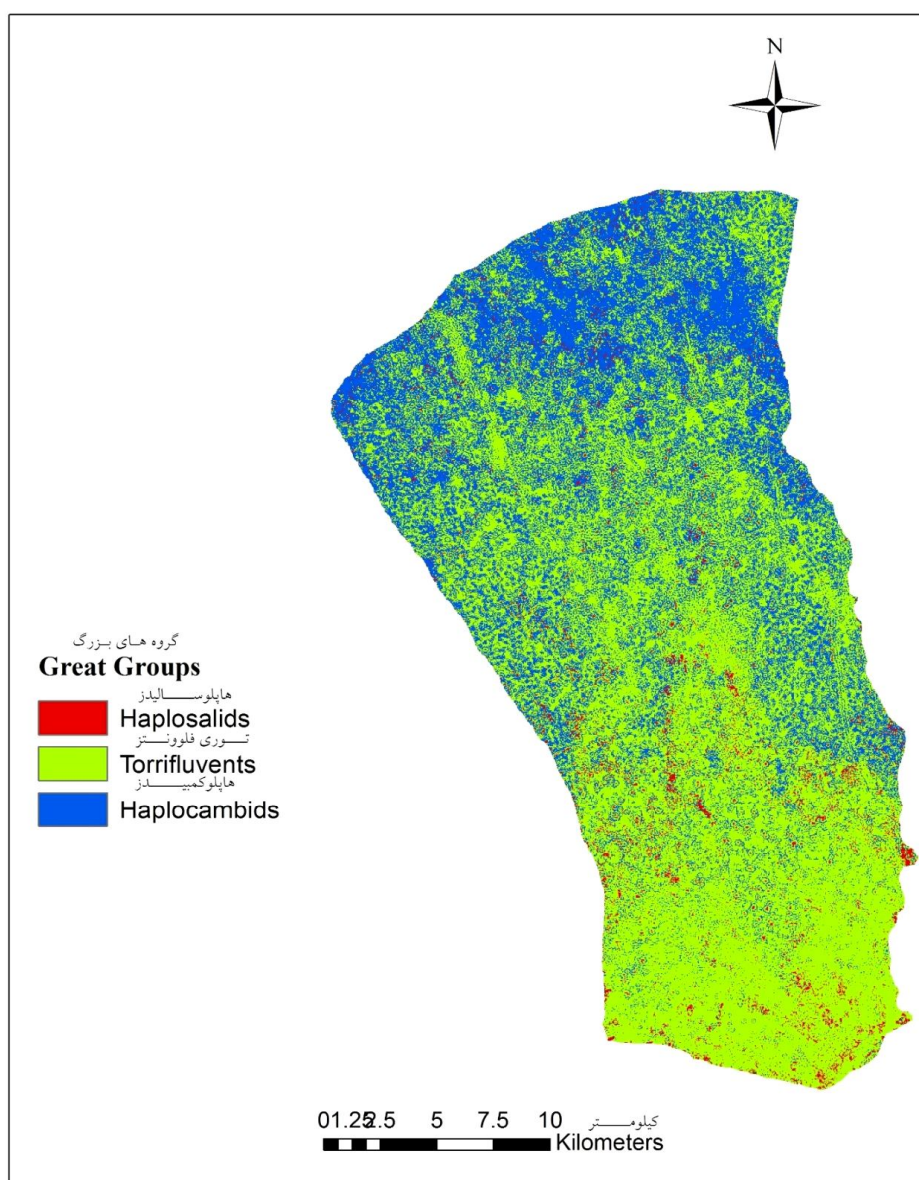
نتایج اعتبارسنجی مستقل نشان داد که بهترین دقت به‌دست آمده برای گروه‌های بزرگ و زیرگروه‌های بزرگ خاک به ترتیب ۴۸ و ۵۳ درصد بود. این نتایج، نشان‌دهنده دقت قابل قبول نقشه‌های تولیدشده در منطقه مطالعاتی است. پهلوان‌راد (۲۰۱۴) در مدل‌سازی کلاس‌های خاک با روش RF در سطح ۸۵۰۰۰ هکتار در استان گلستان با روش اعتبارسنجی مستقل، در سطح گروه‌های بزرگ، دقت کلی ۳۵ تا ۵۰ درصد و در سطح زیرگروه‌های بزرگ، دقت کلی ۱۰ تا ۴۰ درصد را مشاهده کرد (۳۰).

توزیع مکانی گروه‌ها و زیرگروه‌های خاک: نقشه پیش‌بینی گروه‌های بزرگ و زیرگروه‌های خاک در

جدول ۴- خطای کلاس‌های پیش‌بینی شده در گروه‌های بزرگ و زیرگروه‌های بزرگ خاک با استفاده از داده‌های خارج از سید.

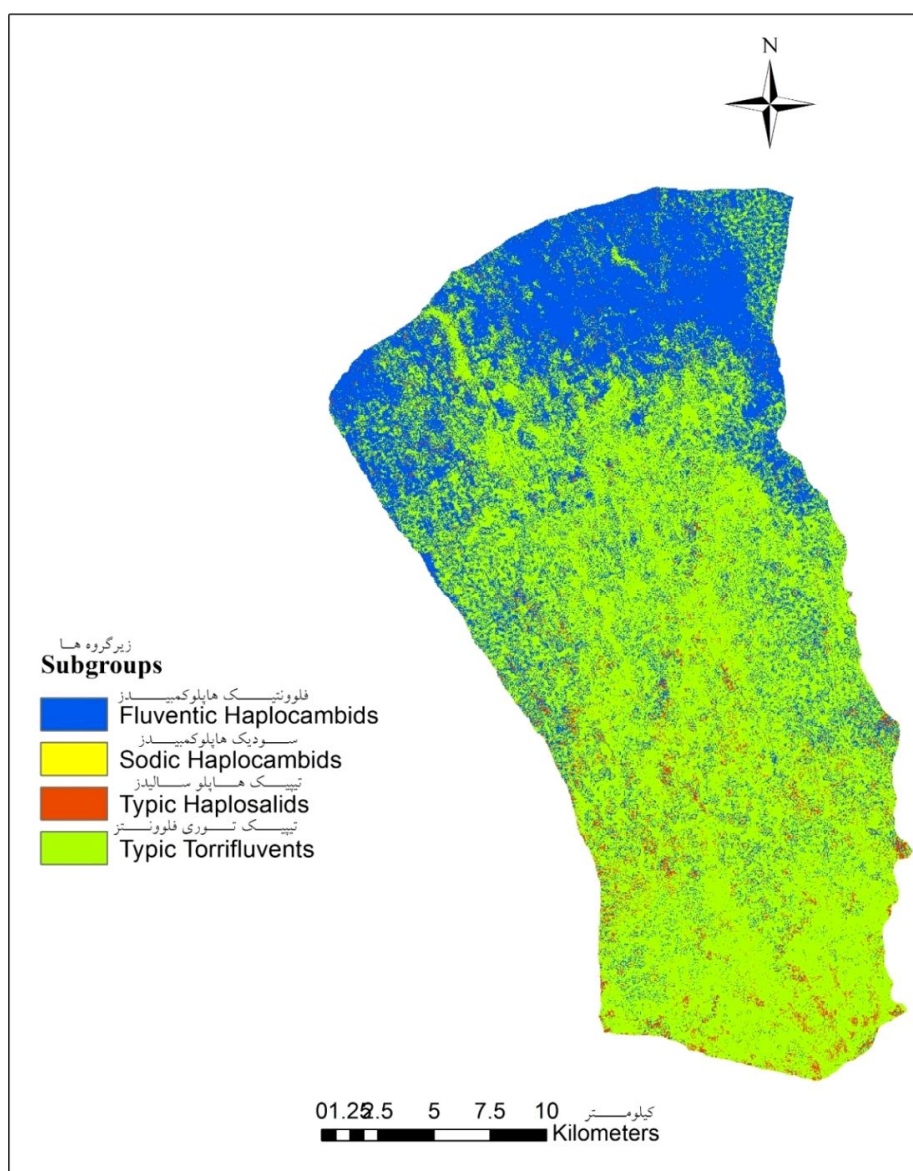
Table 4. Predicting classes error in soil great groups and subgroups by using OOB data.

خطای کلاس Class error	زیرگروه Subgroup	خطای کلاس Class error	گروه بزرگ Great Group
23	Typic Torrifluents	26	Torrifluents
66	Fluentic Haplocambids	48	Haplocambids
92	Typic Haplosalids	92	Haplosalids
100	Sodic Haplocambids		



شکل ۳- نقشه گروه‌های بزرگ خاک در منطقه مطالعاتی.

Figure 3. Soil great groups map of studied area.

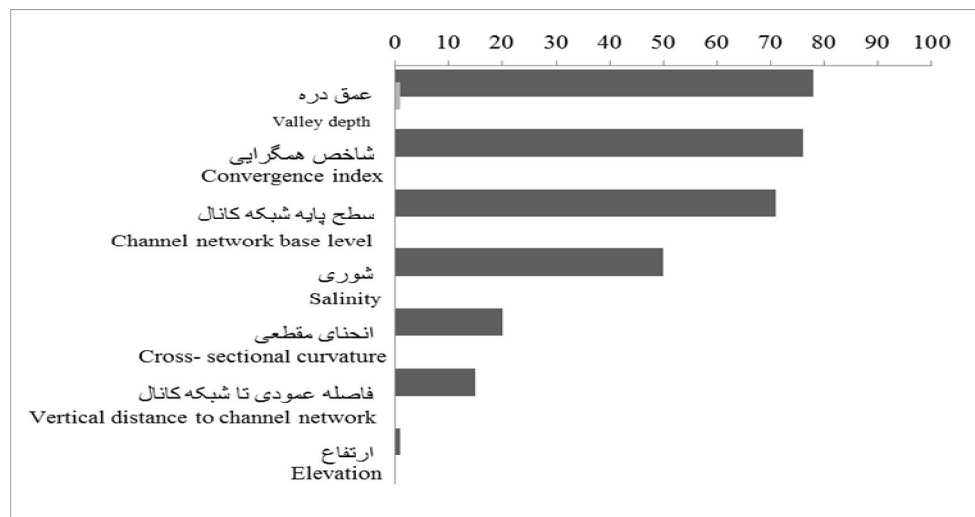


شکل ۴- نقشه زیرگروه‌های خاک در منطقه مطالعاتی.
Figure 4. Soil subgroups map in studied area.

کانال، ارتفاع، موقعیت نسبی شیب و انحناى مقطعى کم‌ترین اثر را در بیان تغییرات خاک نشان دادند (شکل ۵). همچنین اهمیت متغیرها در زیرگروه‌های بزرگ خاک نشان می‌دهد که عمق دره، ارتفاع، سطح ویژه حوزه، انحناى مقطعى بیش‌ترین تأثیر را در بیان تغییرات خاک نشان می‌دهند (شکل ۶). این نتایج بیانگر این است که متغیرهایی مانند عمق دره و سطح پایه شبکه کانال‌ها که متغیرهای مربوط به شبکه‌های توزیع آب در منطقه مطالعاتی هستند، تأثیر به‌سزایی در توزیع مکانی کلاس‌های دارند و مدل RF به‌خوبی توانسته این موضوع را بیان کند.

نتایج تخمین کلاس‌های پیش‌بینی‌شده در جدول ۴ در گروه‌های بزرگ و زیرگروه‌های بزرگ خاک نشان می‌دهد که کلاس‌هایی که تعداد نقاط مشاهداتی بیش‌تری داشتند دارای خطای پیش‌بینی کم‌تری هستند. برای افزایش دقت کلاس‌هایی که دارای دقت کمی هستند می‌توان با افزایش نقاط نمونه‌برداری، دقت کلاس‌ها و در نتیجه دقت نقشه تولیدی را افزایش داد.

نتایج اهمیت متغیرها نشان می‌دهد که عمق دره، شاخص همگرایی، سطح پایه شبکه کانال‌ها، و شوری بیش‌ترین تأثیر را در بیان تغییرات گروه‌های بزرگ خاک داشتند و متغیرهای فاصله عمودی تا شبکه



شکل ۵- مقدار اهمیت متغیرهای استفاده شده در گروه‌های بزرگ خاک.

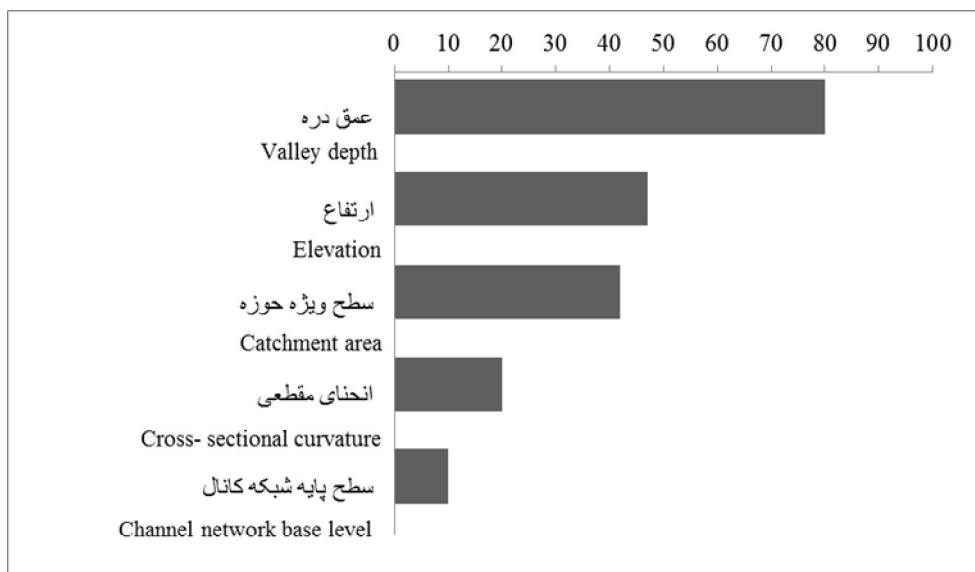
Figure 5. The importance of variables used in soil great groups.

مونتینی (۲۰۱۵) نیز بیان نمودند که در مناطق خشک و بیابانی آبرفتی عوامل مؤثر در سیلاب مانند شبکه کانال‌های سیلاب نقش اصلی در شرایط ژئومورفیک و خصوصیات آن دارند (۱). البته باید توجه داشت که برنس و همکاران (۲۰۱۰) تفاوت ارتفاع و ژئونگ و همکاران (۲۰۱۲) استفاده از متغیرهای کمکی مانند چشم‌انداز آتش و منابع ابزارهای مدیریت پروژه و نقشه پوشش گیاهی و پوشش مکانی زمین به‌عنوان

نتایجی که به آن اشاره شد، با مشاهدات میدانی مطابقت دارد زیرا اراضی منطقه دارای پستی و بلندی کم و شیب یک تا دو درصد هستند و خاک‌های اراضی منطقه بیش‌ترین تأثیر را از رسوبات رودخانه هیرمند و انشعابات دارند. هانگ و همکاران (۲۰۱۴) شاخص فاصله تا نزدیک‌ترین جریان/ رودخانه و شیب هیدرولیکی را به‌عنوان متغیرهای کمکی مربوط به پستی و بلندی استفاده کردند (۱۷). المسراهی و

به‌طور موفق‌تری متغیرهای مؤثر در تشکیل خاک در اراضی مورد مطالعه (دشت‌های سیلابی) را بیان نموده است و برای انتخاب متغیرهای تأثیرگذار در نقشه‌برداری رقومی خاک مورد استفاده قرار گیرد.

متغیرهای کمکی مربوط به پوشش گیاهی را استفاده کردند (۴، ۳۹). این نتایج نقش کلیدی شرایط آبرفتی و عوامل زمینی تحت تأثیر آن‌ها را در تشکیل و تکامل خاک در دشت‌های سیلابی و با پستی و بلندی کم نشان می‌دهد. بنابراین به‌نظر می‌رسد که روش RF



شکل ۶- مقدار اهمیت متغیرهای استفاده در زیرگروه‌های بزرگ خاک.

Figure 6. The importance of variables used in soil subgroups.

داد که استفاده از روش درختان تصمیم‌گیری تصادفی (RF) در نقشه‌برداری خاک می‌تواند برای پیش‌بینی کلاس‌های خاک در اراضی با پستی و بلندی کم و سیلابی هم مورد استفاده قرار گیرد. در این روش می‌توان با بهره‌گیری از متغیرهای کمکی مؤثر در تشکیل خاک، نتایج قابل‌قبولی به‌دست آورد. به‌عبارت دیگر روش RF به‌عنوان یک روش کارا می‌تواند در انتخاب متغیرهای تأثیرگذار در تشکیل خاک عمل کرده و متغیرهای مهم برای استفاده در تهیه نقشه‌های رقومی خاک را نیز معرفی نماید.

نتیجه‌گیری

مطالعات صحرایی نشان از تکامل بسیار کم خاک منطقه مورد مطالعه دارد به‌طوری‌که در دو رده اریدی‌سول و انتی‌سول قرار گرفتند. افق‌های شناسایی اکریک، سالیک و کمبیک تنها افق‌های تکاملی بودند که در منطقه مشاهده شدند. فعالیت رودخانه هیرمند در منطقه و همچنین وجود تالاب‌های هامون در شمال دشت سیستان، باعث شدند که خصوصیات پستی و بلندی متأثر از شرایط آبرفتی مانند شبکه کانال و عمق دره بیش‌ترین تأثیر را در بیان تغییرات خاک در منطقه داشته باشند. نتایج این پژوهش نشان

منابع

1. Al-Masrahy, M.A., and Mountney, N.P. 2015. A classification scheme for fluvial–aeolian system interaction in desert-margin settings. *Aeolian Research*. 17: 67-88.
2. Barthold, F.K., Wiesmeier, M., Breuer, L., Frede, H.G., Wu, J., and Blank, F.B. 2013. Land use and climate control the spatial distribution of soil types in the grasslands of Inner Mongolia. *J. Arid Environ.* 88: 194-205.
3. Behrens, T., Förster, H., Scholten, T., Steinrücken, U., Spies, E.D., and Goldschmitt, M. 2005. Digital soil mapping using artificial neural networks. *J. Plant Nutr. Soil Sci.* 168: 21-33.
4. Behrens, T., Schmidt, K., Zhu, A.X., and Scholten, T. 2010. The ConMap approach for terrainbased digital soil mapping. *Eur. J. Soil Sci.* 61: 133-143.
5. Boer, M., DelBarrio, G., and Puigdefabregas, J. 1996. Mapping soil depth classes in dry Mediterranean areas using terrain attributes derived from a digital elevation model. *Geoderma*. 72: 99-118.
6. Breiman, L., and Cutler, A. 2004. Random Forests. Department of Statistics, University of Berkeley. http://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm.
7. Brungard, C.B., and Boettinger, J.L. 2012. Spatial prediction of biological soil crust classes; value added DSM from soil survey. P 57-60, In: B. Minasny, B.P. Malone and A. McBratney (Eds.), *Digital Soil Assessments and Beyond Proceedings of the 5th Global Workshop on Digital Soil Mapping*. CRC Press, Sydney.
8. Brungard, C.W. 2009. Alternative Sampling and Analysis Methods for Digital Soil Mapping in Southwestern Utah. Thesis for Master of Science, Utah State University. USA.
9. 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.
10. Buol, S.W., Southard, R.J., Graham, R.C., and McDaniel, P.A. 2011. Soil genesis and classification. 6th edition. Iowa State Univ. Press. Ames. Iowa, 556p.
11. Campling, P., Gobin, A., and Feyen, J. 2002. Logistic modeling to spatially predict the probability of soil drainage classes. *Soil Sci. Soc. Am. J.* 66: 1390-1401.
12. Cook, S.E., Jarvis, A., and Gonzalez, J.P. 2008. A New Global Demand for Digital Soil Information. P 31-43, In: A.E. Hartemink, A. McBratney and M.L. Mendonca-Santos (Eds.), *Digital Soil Mapping with Limited Data*. Springer, Dordrecht Heidelberg London New York.
13. Grunwald, S. 2010. Current State of Digital Soil Mapping and What Is Next. P 3-12, In: J.L. Boettinger, D.W. Howel, A.C. Moore, A.E. Hartemink and S. Kienast-Brown (Eds.), *Digital Soil Mapping: Bridging Research, Environmental Application and Operation*. Springer. Dordrecht Heidelberg London New York.
14. Hastie, T., Tibshirani, R., and Friedman, J.H. 2001. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer, New York.
15. Hengl, T., and Reuter, H.I. 2008. *Geomorphometry. Concepts, Software, Applications*. Developments in Soil Science. Elsevier, Amsterdam.
16. Hengl, T., Toomanian, N., Reuter, H.I., and Malakouti, M.J. 2007. Methods to interpolate soil categorical variables from profile observations: lessons from Iran. *Geoderma*. 140: 417-427.
17. Heung, B., Bulmer, C.E., and Schmidt, M.G. 2014. Predictive soil parent material mapping at a regional–Scale: A random forest approach. *Geoderma*. 214-215: 141-154.
18. Jafari, A., Ayoubi, S., Khademi, H., Finke, P.A., and Toomanian, N. 2013. Selection of a taxonomic level for soil mapping using diversity and map purity indices: a case study from an Iranian arid region. *Geomorphology*. 201: 86-97.
19. Jafari, A., Finke, P.A., Van deWauw, J., Ayoubi, S., and Khademi, H. 2012. Spatial prediction of USDA-great soil groups in the arid Zarand region, Iran: comparing logistic regression approaches to predict diagnostic horizons and soil types. *Eur. J. Soil Sci.* 63: 284-298.
20. Jenny, H. 1941. *Factors of Soil Formation: A System of Quantitative Pedology*. McGraw-Hill, New York.

21. Lieb, M., Glaser, B., and Huwe, B. 2012. Uncertainty in the spatial prediction of soil texture comparison of regression tree and random forest models. *Geoderma*. 170: 70-79.
22. Liu, J., Pattey, E., Nolin, M.C., Miller, J.R., and Ka, O. 2008. Mapping within-field soil drainage using remote sensing, DEM and apparent soil electrical conductivity. *Geoderma*. 143: 261-272.
23. McBratney, A.B., Mendonça Santos, M.L., and Minasny, B. 2003. On digital soil mapping. *Geoderma*. 117: 1-2. 3-52.
24. Minasny, B., McBratney, A.B., and Hartemink, A.E. 2010. Global pedodiversity, taxonomic distance and the World Reference Base. *Geoderma*. 155: 132-139.
25. Moonjun, R., Farshad, A., Shrestha, D.P., and Vaiphasa, C. 2010. Artificial neural network and decision tree in predictive soil mapping of Hoi Num Rin sub-watershed, Thailand. P 151-164, In: J.L. Boettinger, D.W. Howell, A.C. Moore, A.E. Hartemink and S. Kienast-Brown (Eds.), *Digital Soil Mapping: Bridging Research, Environmental Application and Operation*. Springer, Dordrecht.
26. National soil survey center. 2012. Field book for describing and sampling soils, Ver. 3. U.S. department of agriculture, Natural resources conservation service.
27. Pahlavan Rad, M.R., Toomanian, N., Khormali, F., Brungard, C.W., Komaki, C.B., and Bogaert, P. 2014. Updating soil survey maps using random forest and conditional latin hypercube sampling in the loess soil of northern Iran. *Geoderma*. 232-234: 97-106.
28. Pahlavan Rad, M.R., Toomanian, N., Khormali, F., Brungard, C.W., Komaki, C.B., and Bogaert, P. 2016. Legacy soil maps as a covariate in digital soil mapping: A case study from northern Iran. *Geoderma*. 279: 141-148.
29. Pahlavan Rad, M.R. 2014. Mapping and Updating Soil Map Using Random Forest and Multinomial Logistic Regression in Golestan Province. Phd Thesis, Gorgan University of Agricultural Sciences and Natural Resources, 114p.
30. Pahlavan Rad, M.R., Toomanian, N., Khormali, F., Brungard, C.W., Komaki, C.B., and Bogaert, P. 2014. Digital soil mapping using random decision tree models in Golestan province. *J. Water Soil Cons.* 21: 6. 73-93. (In Persian)
31. Poggio, L., Gimona, A., and Brewer, M.J. 2013. Regional scale mapping of soil properties and their uncertainty with a large number of satellite-derived covariates. *Geoderma*. 209-210: 1-14.
32. Roecker, S.M., Howell, D.W., Haydu-Houdeshell, C.A., and Blinn, C. 2010. A Qualitative Comparison of Conventional Soil Survey and Digital Soil Mapping Approaches. P 369-384, In: J.L. Boettinger, D.W. Howell, A.C. Moore, E.A. Hartemink and S. Kienast-Brown (Eds.), *Digital Soil Mapping: Bridging Research, Environmental Application and Operation*. Progress in Soil Science. Springer, New York.
33. Schatzl, R.J., and Anderson, S. 2005. *Soils: Genesis and Geomorphology*. Cambridge University Press, 833p.
34. Soil Survey Staff. 2014. *Keys to soil Taxonomy*, 12th ed. U.S. department of agriculture, Natural resources conservation service.
35. Stum, A.K., Boettinger, J.L., White, M.A., and Ramsey, R.D. 2010. Random Forests applied as a soil spatial predictive model in arid Utah. P 179-189, In: J.L. Boettinger, D. Howell, A.C. Moore, A. Hartemink and E.S. Kienast-Brown (Eds.), *Digital Soil Mapping: Bridging Research, Environmental Application and Operation*. Progress in Soil Science. Springer, Logan, USA.
36. Taghizadeh-Mehrjardi, R., Minasny, B., Sarmadian, F., and Malone, B.P. 2014. Digital mapping of soil salinity in Ardakan region, central Iran. *Geoderma*. 213: 15-28.
37. Were, K., Bui, D.T., Disk, B., and Singl, B.R. 2015. A comparative assessment of support vector regression, artificial neural networks and random forest for predicting soil organic carbon stocks across an afro-monkane land scape. *Ecological indicator*. Pp: 394-403.

38. Wilson, J.P., and Gallant, J.C. 2000. Terrain Analysis: Principles and Applications. In: G.J. Wilson JP (Ed.), Digital terrain analysis. John Wiley, New York, 478p.
39. Xiong, X., Grunwald, S., Myers, D.B., Kim, J., Harris, W.G., and Comerford, N.B. 2012. Which soil, environmental and anthropogenic covariates for soil carbon models in Florida are needed? P 335-339, In: B. Minasny, B.P. Malone and A. McBratney (Eds.), Digital Soil Assessments and Beyond: Proceedings of the 5th Global Workshop on Digital Soil Mapping. CRC Press, Sydney.
40. Yokoyama, R., Shirasawa, M., and Pike, R.J. 2002. Visualizing topography by openness: a new application of image processing to digital elevation models. Photogramm. Eng. Remote Sens. 68: 257-266.



Gorgan University of Agricultural
Sciences and Natural Resources

J. of Water and Soil Conservation, Vol. 24(1), 2017
<http://jwsc.gau.ac.ir>

Application of random forest method for predicting soil classes in low relief lands (Case study: Hirmand county)

Kh. Mirakzahi¹, *A. Shahriari¹, M.R. Pahlavan Rad² and A. Bameri¹

¹Department of Soil Science and Engineering, Faculty of Water and Soil, University of Zabol, Zabol, Iran, ²Soil and Water Department, Agriculture and Resources Research and Education Center of Sistan, AREEO, Zabol, Iran

Received: 12/21/2016; Accepted: 05/16/2017

Abstract

Background and Objectives: Soil survey and mapping are base of soil information for environmental modeling as a way to determine soil distribution patterns, describe and display it to understood and interpreted for different users. Digital soil mapping creates link between classes or soil characteristics and environmental factors affected soil formation and development by using mathematical models which can provide more precise and accurate soil maps and reducing costs of soil survey and mapping projects. This study was done to mapping soil great groups and subgroups by using Random Forest technique in the Hirmand county lands in Sistan plain.

Materials and Methods: In this study 108 soil profiles were dug on about 60.000 hectares of Hirmand county lands. Sixteen environmental variables were used as estimators for soil mapping including land properties, salinity and vegetation index. After classification of soil profiles to great groups and subgroups, soil classes map produced by using random forest (RF) method. It should be mentioned 80 percent of data was used for model training and 20 percent for independent validation.

Results: Pedological studies showed soils of Sistan plain haven't high development and most of them are Entisols and Aridisols. Most soil profiles classified as Torrifluvents on great groups level and Typic Torrifluvents as a subgroup. Also the result of RF showed the lowest out of bag error (OOB) samples in soil great groups and subgroups were 43.53 and 50.59 respectively. Independent validation results showed the best accuracy obtained for soil great groups and subgroups were 48 and 53 percent respectively. Valley depth, convergence index, channel network between and salinity in soil great groups and valley depth, elevation and catchment area in soil subgroups were the most important environmental variables to estimate soil classes.

Conclusion: The results showed most soils are young in the low relief lands in arid regions and these regions have also low soil diversity. Soil digital mapping and random forest technique could be useful for soil classes prediction and soil mapping in this kind of regions.

Keywords: Soil digital mapping, Random forest technique, Map accuracy, Arid regions, Sistan plain

* Corresponding Author; Email: shahriari.ali@uoz.ac.ir