



دانشگاه گوارزی و منابع طبیعی گران

نشریه پژوهش‌های حفاظت آب و خاک
جلد بیست و یکم، شماره ششم، ۱۳۹۳
<http://jwsc.gau.ac.ir>

پهنه‌بندی رقومی واحدهای خاک با استفاده از مدل درختان تصمیم‌گیری تصادفی در استان گلستان

*محمدرضا پهلوان‌راد^۱، فرهاد خرمالی^۲، نورایر تومانیان^۳، فرشاد کیانی^۴ و چوقی بایرام کمکی^۵

^۱دانش‌آموخته دکتری گروه علوم خاک، دانشگاه علوم کشاورزی و منابع طبیعی گرگان، آستاد گروه علوم خاک،

دانشگاه علوم کشاورزی و منابع طبیعی گرگان، آستادیار بخش خاک و آب، مرکز تحقیقات کشاورزی و

منابع طبیعی اصفهان، ^۲آستادیار گروه علوم خاک، دانشگاه علوم کشاورزی و منابع طبیعی گرگان،

^۳آستادیار گروه مدیریت مناطق بیابانی، دانشگاه علوم کشاورزی و منابع طبیعی گرگان

تاریخ دریافت: ۹۲/۸/۱۲؛ تاریخ پذیرش: ۹۳/۱/۹

چکیده

نقشه‌های خاک از منابع عمده اطلاعات برای مدیریت اراضی، منابع طبیعی و محیط زیست هستند. استفاده از روش‌های نقشه‌برداری رقومی باعث کاهش هزینه و صرفه‌جویی در زمان در مطالعات شناسایی خاک می‌گردد. در این مطالعه روش نمونه‌برداری مکعب لاتین مشروط (cLHS) برای تعیین محل نقاط نمونه‌برداری و تکنیک درختان تصمیم‌گیری تصادفی (RF) برای تخمین نقشه زیرگروه‌های خاک در سطحی حدود ۸۵۰۰۰ هکتار از اراضی استان گلستان ارزیابی شدند. ۱۲ متغیر محیطی شامل خصوصیات زمین، واحدهای ژئومورفولوژی و شاخص پوشش گیاهی به‌عنوان تخمین‌گر مورد استفاده قرار گرفته شدند. نتایج نشان داد که تعداد کلاس‌های خاک موجود در سطح زیرگروه‌های بزرگ خاک در محدوده مورد مطالعه ۱۱ کلاس بودند. کم‌ترین مقدار خطای تخمین نمونه‌های اعتبارسنجی در مدل‌سازی ۵۲/۵۳ به‌دست آمد. نتایج نشان داد که کلاس‌هایی که نمونه بیش‌تری داشتند خطای کم‌تری داشتند. کلاس‌های جیپسیک آکویب سالدز با ۲۲، آکوئیک کلسی‌زپتیز با ۹ و تیپیک هاپلوزرپتیز با ۲۲ نقطه مشاهداتی به‌ترتیب دارای خطای ۲۲، ۲۵ و ۳۷ درصد بودند. کلاس‌های جیپسیک هاپلوزرپتیز با ۶، تیپیک آکویب سالدز با ۵، تیپیک کلسی‌زپتیز و تیپیک کلسی‌زولز با ۴ و تیپیک هالوآکوئیتیز با

* مسئول مکاتبه: pahlevan354@yahoo.com

۳ نقطه مشاهداتی دارای خطای ۱۰۰ درصد بودند. بین متغیرهای مختلف شاخص پوشش گیاهی، ژئومورفولوژی، ارتفاع و جهت شیب به ترتیب دارای بیش‌ترین اهمیت در تخمین کلاس بودند. نتایج نشان داد که روش RF می‌تواند یک روش قابل اعتماد و با دقت مناسب باشد که با تعداد نمونه کم تخمین قابل قبول ارائه می‌دهد.

واژه‌های کلیدی: نقشه‌برداری رقومی، درختان تصمیم‌گیری تصادفی، دقت نقشه، استان گلستان

مقدمه

وجود نقشه‌های خاک از نیازهای اساسی در مطالعات علوم زمین می‌باشد. با وجودی که بیش از ۵۰ سال از شروع مطالعات خاکشناسی در ایران می‌گذرد، نقشه‌های خاک کشور به‌طور کامل تهیه نشده‌اند. استفاده از روش‌های مرسوم نقشه‌برداری خاک برای تهیه نقشه‌های خاک کل مناطق کشور با مقیاس مناسب، به دلیل هزینه‌بر و وقت‌گیر بودن بسیار مشکل است. بنابراین، یافتن روش‌های دیگر که زمان و هزینه کم‌تری نیاز داشته و از دقت قابل‌قبولی نیز برخوردار باشند ضروری به‌نظر می‌رسد. نقشه‌برداری رقومی خاک (Digital Soil Mapping) می‌تواند جایگزین روش‌های معمولی نقشه‌برداری گردد (کوک و همکاران، ۲۰۰۸). در نقشه‌برداری مرسوم خاک دانش پدولوژی به‌طور دقیقی برای تولید نقشه‌های خاک استفاده می‌شود؛ اما هنگامی که با روش‌های در حال توسعه نقشه‌برداری رقومی خاک مقایسه می‌شوند زمان‌بر و هزینه‌بر نشان می‌دهند (گرونوالد، ۲۰۱۰).

نقشه‌برداری رقومی خاک (DSM)^۱ شامل روش‌ها و مدل‌هایی است که بین توزیع خاک (کلاس‌ها یا خصوصیات خاک) و داده‌هایی که به آسانی و با قیمت ارزان از طریق روش‌های سنجش از دور، تصاویر و عکس‌های ماهواره‌ای و داده‌های ژئومورفومتری به‌دست می‌آیند و تحت عنوان متغیرهای کمکی محیطی^۲ نامیده می‌شوند ارتباط برقرار می‌کند. اگر چه تکنیک‌های DSM بر روی محاسبات دیجیتالی استوار است ولی اساس آن بر روی معادلات تشکیل خاک (جنی، ۱۹۴۱؛ مک‌براتنی و همکاران، ۲۰۰۳) قرار دارد. در همه این مدل‌ها نباید اثر دانش نقشه‌بردار روی کیفیت نتایج نادیده گرفته شود. جنی (۱۹۴۱) مدل معروف خود را که براساس آن خاک بر روی یک زمین‌نما تابعی از

1- Digital Soil Mapping

2- Environmental Covariates

۵ فاکتور محیطی است معرفی کرد. این مدل به صورت معادله $S=f(CL, O, R, P, T)$ تعریف شده است، که S: خاک، CL: اقلیم، O: ارگانیزم که بیش تر به صورت پوشش گیاهی بیان می شود، R: توپوگرافی یا پستی و بلندی، P: مواد مادری و T: زمان می باشد. مک براتنی و همکاران (۲۰۰۳) مدل پدولوژیکی CIORPT را بازبینی کردند و پارامترهای خاک (کلاس ها و یا خصوصیات خاک) و موقعیت مکانی را به آن اضافه نموده و مدل جدیدی تحت عنوان مدل SCORPAN را برای کمی کردن ارتباط بین داده های دقیق مکانی و خاک ارایه کردند. اجزای این مدل تحت عنوان متغیرهای کمکی محیطی معرفی شده اند. مدل جدید به صورت $S=f(s, c, o, r, p, a, n)$ است که فاکتورهای S داده های صحرایی و یا آزمایشگاهی خاک در نقاط نمونه برداری یا در نقشه های خاک هستند و n موقعیت مکانی نقاط نمونه برداری می باشد که به پارامترهای مدل جنی (۱۹۴۱) اضافه شدند.

روش های فراوان و مختلفی در DSM استفاده می شوند. این ها شامل توابع پدوترانسفر^۱، زمین آمار، مدل های آماری تجربی، طبقه بندی فازی، تصمیم گیری درختی^۲ و شبکه های عصبی مصنوعی هستند (مک براتنی و همکاران، ۲۰۰۳). جعفری و همکاران (۲۰۱۲) روش رگرسیون دوتایی^۳ و رگرسیون چندمتغیره^۴ را با ۱۲۶ پروفیل در سطح ۹۰۰۰۰ هکتار، برای پیش بینی توزیع گروه های بزرگ در منطقه زرنند کرمان مورد استفاده قرار دادند. تقی زاده مهرجردی و همکاران (۲۰۱۴) روش تصمیم گیری درختی را برای پیش بینی مکانی زیرگروه های بزرگ خاک در اردکان یزد به کار بردند.

مدل درختان تصمیم گیری تصادفی (RF)^۵ مدل توسعه یافته از مدل طبقه بندی و رگرسیون درختی (CART)^۶ می باشد. روش CART (برایمن و همکاران، ۱۹۸۴) روشی است که داده ها را به طور تکراری برای به دست آوردن ارتباط بین متغیر پاسخ و متغیرهای مستقل و انجام تخمین جداسازی می کند. در روش RF به جای رشد یک درخت، صدها یا هزاران درخت طبقه بندی تولید می شود (برایمن و کاتلر، ۲۰۰۴). در این روش یک سوم داده های اصلی در ایجاد هر درخت استفاده نمی شوند که تحت عنوان نمونه های بیرون از سبد (OOB)^۷ نامیده می شوند. چون این نمونه ها در آموزش

- 1- Pedotransfer
- 2- Decision Tree
- 3- Binary Logistic Regression
- 4- Multinomial Logistic Regression
- 5- Random Forest
- 6- Classification and Regression Trees
- 7- Out of Bag

درخت استفاده نشده‌اند در تست 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. 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 بیش‌تر مطالعات استفاده شده، مقادیر درست متغیرها با مقادیر 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 (بريمن و کاتلر، ۲۰۰۴).

کاربردهای RF در علوم خاک هنوز کم بوده و مطالعات محدودی از آن در علوم خاک انجام شده است. 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. گریم و همکاران (۲۰۰۸) RF را برای پیش‌بینی و منشأ کربن آلی خاک در محیط تروپیکال استفاده کردند. آن‌ها RF را یک ابزار نیرومند برای تخمین مکانی و درک ارتباطات خاک-زمین یافتند. لایب و همکاران (۲۰۱۲) RF و RT^۳ را برای پیش‌بینی توزیع بافت خاک استفاده کرده و مشاهده کردند که روش RF دارای دقت بالاتری نسبت به RT بود. 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 می‌باشد. بارتولد و همکاران (۲۰۱۳) با استفاده از RF توزیع تیپ‌های مختلف خاک با روش رده‌بندی WRB^۴ را در سطحی حدود ۳۶۰۰ کیلومترمربع با ۱۴۵ پروفیل 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 ۵۱/۶ درصد بود.

انتخاب روش نمونه‌برداری در مطالعات خاکشناسی مهم می‌باشد. یک روش نمونه‌برداری تقریباً مدرن، روش نمونه‌برداری مکعب لاتین شرطی (cLHS)^۵ است (میناسنی و مک‌برانتسی، ۲۰۰۶)، که

- 1- Mean Decrease in Accuracy
- 2- Mean Decrease in Gini Coefficient
- 3- Regression Tree
- 4- World Reference Base for Soil Resources
- 5- Conditioned Latin Hypercube Sampling

روشی طبقه‌بندی شده^۱ یا لایه‌ای است. در این روش متغیرهای کمکی به دسته‌های یکنواخت تقسیم‌بندی می‌شوند. روش کار cLHS به این صورت است که کاربر تصمیم می‌گیرد تعداد نمونه مشخصی را استفاده کند، این تکنیک، داده‌های متغیرهای محیطی را به همان تعداد نمونه‌های خواسته به صورت خوشه تقسیم می‌کند و یک نمونه تصادفی از داده‌های ورودی متغیرهای محیطی هر خوشه را طوری انتخاب می‌کند که شرط مربع لاتین را تامین کند و این شرط این است که در هر بعد a و b فقط یک نمونه وجود دارد. در روش‌های جاری نمونه‌برداری برای نقشه خاک، انتخاب محل نمونه‌برداری به تصمیمات شخصی نقشه‌برداران بستگی دارد. میناسنی و مک‌برانتی (۲۰۰۷) روش cLHS را با روش نمونه‌برداری تصادفی ساده^۲، نمونه‌برداری تصادفی طبقه‌بندی شده و نمونه‌برداری هایپرکیوب لاتین مکانی^۳ مقایسه کرده و مشاهده کردند که روش cLHS بیش‌تر به‌طور دقیقی توزیع واقعی متغیرهای محیطی را نشان می‌دهد. تقی‌زاده‌مهرجردی و همکاران (۲۰۱۴) روش cLHS را برای تعیین محل نمونه‌برداری ۱۷۳ نقطه مطالعاتی در ارکان یزد استفاده کردند. مودلر و همکاران (۲۰۱۳) نشان دادند که cLHS روش مناسبی برای نشان دادن تغییرات خاک می‌باشد.

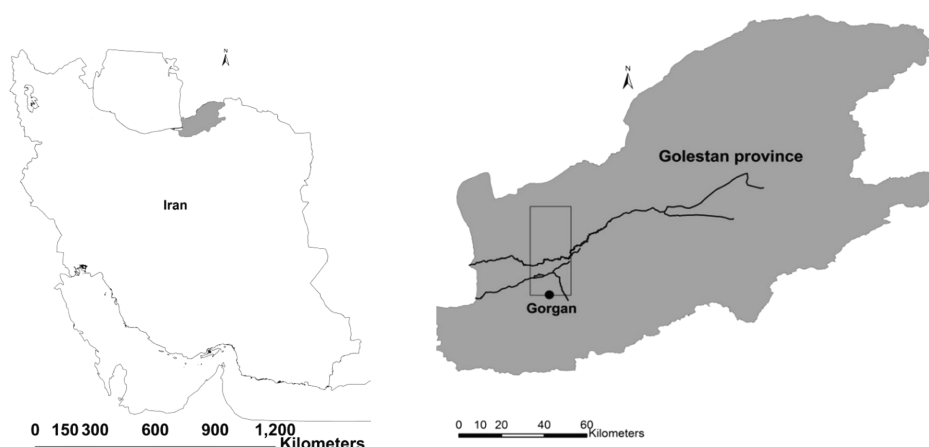
این پژوهش با هدف ارزیابی دقت تکنیک RF در تخمین نقشه کلاس‌های خاک در سطح زیرگروه‌های خاک در استان گلستان که از قطب‌های کشاورزی کشور می‌باشد انجام گرفت.

مواد و روش‌ها

محدوده مورد مطالعه دارای مساحت حدود ۸۵۰۰۰ هکتار و در استان گلستان واقع شده است (شکل ۱). این محدوده با عرض ۱۹ کیلومتر از شمال شهر گرگان شروع و تا ۴۵ کیلومتر به طرف شمال ادامه دارد. ارتفاع محدوده بین ۱۵۸ متر در جنوب منطقه مورد مطالعه تا ۱۸ متر زیر سطح دریا می‌باشد. مواد مادری منطقه به‌طور عمده رسوبات لس و لس‌های منتقل شده با آب هستند. واحدهای ژئومورفولوژی در محدوده مورد مطالعه شامل دشت رسوبی دامنه‌ای، دشت رسوبی رودخانه‌ای، اراضی پست و تپه‌های لسی هستند. رژیم رطوبتی منطقه از زیریک در قسمت‌های جنوبی و مرکزی تا اریدیک در نزدیک مرز شمالی متفاوت است. رژیم حرارتی منطقه مورد مطالعه ترمیک است. تقریباً

- 1- Stratified random sampling
- 2- Simple random sampling
- 3- Spatial Latin Hypercube sampling

۸۵ درصد منطقه مورد مطالعه اراضی قابل کشت و بقیه را اراضی مرتعی تشکیل می‌دهد. بیش‌تر اراضی کشت شده مسطح بوده و تحت کشت گندم، جو، برنج، ذرت، سویا، سیب‌زمینی و پنبه هستند. رده‌های عمده خاک مورد انتظار در این محدوده مالی‌سولز، آلفی‌سولز، اینسپتی‌سولز و انتی‌سولز است.



شکل ۱- موقعیت منطقه مطالعاتی.

۱۲ خصوصیت متغیر محیطی که شامل خصوصیات زمین (ویلسون و گالانت، ۲۰۰۰) شاخص پوشش گیاهی و واحدهای ژئومورفولوژی بودند در این مطالعه استفاده شد. خصوصیات زمین شامل شیب، ارتفاع، جهت، انحنای سطح^۱، انحنای نیم‌رخ^۲، شاخص حمل رسوب، شیب میانگین شیب بالایی^۳، طول جریان شیب پایینی^۴، بیش‌ترین شیب به سمت پایین^۵ و شاخص خیزی^۶ از نقشه DEM (شکل ۲) با بزرگ‌نمایی ۳۰ متر با استفاده از نرم‌افزارهای ArcGIS 9.3 و TAS^۷ استخراج گردیدند. جدول ۱ تعریف این خصوصیات و شکل ۳ نمای برخی از آن‌ها را نشان می‌دهد. از SAVI به‌عنوان

- 1- Plan Curvature
- 2- Profile Curvature
- 3- Mean Upslope Slope
- 4- Downslope Flowpath Length
- 5- Max Downward Slope
- 6- Wetness Index
- 7- Terrain Analysis System

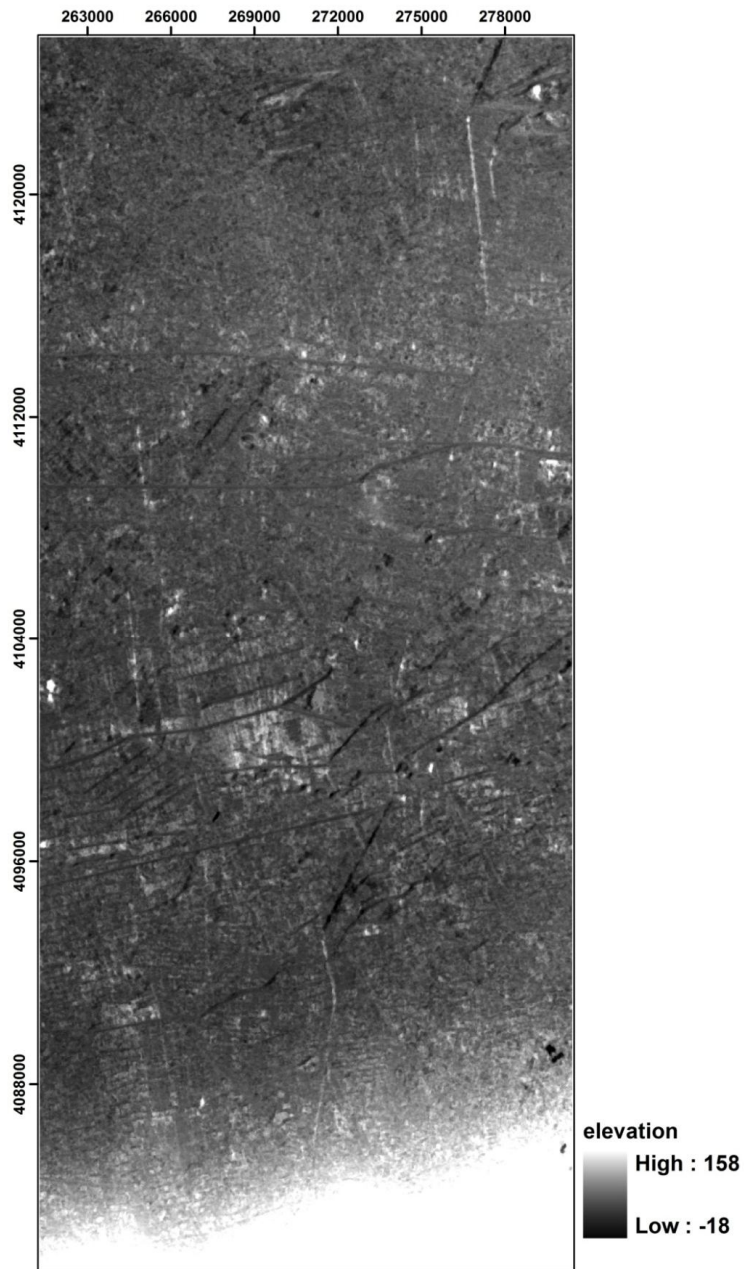
شاخص پوشش گیاهی استفاده شد چون به زمینه خاک حساسیت ندارد (هیوت، ۱۹۸۸). رابطه ۱ روش محاسبه SAVI را نشان می‌دهد.

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

که در آن، NIR و Red: مقادیر انعکاس امواج مادون قرمز نزدیک و قرمز هستند و L: یک فاکتور تصحیح می‌باشد و به نظر می‌رسد استفاده از مقدار ۰/۵ برای زمینه خاک‌های مختلف کافی باشد (هیوت، ۱۹۸۸). SAVI با استفاده از تصاویر ماهواره‌ای لندست TM ۵ که مربوط به هشتم فروردین سال ۱۳۹۰ می‌باشد (شکل ۴) و با کمک نرم‌افزار ERDAS IMAGINE 9.2 به دست آمد. واحدهای ژئومورفولوژی با استفاده از روش تفسیر عکس‌های هوایی (تومانیان و همکاران، ۲۰۰۶) تعیین شدند (شکل ۵ و جدول ۲).

روش cLHS که توسط میناسنی و مک‌براتی (۲۰۰۶) معرفی شده است با استفاده از نرم‌افزار Matlab (مت‌ورکس، ۲۰۰۹) برای انتخاب محل‌های نمونه‌برداری پروفیل‌های خاک اجرا گردید. مقادیر همه متغیرهای کمی در پیکسل‌های ۳۰ در ۳۰ متر در نرم‌افزار ArcGIS به دست آمد و این مقادیر به نرم‌افزار Matlab منتقل و با تکنیک cLHS محل ۱۰۵ پروفیل مشخص گردید (شکل ۴). در عملیات میدانی ۹۹ پروفیل قابل نمونه‌برداری بود که تا عمق ۱۵۰-۱۰۰ سانتی‌متر حفر گردید. پروفیل‌های خاک با روش‌های استاندارد تشریح شدند (اسکائنیگر و همکاران، ۲۰۰۲) و نمونه‌های خاک از هر افق تهیه شد و پس از هوا خشک شدن و عبور از الک ۲ میلی‌متر آزمایش‌های فیزیکوشیمیایی لازم انجام و سپس پروفیل‌ها براساس دستورالعمل نقشه‌برداری خاک (اعضای نقشه‌برداری خاک، ۱۹۹۹) و کلید طبقه‌بندی خاک (اعضای نقشه‌برداری خاک، ۲۰۱۰) تا سطح زیرگروه بزرگ خاک طبقه‌بندی شدند.

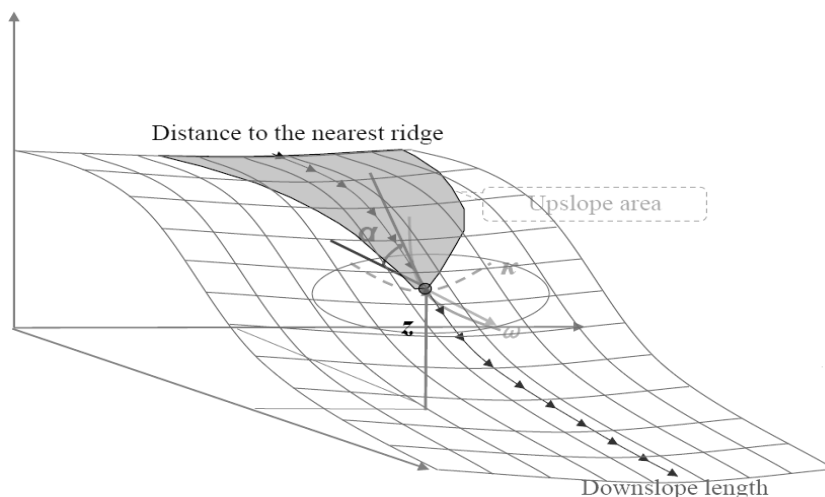
لایه‌های ۱۲ متغیر با پیکسل‌های ۳۰ مترمربعی و کلاس‌های خاک وارد نرم‌افزار (۲۰۱۳) R 3.0.1 و R-Studio0.97.551 گردید و مدل RF اجرا گردید. در مدل RF تعداد درختان در جنگل و تعداد متغیرهای محیطی در گره هر درخت توسط کاربر مشخص می‌شود و مناسب‌ترین مقدار این دو پارامتر با روش سعی و خطا برای به دست آوردن کم‌ترین مقدار خطا به دست آمد.



شکل ۲- نقشه ارتفاعی رقومی (DEM) منطقه مطالعاتی.

جدول ۱- توصیف خصوصیات زمین استفاده شده در پژوهش (ویلسون و گالات، ۲۰۰۰).

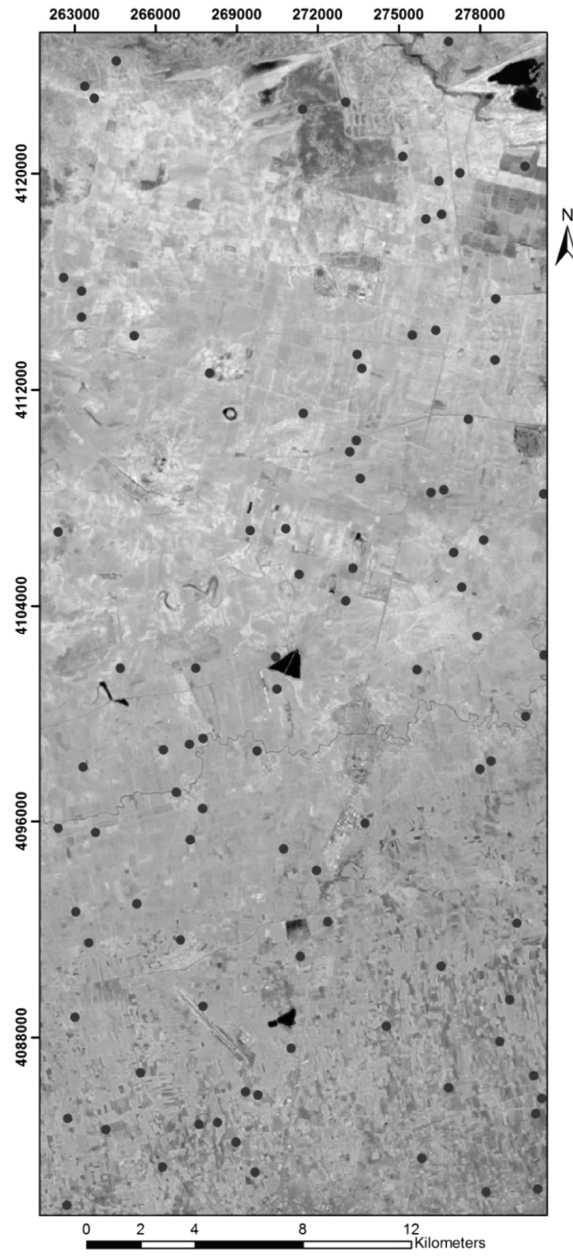
تعریف	واحد	خصوصیت
ارتفاع نسبت به سطح دریا	متر	ارتفاع (Elevation)
زاویه بین سطح تماس و سطح افقی در یک نقطه در سطح زمین	درصد	شیب (Slope)
بیشترین سرعت تغییر در ارتفاع در هر سلول از DEM انحنای در صفحه عمودی خط جریان است، بیانگر اندازه افزایش و کاهش سرعت جریان است، بنابراین در جریان آب و انتقال رسوب می‌باشد.	RAD	جهت شیب (Aspect)
انحنای در صفحه افقی یک خط کنتور. نشان‌دهنده انحراف جریان مواد و بیانگر اندازه همگرایی و واگرایی جریان می‌باشد.	یک بر متر	انحنای قائم (Profile Curvature)
انحنای در صفحه افقی یک خط کنتور. نشان‌دهنده انحراف جریان مواد و بیانگر اندازه همگرایی و واگرایی جریان می‌باشد.	یک بر متر	انحنای افقی (Plan Curvature)
پارامتری که توزیع مکانی منطقه اشباع و مقدار آب خاک در زمین‌نما را بیان می‌کند.	-	شاخص خیزی (TWI)
نمایانگر فرآیندهای فرسایش و رسوب در زمین‌نما بوده و اثر توپوگرافی را بر فرسایش نشان می‌دهد.	-	شاخص حمل رسوب (STI)
متوسط شیب ناحیه بالایی شیب	درصد	شیب میانگین شیب بالایی (Mean upslope slope)
بیشترین شیب به طرف پایین‌ترین همسایه در یک پنجره ۳×۳ پیکسلی	درصد	بیشترین شیب به سمت پایین (Max downward slope)
فاصله پایین شیب با مسیر جریان آب به بیرون	متر	طول جریان به طرف پایین شیب (Downslope flowpath length)



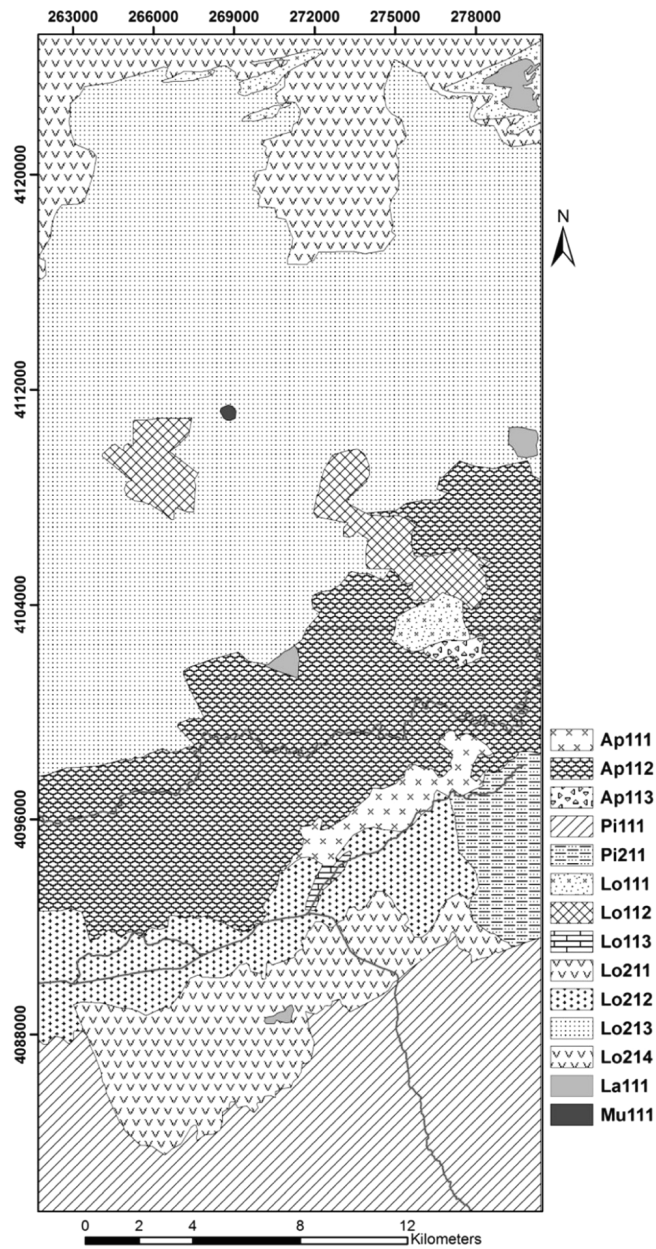
شکل ۳- ارتفاع (Z)، شیب (α)، جهت (ω)، انحناءها (κ)، فاصله به نزدیک‌ترین لبه (Distance to the nearest ridge)، ناحیه بالای شیب (Upslope area) و طول پایین شیب (Downslope length) (ویلسون، ۲۰۰۹).

جدول ۲- راهنمای واحدهای ژئومورفولوژی تفکیک شده.

زمین‌نما	لندفرم	لیتولوژی	سطح ژئومورفیک	علامت
آبرفت دامنه‌ای	فن بادبزنی شکل	مواد لسی منتقل شده با آب	با شیب کم و کشت شده	Pi111
	آبرفت دامنه‌ای مسطح	مواد لسی منتقل شده با آب	به نسبت مسطح و کشت شده	Pi211
دشت رودخانه‌ای	دشت رودخانه‌ای با مواد منتقل شده لسی	رسوبات رودخانه	مسطح، تحت تأثیر سیلاب و کشت شده	Ap111
			مسطح با شوری کم و توسط کشت شده	Ap112
			مسطح با شوری زیاد و کشت نشده	Ap113
مواد لسی	مواد لسی دارای پستی بلندی	رسوبات ریز	کشت نشده با شیب متوسط	Lo111
			کشت شده با شیب متوسط	Lo112
			پوشش مرتعی با شیب کم	Lo113
			مسطح کشت شده غیر شور	Lo211
			مسطح کشت شده با شوری کم با آب زیرزمینی بالا	Lo212
	اراضی پست	رسوبات ریز	مسطح شور کشت شده	Lo213
			مسطح خیلی شور با زهکشی ضعیف با پوشش مرتعی	Lo214
گلفشان	رسوبات گلفشان	سیلتی، رسی		Mul11
آبگیر	رسوبات آبگیرها	سیلتی رسی		Lal11



شکل ۴- تصویر ماهواره‌ای و محل نقاط نمونه برداری.



شکل ۵- نقشه واحدهای ژئومورفولوژی منطقه مورد مطالعه.

نتایج و بحث

اجرای مدل RF: کمترین مقدار خطای تخمین نمونه‌های OOB (نمونه‌های خارج از سبد) که توسط تکنیک در مدل‌سازی استفاده نشده‌اند و جهت اعتبارسنجی مستقل به کار می‌روند، ۵۲/۵۳ درصد بود که با تعداد ۳۰۰ درخت و ۷ متغیر در هر درخت به دست آمد. گریم و همکاران (۲۰۰۸) مقدار خطای تخمین OOB ۹۵-۷۴ درصد را در تخمین مقدار ذخایر کربن آلی خاک با روش RF مشاهده کردند. بارتولد و همکاران (۲۰۱۳) به مقدار خطای تخمین OOB ۵۱ درصد در پیش‌بینی کلاس‌های خاک در سطح اول روش WRB دست یافتند. استام و همکاران (۲۰۱۰) با ۶۷۲ مشاهده پروفیلی 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. راکر و همکاران (۲۰۱۰) در پیش‌بینی کلاس خاک با روش RF مقدار خطای ۵۱ درصد را گزارش کردند. برونگارد (۲۰۰۹) با ۳۰۰ نمونه پروفیل خاک در سطح ۳۰۰۰۰ هکتار در سطح زیرگروه بزرگ خاک مقدار خطای تخمین ۶۲ درصد را به دست آوردند. مقایسه نتایج به دست آمده با نتایج سایر پژوهشگران که در بالا ذکر گردید نشان می‌دهد که خطای به دست آمده کم‌تر از بیش‌تر آن‌ها بوده است و این در شرایطی به دست آمده است که از نظر تعداد مشاهدات پروفیلی کم‌تر از سایر پژوهشگران بوده است و سطح رده‌بندی خاک هم در سطوح پایین رده‌بندی بوده است.

جدول ۳ کلاس‌های خاک، تعداد پروفیل‌ها و درصد سطح مربوط به هر کلاس، در منطقه مورد مطالعه را نشان می‌دهد. براساس نتایج این جدول، بیش‌ترین تعداد پروفیل‌ها مربوط به کلاس‌های *Gypsic Aquisalids* و *Typic Haploxerepts* بود که بیش از نصف نقشه تولیدی را تشکیل می‌دهند.

نتایج خطای تخمین کلاس‌های پیش‌بینی شده در جدول ۴ نشان داد که عموماً کلاس‌هایی که تعداد نقاط مشاهداتی بیش‌تری داشتند (جدول ۳) دارای خطای پیش‌بینی کم‌تری هستند. براساس نتایج جدول ۳، زیرگروه بزرگ *Gypsic Aquisalids* که ۲۲ درصد پروفیل‌های مشاهداتی را به خود اختصاص داد با مقدار خطای ۲۲ درصد دارای کم‌ترین خطای پیش‌بینی بود (جدول ۴) و پس از آن زیرگروه *Aquic Calcixerepts* دارای خطای تخمین ۲۵ درصد بود (جدول ۴) که ۱۲ مشاهده

پروفیلی (۱۲ درصد نقاط مطالعاتی) داشت. زیرگروه‌هایی مانند Typic Calcixerolls و Typic Haloaqepts که دارای نقاط مشاهداتی کم‌تری بودند (جدول ۳) دارای خطای بالایی بودند (جدول ۴). برونگارد (۲۰۰۹) مشاهده کرد که کلاس‌هایی از خاک که دارای نقاط پروفیلی بیش‌تری داشت دارای خطای کم‌تری بود. جعفری و همکاران (۲۰۱۲) نیز به چنین نتیجه‌ای دست یافتند که کلاس‌های خاکی که نمونه کم‌تری داشتند دارای خطای پیش‌بینی بیش‌تری بودند. نتایج به‌دست آمده نشان می‌دهد که روش RF می‌تواند یک روش مناسب برای تخمین کلاس‌های خاک باشد. بارتولد و همکاران (۲۰۱۳)، برونگارد (۲۰۰۹) و استام و همکاران (۲۰۱۰) RF را یک روش قدرتمند در تخمین کلاس‌های خاک معرفی کردند. برای افزایش دقت کلاس‌هایی که دارای دقت کمی هستند می‌توان با افزایش تعداد نقاط نمونه‌برداری، دقت کلاس‌ها و در نتیجه دقت نقشه تولیدی را افزایش داد. استفاده از سایر متغیرهای محیطی که در این مطالعه استفاده نگردید ممکن است سبب افزایش دقت نقشه تولیدی گردد.

جدول ۳- نتایج رده‌بندی زیرگروه‌های غالب منطقه مورد مطالعه به همراه فراوانی نسبی آن‌ها.

نام زیرگروه بزرگ	تعداد پروفیل	سطح (هکتار)	درصد سطح
Aquic Calcixerepts	۱۲	۱۱۱۳۴	۱۳/۱
Aquic Calcixerolls	۵	۳۳۱۳	۳/۹
Fluventic Haploxerepts	۷	۹۸۳	۱/۱۶
Gypsic Aquisalids	۲۲	۳۰۵۰۰	۳۶
Gypsic Haploxerepts	۶	۲۸۳۸	۳/۳۵
Typic Aquisalids	۵	۲۰۳	۰/۲۴
Typic Calcixerepts	۴	۷۹	۰/۰۹
Typic Calcixerolls	۴	۱۱۲	۰/۱۳
Typic Haloaqepts	۳	۶۴۲	۰/۶۷
Typic Haploxerepts	۲۲	۲۷۷۰۰	۳۲/۷
Typic Xerofluvents	۸	۷۴۰۲	۸/۷۴

جدول ۴- آرایه عدم قطعیت تخمین OOB و خطای کلاس‌های پیش‌بینی شده.

خطای کلاس	مشاهده شده											
	۱۱	۱۰	۹	۸	۷	۶	۵	۴	۳	۲	۱	زیرگروه بزرگ شاکی
۲۵	۰	۲	۰	۰	۰	۰	۰	۱	۰	۰	۹	(۱) Aquic Calcixereps
۶۰	۲	۰	۰	۰	۰	۰	۰	۱	۲	۰	۰	(۲) Aquic Calcixerolls
۷۱	۱	۰	۰	۰	۰	۱	۲	۲	۱	۰	۰	(۳) Fluventic Haploxerepts
۲۲	۰	۴	۰	۰	۰	۱	۱۷	۰	۰	۰	۰	(۴) Gypsic Aquisalids
۱۰۰	۱	۲	۰	۰	۰	۰	۱	۱	۱	۰	۰	(۵) Gypsic Haploxerepts
۱۰۰	۱	۳	۰	۰	۰	۰	۱	۰	۰	۰	۰	(۶) Typic Aquisalids
۱۰۰	۰	۴	۰	۰	۰	۰	۰	۰	۰	۰	۰	(۷) Typic Calcixereps
۱۰۰	۰	۲	۰	۰	۰	۰	۰	۰	۱	۱	۰	(۸) Typic Calcixerolls
۱۰۰	۱	۱	۰	۰	۰	۰	۰	۱	۰	۰	۰	(۹) Typic Halaquepts
۴۵	۱	۱۲	۰	۱	۰	۰	۳	۱	۱	۳	۰	(۱۰) Typic Haploxerepts
۳۷	۵	۲	۰	۰	۰	۰	۰	۰	۱	۰	۰	(۱۱) Typic Haploxerepts

تخمین زده شده

اهمیت متغیرها در تخمین کلاس‌های خاک با RF: نتایج اهمیت متغیرها در نقشه تولید شده در شکل ۶ ارایه شده است که براساس آن، شاخص پوشش گیاهی (SAVI) دارای بیش‌ترین اهمیت در تخمین نقشه تولیدی بود و بیش‌ترین اثر را روی نقشه تولیدی داشت. به‌نظر می‌رسد نقشه پوشش گیاهی علاوه‌بر این‌که پوشش گیاهی را نشان می‌دهد یک شاخص اقلیمی هم در این پژوهش باشد. به‌علت کمبود تعداد ایستگاه‌های هواشناسی در منطقه مورد مطالعه امکان تهیه نقشه اقلیمی فراهم نگردید. در منطقه مطالعاتی در قسمت‌های جنوبی بارندگی بیش‌تر است و در نتیجه این قسمت پوشش گیاهی بهتری دارد و خاک‌های این قسمت ماده آلی بیش‌تری دارند و هرچه به‌سمت شمال حرکت می‌کنیم مقدار بارندگی کم‌تر و مقدار پوشش هم کم‌تر می‌شود و خاک‌هایی با شوری متوسط تا زیاد و ماده آلی کم‌تر داریم. در استان گلستان اقلیم بیش‌ترین نقش را در تشکیل خاک دارد بنابراین در این‌جا اهمیت زیاد شاخص پوشش گیاهی می‌تواند نشان‌دهنده اهمیت اقلیم در کلاس‌های پیش‌بینی شده باشد. بعد از شاخص پوشش گیاهی، ژئومورفولوژی و ارتفاع در هر دو روش تعیین اهمیت متغیرها در تخمین نقشه بودند (شکل ۶). نقش مهم ژئومورفولوژی در پیش‌بینی و دقت نقشه توسط جعفری و همکاران (۲۰۱۲) نیز گزارش شده است، آن‌ها دریافتند که ژئومورفولوژی مهم‌ترین متغیر در تخمین نقشه کلاس‌های خاک بوده است. گریم و همکاران (۲۰۰۸) مشاهده کردند که توپوگرافی مهم‌ترین متغیر در تخمین ذخیره کربن آلی در مطالعه آنان بوده است.

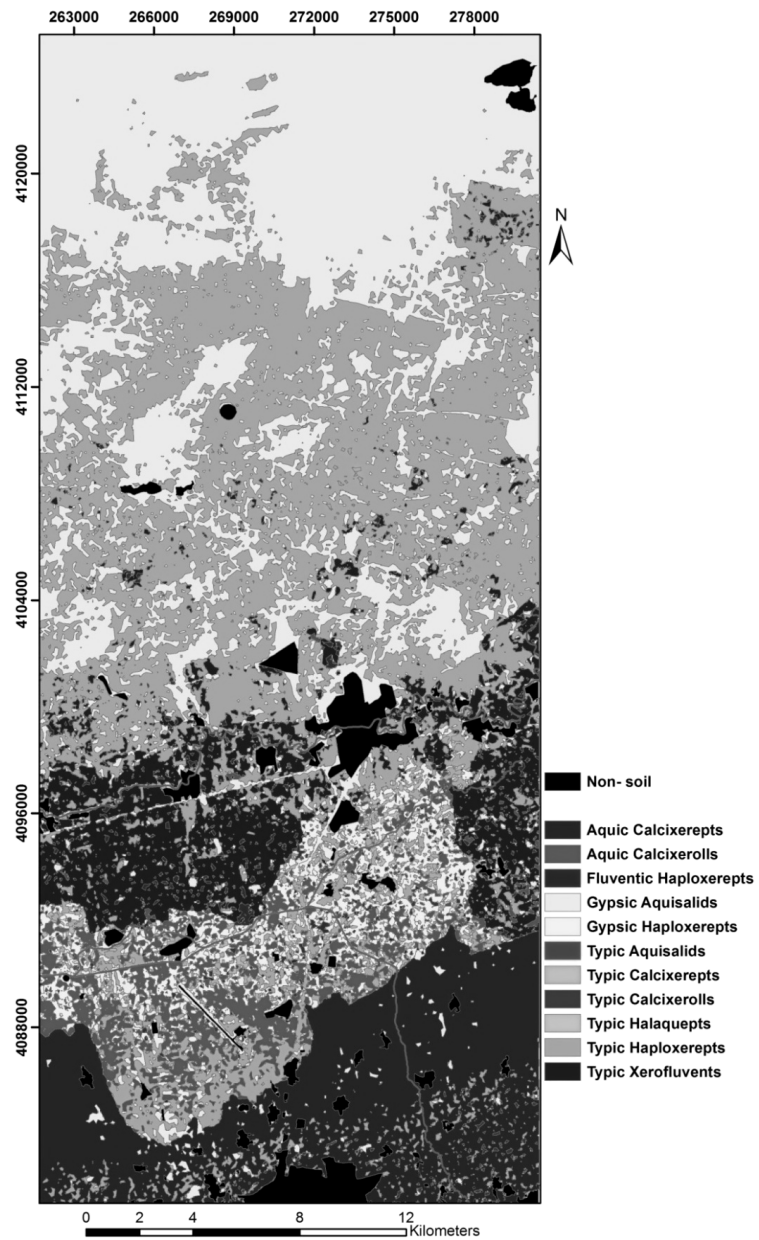


شکل ۶- مقدار اهمیت متغیرهای استفاده شده در مدل‌سازی.

توزیع مکانی کلاس‌های خاک با RF: نقشه پیش‌بینی زیرگروه‌های خاک در منطقه مورد مطالعه در شکل ۷ نشان داده شده است. همان‌طور که در شکل نشان داده شده است بیش‌ترین سطح مربوط به زیرگروه‌های Gypsic Aquisalids و Typic Haploxerepts می‌باشد. زیرگروه Gypsic Aquisalids در اراضی دیده می‌شود که سطح آب زیرزمینی به نسبت بالا بود و مشکل زهکشی داشتند و جزو اراضی پست می‌باشند. قسمت زیادی از زیرگروه بزرگ Typic Haploxerepts در قسمت شمالی محدوده مطالعاتی قرار دارند و قسمت کمی نیز در قسمت‌های جنوبی مشاهده می‌شوند. این دو کلاس خاک تعداد نقاط مشاهداتی بیش‌تری نیز نسبت به سایر زیرگروه‌ها داشتند (جدول ۳). زیرگروه Aquic Calcixerepts در قسمت‌های جنوبی نقشه تخمینی جایی که مقدار بارندگی آن بیش‌تر از سایر قسمت‌های محدوده مطالعه قرار دارد. به دلیل بارندگی بیش‌تر آهک اولیه در آب باران حل و در قسمت‌های عمقی خاک رسوب کرده و افق کلسیک را تشکیل داده است. زیرگروه Typic Xerofluvents در اطراف رودخانه گرگانرود قرار دارد و تحت تأثیر رسوبات این رودخانه می‌باشند. زیرگروه‌های بزرگ Aquic Calcixerolls، Typic Aquisalids، Typic Calcixerepts، Typic Halaquepts مقدار کمی از محدوده مطالعاتی را در نقشه تخمینی تشکیل می‌دهد و هر کدام در نقشه تخمینی سطحی زیر ۱ درصد را تشکیل می‌دهد (جدول ۳). این کلاس‌های خاک همچنین تعداد نقاط مشاهداتی کم‌تری نسبت به سایر کلاس‌ها داشتند.

نتیجه‌گیری

نتایج این پژوهش نشان داد که تکنیک RF یک روش مناسب و با دقت قابل می‌باشد که می‌تواند با مقدار کم نمونه تخمین مناسبی ارائه دهد. مقدار دقت نقشه با نمونه‌برداری بیش‌تر از کلاس‌هایی که دارای دقت کم و تعداد نمونه کم هستند می‌توانست با انجام نمونه‌برداری بیش‌تر افزایش یابد. استفاده از سایر متغیرهای محیطی که در این مطالعه استفاده نشد و استفاده از متغیرهای با بزرگ‌نمایی بالاتر مانند استفاده از نقشه DEM با بزرگ‌نمایی ۱۰ متر در این گونه مطالعات سبب بهبود دقت نقشه تولیدی می‌گردد. دقت نقشه‌های تولیدی با روش‌های نقشه‌برداری رقومی، بستگی زیادی به دقت ورودی‌ها دارد.



شکل ۷- نقشه پیش‌بینی شده زیرگروه‌های بزرگ خاک.

منابع

1. 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.
2. Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J. 1984. *Classification and Regression Trees*. Chapman and Hall, New York. 368p.
3. Breiman, L., and Cutler, A. 2004. *Random Forests*. Department of Statistics, University of Berkeley. http://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm.
4. 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.
5. Cook, S.E., Jarvis, A., and Gonzalez, J.P. 2008. A New Global Demand for Digital Soil Information, P 31-43. In: Hartemink, A.E., McBratney, A., Mendonca-Santos, M.L., (eds.), *Digital Soil Mapping with Limited Data*. Springer, Dordrecht Heidelberg London New York.
6. Grimm, R., Behrens, T., Marker, M., and Elsenbeer, H. 2008. Soil organic carbon concentrations and stocks on Barro Colorado Island - Digital soil mapping using random forests analysis. *Geoderma*. 146: 1-2: 102-113.
7. Grunwald, S. 2010. Current State of Digital Soil Mapping and What Is Next, P 3-12. In: Boettinger, J.L., Howel, D.W., Moore, A.C., Hartemink, A.E., and Kienast-Brown, S., (eds.), *Digital Soil Mapping: Bridging Research, Environmental Application, and Operation*. Springer. Dordrecht Heidelberg London New York.
8. Huete, A.R. 1988. A soil adjusted vegetation index (SAVI). *Remote Sensing of Environment*. 25: 295-309.
9. Jafari, A., Finke, P.A., Van de Wauw, 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. *Europ. J. Soil Sci.* 63: 284-309.
10. Jenny, H. 1941. *Factors of Soil Formation: A System of Quantitative Pedology*. McGraw-Hill, New York.
11. Ließ, 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.
12. Mathworks. 2009. *Matlab The Mathworks. Inc, Natick, MA*.
13. McBratney, A.B., Mendonça Santos, M.L., and Minasny, B. 2003. On digital soil mapping. *Geoderma*. 117: 1-2. 3-52.
14. Minasny, B., and McBratney, A.B. 2006. A conditioned Latin hypercube method for sampling in the presence of ancillary information. *Computers and Geosciences*. 32: 9. 1378-1388.
15. Minasny, B., and McBratney, A.B. 2007. Latin hypercube sampling as a tool for digital soil mapping, P 153-165. In: Lagacherie, P., McBratney, A.B., Voltz, M. (Eds.), *Digital Soil Mapping: An introductory perspective*. Elsevier, Amsterdam.

16. Mudler, V.L., Bruin, S., and Schaepman, M.E. 2013. Representing major soil variability at regional scale by constrained Latin Hypercube Sampling of remote sensing data. *Inter. J. Appl. Earth Observ. Geoinfo.* 21: 301-310.
17. R Development Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
18. 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, S. Kienast-Brown (Eds.), *Digital Soil Mapping: Bridging Research, Environmental Application and Operation*. Progress in Soil Science. Springer, New York.
19. Schoeneberger, P.J., Wysocki, D.A., Benham, E.C., and Broderson, W.D. (Eds.). 2002. Field book for describing and sampling soils, Version 2.0. National Soil Survey Center Natural Resources Conservation Service U.S. Department of Agriculture.
20. Scull, P., Franklin, J., Chadwick, O.A., and McArthur, D. 2003. Predictive soil mapping: a review. *Progress in Phys. Geogr.* 27: 171-197.
21. Soil Survey Staff. 1999. *Soil Taxonomy*. 2nd ed. U.S. Department of Agriculture, Natural Resources Conservation Service, Washington, D.C.
22. Soil Survey Staff. 2010. *Keys to Soil Taxonomy*. U.S. Department of Agriculture, Natural Resources Conservation Service.
23. 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, W. A.C. Moore, A. Hartemink, E. S. Kienast-Brown (Eds.), *Digital Soil Mapping: Bridging Research, Environmental Application, and Operation*. Progress in Soil Science. Springer, Logan, USA.
24. 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.
25. Taghizadeh-Mehrjardi, R., Minasny, B., Triantafilis, J., Sarmadian, F., and Omid, M. 2014. Digital mapping of soil classes using decision tree and auxiliary data in the Ardakan region, Iran. *Arid Land Research and Management* (In Press)
26. Toomanian, N., Jalalian, A., Khademi, H., and Karimian Eghbal, M. 2006. Pedodiversity and pedogenesis in Zayandeh-rud Valley, Central Iran. *Geomorphology*. 81: 376-393.
27. Wilson, J.P. 2009. The role of geocomputation in the hydrological sciences. International Symposium on Geocomputation and Analysis. Wuhan, China.
28. 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.



Gorgan University of Agricultural
Sciences and Natural Resources

J. of Water and Soil Conservation, Vol. 21(6), 2015
<http://jwsc.gau.ac.ir>

Digital soil mapping using Random Forest model in Golestan province

***M.R. Pahlavan Rad¹, F. Khormali², N. Toomanian³,
F. Kiani⁴ and Ch.B. Komaki⁵**

¹Ph.D. Graduate, Dept. of Soil Science, Gorgan University of Agricultural Sciences and Natural Resources, ²Professor, Dept. of Soil Science, Gorgan University of Agricultural Sciences and Natural Resources, ³Assistant Prof., Dept. of Soil and Water, Agricultural and Natural Resources Research Center of Isfahan, ⁴Assistant Prof., Dept. of Soil Science, Gorgan University of Agricultural Sciences and Natural Resources, ⁵Assistant Prof., Dept. of Arid Zone Management, Gorgan University of Agricultural Sciences and Natural Resources

Received: 11/03/2013; Accepted: 03/29/2014

Abstract

Soil maps are the major sources for management of lands, natural resources and environmental aspects. Digital soil mapping methods can decrease the cost of mapping and are not time-consuming. Conditioned Latin Hypercube Sampling for determining the sites of sampling and Random Forest to predict soil sub group classes were evaluated at an area of 85000 ha in Golestan province. 12 environmental variables including terrain attribute, geomorphology units and vegetation cover index were used as predictor parameters. There were 11 sub group classes in the soils of study area. The results revealed that the lowest OOB estimate error rate in modeling was 52.53%. The soil classes with higher frequency had lower OOB error. Gypsic Aquisalids with 22 samples and Aquic Calcixerepts with 9 and Typic Haploxerepts with 22 samples had the OOB error rates 22, 25 and 37%, respectively. Gypsic Haploxerepts (6 samples), Typic Aquisalids (5 samples), Typic Calcixerepts (4 samples), Typic Calcixerolls (4 samples) and Typic Halaquepts (3 samples) had OOB error rate of 100%. SAVI, geomorphology, elevation and aspect were the most importance parameters in prediction of soil map and increasing model accuracy. Results indicate that RF technique can be a reliable and appropriate method to give satisfactory results with small samples.

Keywords: Digital soil mapping, Random forest, Map accuracy, Golestan province

* Corresponding Author; Email: pahlevan354@yahoo.com

