

نقشه‌برداری رقومی برخی ویژگی‌های خاک سطحی با استفاده از دو رویکرد جنگل تصادفی و منطق فازی

(مطالعه موردی: بخشی از اراضی کوهین، استان قزوین)

اصغر رحمانی^{۱*}، فریدون سرمدیان^۲، سید روح‌اله موسوی^۱، سید عرفان خاموشی^۱
^۱ دانشجوی دکتری گروه علوم و مهندسی خاک پردیس کشاورزی و منابع طبیعی، دانشگاه تهران
^۲ استاد گروه علوم و مهندسی خاک پردیس کشاورزی و منابع طبیعی، دانشگاه تهران

چکیده

روش معمول نقشه برداری ویژگی‌های خاک، وابسته به برداشت‌های میدانی بوده و اطلاعات حاصل از تغییرات مکانی آنها متأثر از مقیاس مطالعاتی و نیاز به دانش کارشناسان باتجربه دارد. لذا استفاده از رویکردهای جدید داده‌کاوی در تهیه نقشه رقومی ویژگی‌های خاک ضرورت دارد. در این پژوهش از ۶۲ نقطه خاک نمونه‌برداری و ویژگی‌های درصد کربن آلی، رس و آهک اندازه‌گیری گردید. داده‌ها به دو دسته ۸۰ درصد (۵۰ نقطه) برای آموزش و ۲۰ درصد (۱۲ نقطه) جهت اعتبارسنجی دسته‌بندی گردید. براساس آنالیز مولفه‌های اصلی، سه متغیر شاخص موقعیت توپوگرافی، شاخص شدت پستی و بلندی، ارتفاع به عنوان مؤثرترین متغیرها شناسایی شدند. نتایج مدل‌سازی جنگل تصادفی (RF) برای ویژگی‌های کربن آلی، رس و آهک براساس آماره‌های ضریب تبیین (R^2) به ترتیب مقادیر ۰/۶۳، ۰/۷۵ و ۰/۶۳ و ریشه میانگین مربعات خطا (RMSE) مقادیر ۰/۱۷، ۷/۵ و ۵/۷۷ درصد و برای رویکرد SoLIM مقادیر ضریب تبیین ۰/۴۷، ۰/۴۲ و ۰/۴۲ و مقادیر ریشه میانگین مربعات خطای ۰/۲، ۸/۰۸ و ۴/۶۸ درصد حاصل گردید. رویکرد جنگل تصادفی می‌تواند نقشه‌های رقومی را با دقت مناسب برای مدیریت و بهره‌برداری پایدار از اراضی پیش‌بینی نماید.

کلمات کلیدی: داده‌کاوی، جنگل تصادفی، مدل استنباطی خاک-زمین نما، منطق فازی.

مقدمه

نقشه‌برداری خاک‌ها به روش معمول وابسته به برداشت‌های میدانی بوده و اطلاعات حاصل از تغییرات مکانی آنها، متأثر از مقیاس مطالعاتی می‌باشد (McBratney و همکاران ۲۰۰۳). ارتباط بین ویژگی‌های خاک و متغیرهای محیطی بصورت یک رابطه خطی، چند جمله‌ای یا حتی بسیار پیچیده‌تر می‌تواند باشد (Taghizadeh-Mehrjardi et al., 2016). روش‌های یادگیری ماشین^۱ متعددی از قبیل کوبیست^۲ (Taghizadeh-Mehrjardi et al., 2016)، جنگل تصادفی^۳ (Pahlavan-Rad and Akbarimoghaddam, 2018)، رگرسیون درختی^۴ (Taghizadeh-Mehrjardi et al., 2014) به‌منظور شناخت روابط بین ویژگی‌های خاک و متغیرهای محیطی مورد استفاده واقع شده‌اند. مطالعات متعددی توسط محققان در سال‌های اخیر بر روی مدل‌سازی و پیش‌بینی مکانی کلاس‌های خاک صورت گرفته است (Taghizadeh et al., 2015؛ Pelegrino et al., 2016) اما به نسبت در مطالعات کمتری به نقشه‌برداری رقومی ویژگی‌های خاک پرداخته شده است. Hengl و همکاران (۲۰۱۵) به پیش‌بینی مکانی ویژگی‌های خاک در کشور آفریقا پرداختند و نتیجه گرفتند که مدل جنگل تصادفی نسبت به روش رگرسیون خطی دارای صحت بیشتری می‌باشد. مدل استنباطی خاک-زمین نما^۵ و درخت تصمیم^۶ برای استخراج دانش صریح از طریق قوانین^۷

* ایمیل نویسنده مسئول: a.rahmani@ut.ac.ir

¹ Machine learning

² Cubist

³ Random Forest

⁴ Regression Tree

⁵ Soil Land Inference Model (SoLIM)

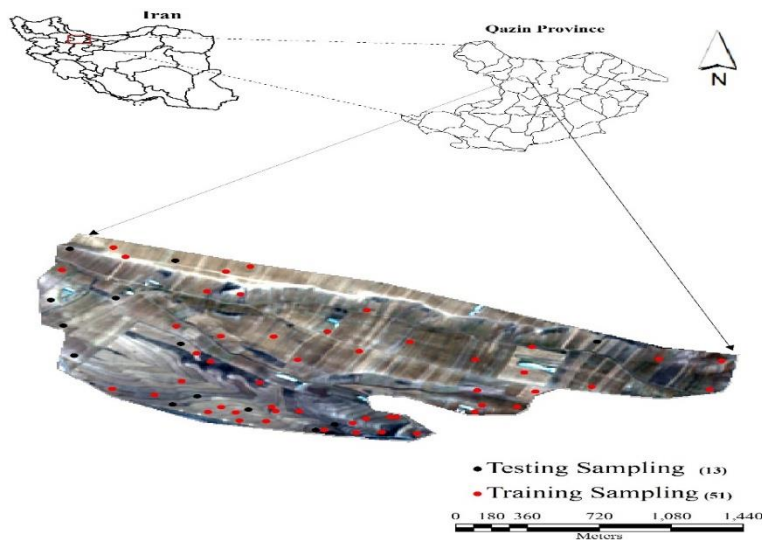
⁶ Decision Tree

⁷ Rule

و توابع عضویت فازی^۱ در فهم پیدایش و نقشه برداری رقومی خاکها به کار گرفته می شوند (Zeng و همکاران، ۲۰۱۷). توابع عضویت فازی به دو صورت قیاسی^۲ (بر اساس دانش کارشناس) و استقرایی^۳ (بر اساس داده) طبقه بندی می گردند (Kaufmann و همکاران، ۲۰۱۵). در پژوهش Zeng و همکاران (۲۰۱۷) از منحنی های پاسخ کلاس های خاک نسبت به شرایط محیطی مستخرج شده از مدل جنگل تصادفی برای تعریف توابع عضویت فازی استفاده نمودند. با توجه به بررسی های بعمل آمده تاکنون از روشهای داده کاوی و مدل استنباطی خاک-زمین نما به صورت جداگانه برای تهیه نقشه رقومی ویژگی های خاک استفاده شده است. بنابراین، مقایسه ای این دو رویکرد داده کاوی برای پیش بینی مکانی درصد ماده آلی، رس و آهک خاک سطحی با توجه به تاثیر گذاری آنها در انباشت و قابلیت استفاده رطوبت و عناصر غذایی برای افزایش عملکرد در اراضی دیم کوهین قزوین به عنوان هدف اصلی مورد بررسی قرار می گیرد.

مواد و روشها

منطقه مورد مطالعه بین مدار ۳۶ درجه و ۲۳ دقیقه تا ۳۶ درجه و ۲۰ دقیقه عرض شمالی و ۴۹ درجه و ۳۴ دقیقه تا ۴۹ درجه و ۳۸ دقیقه طول شرقی واقع شده است، مساحت ایستگاه به همراه اراضی مورد مطالعه ۳۷۰ هکتار می باشد (شکل ۱). محل ۶۲ نمونه خاک سطحی در عمق ۰-۲۰ سانتی متر بر اساس روش نمونه برداری شبکه و نظر کارشناس تعیین و با استفاده از سامانه موقعیت یاب جهانی^۴ مدل گارمین در سیستم تصویر UTM متریک ثبت و برداشت گردید. درصد کربن آلی (والکی بلک)، کربنات کلسیم معادل (به روش حجم سنجی) و درصد رس (روش هیدرومتری) بر اساس روش های استاندارد اندازه گیری گردید.



شکل ۱- نقاط مشاهداتی آموزش و آزمون بر روی تصویر ماهواره ای لندست ۸

برای استخراج مشتقات اولیه و ثانویه ژئومورفومتری از مدل رقومی ارتفاع با قدرت تفکیک مکانی ۱۰ متر تهیه شده از سازمان نقشه برداری کشور استفاده شد، پارامترهای ژئومورفومتری شامل مساحت زهکشی^۵، فاصله تا شبکه آبراهه ها^۶، ارتفاع از سطح شبکه آبراهه ها^۷، فاکتور طول در شیب^۸، ارتفاع، تابش پخشیده^۱، تابش مستقیم^۲، مساحت حوضه زهکشی اصلاح شده^۳، شاخص همواره دره با قدرت تفکیک مکانی بالا^۴،

¹ Fuzzy membership function

² Deductive

³ Inductive

⁴ Global position system

⁵ Catchment area

⁶ Channel network distance

⁷ Channel network base level

⁸ LS Factor

شاخص همواری قله با قدرت تفکیک مکانی بالا^۵، موقعیت نسبی شیب^۶، ارتفاع نرمال شده^۷، شاخص خیزی ساگا^۸، درصد شیب^۹، شاخص قدرت جریان^{۱۰}، شاخص موقعیت توپوگرافیکی^{۱۱}، شاخص زبری پستی و بلندی^{۱۲}، عمق دره^{۱۳}، شاخص در معرض باد قرارگرفتن^{۱۴} در نرم افزار SAGA GIS نسخه 7.3 محاسبه و استخراج گردید. از روش آنالیز چند متغیره تجزیه مولفه‌های اصلی^{۱۵} برای انتخاب متغیرهای موثر در نرم افزار SPSS نسخه ۲۵ استفاده شد.

برای مدلسازی پیش‌بینی مکانی ویژگی‌های خاک در رویکرد داده‌کاوی جنگل تصادفی از بسته "Random Forest" در نرم افزار Rstudio نسخه 1.0.131 و برای آنالیز اهمیت^{۱۶} هم‌متغیرهای محیطی از شاخص درصد افزایش میانگین مربعات خطای پیش‌بینی (InMSE%) از تابع "Var Important" استفاده شد. در رویکرد مدل استنباطی خاک-زمین نما از پروژه‌ی نمونه مدار^{۱۷} نرم افزار SoLIM Solutions نسخه ۲۰۱۵ استفاده شد (Zhu و همکاران، ۲۰۱۵).

نقاط اندازه‌گیری شده با استفاده از روش نمونه برداری بهینه بوت استرپت^{۱۸} الگوریتم جنگل تصادفی (Viscarra Rossel, 2014) در نرم‌افزار RStudio 1.0131 به دو دسته ۸۰ درصد (۵۰ نقطه) برای مرحله آموزش و ۲۰ درصد (۲۰ نقطه) برای اعتبارسنجی تقسیم گردید. نقشه رقومی پیش‌بینی ویژگی‌های درصد ماده آلی، رس و آهک با استفاده از دو آماره‌ی ریشه میانگین مربعات خطا (RMSE) و ضریب تبیین R^2 مورد ارزیابی قرار گرفت.

$$R^2 = \frac{\sum_{i=1}^n (Pi - \bar{O}i)}{\sum_{i=1}^n (Oi - \bar{O}i)} \quad (1)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Oi - Pi)^2} \quad (2)$$

Oi: مقدار اندازه‌گیری شده Pi: مقدار پیش‌بینی شده $\bar{O}i$: میانگین اندازه‌گیری شده n: تعداد مشاهدات

نتایج و بحث

- ¹ Diffuse insolation
- ² Direct insolation
- ³ Modified catchment area
- ⁴ Multiresolution valley bottom flatness(MrVBF)
- ⁵ Multiresolution ridge tope flatness (MRTTF)
- ⁶ Relative slope position (RSP)
- ⁷ Normalized Height
- ⁸ SAGA Wetness index
- ⁹ Slope percent
- ¹⁰ Stream power index
- ¹¹ Topographic slope position
- ¹² Terrain roughness index
- ¹³ Valley depth
- ¹⁴ Wind explosion index
- ¹⁵ Principle component analysis
- ¹⁶ Relative important
- ¹⁷ Sample-based project
- ¹⁸ Bootstrap

نتایج خلاصه آماری ویژگی‌های خاک در جدول ۱ نشان داده شده است. کربن آلی خاک از ۰/۱۳ تا ۱/۳۴ درصد با میانگین ۰/۷۶۱ و ضریب تغییرات ۳۴/۴۸ درصد، ویژگی رس از حداقل ۲۵ تا ۶۰ درصد با میانگین ۴۱/۷۷ و ضریب تغییرات ۱۹/۴۹ درصد و آهک از حداقل ۸/۰۲ تا ۳۱/۴۲ درصد با میانگین ۷/۸۸ و ضریب تغییرات ۲۸/۵۸ درصد و بر اساس استاندارد (Wildig, 1985) در کلاس تغییرپذیری متوسط (۱۵-۳۵ درصد) قرار می‌گیرند. Zeraatpisheh و همکاران (۲۰۱۹) نیز نتایج مشابه با این تحقیق در ارتباط با ضریب تغییرات پارامترهای کربن آلی، رس و آهک گزارش نمودند.

جدول ۱- خلاصه آماری ویژگی‌های خاک سطحی

متغیر	میانگین	میانه	حداقل	حداکثر	انحراف معیار	درصد ضریب تغییرات
کربن آلی	۰/۷۶۱	۰/۸۲	۰/۱۳	۱/۳۴	۰/۲۷۷	۳۴/۴۸
رس	۴۱/۷۷	۴۲/۵	۲۵	۶۰	۸/۱۴	۱۹/۴۹
آهک	۱۷/۸۸	۱۶/۴۵	۸/۰۲	۳۱/۴۲	۵/۱۱	۲۸/۵۸

بر اساس مقدار ارزش ویژه^۱ بالاتر از یک، ۵ مولفه اصلی اول ۸۰/۲۶ درصد واریانس موجود در ویژگی‌های ژئومورفوتری را توجیه می‌نمایند (شکل ۲). نتایج همبستگی ویژگی‌های ژئومورفوتری با مولفه‌های اصلی با ارزش ویژه بالای یک نشان داد که در مولفه‌های اول تا پنجم، پارامترهای شاخص موقعیت توپوگرافی، شاخص شدت پستی و بلندی، ارتفاع، مساحت حوزه زهکشی و موقعیت میانی شیب به ترتیب بالاترین همبستگی را به خود اختصاص دادند و در نهایت با توجه به نظر کارشناس و به دلیل همبستگی بالای موقعیت میانی شیب با موقعیت توپوگرافیکی و عدم تغییر مکانی مساحت حوزه زهکشی، از سایر پارامترها (DEM, Tri, TPI) به همراه نقشه رستری لندفرم برای مدلسازی مورد استفاده گردید.

جدول ۲- نتایج واریانس توجیحی مولفه‌های اصلی

مولفه اصلی	درصد واریانس	درصد تجمعی واریانس
۱	۲۳/۴۵۱	۲۳/۴۵۱
۲	۲۳/۳۲۹	۴۶/۷۸۰
۳	۱۴/۰۵۰	۶۰/۸۳۰
۴	۱۳/۸۸۷	۷۴/۷۱۷
۵	۵/۵۴۰	۸۰/۲۵۸

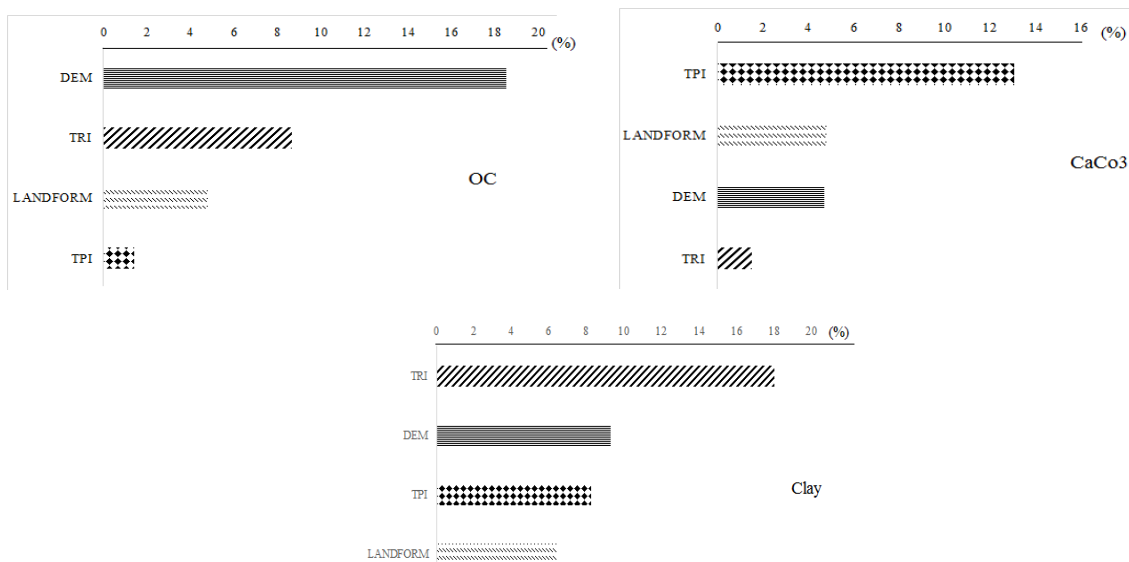
نتایج اعتبارسنجی ارائه شده در جدول ۳ نشان می‌دهد که کربن آلی خاک در دو رویکرد داده‌کاوی جنگل تصادفی (RF) و مدل استنباطی خاک-زمین‌نما (SoLIM) به ترتیب دارای ضریب تبیین ۰/۶۳ و ۰/۴۷ با ریشه میانگین مربعات خطای ۰/۱۷ تا ۰/۲ درصد، ویژگی رس دارای ضریب تبیین ۰/۷۵ و ۰/۴۲ با ریشه میانگین مربعات خطا ۷/۵ تا ۸/۰۸ درصد و ویژگی آهک خاک سطحی دارای ضریب تبیین ۰/۶۳ و ۰/۴۲ همراه با ریشه میانگین مربعات خطا ۵/۷۷ و ۴/۶۸ درصد می‌باشند Zeraatpisheh و همکاران (۲۰۱۹) برای ویژگی‌های کربن آلی، رس و آهک، مقادیر ضرایب تبیین ۰/۵۵، ۰/۱۵ و ۰/۳۰ با ریشه میانگین مربعات خطای ۰/۳۳، ۷/۸۶ و ۹/۵۲ درصد در استفاده از مدل جنگل تصادفی به عنوان بهترین مدل در مقایسه با مدل کوبیست، رگرسیون درختی و رگرسیون چندمتغیره خطی گزارش نمودند. در نقشه-برداری رقومی ویژگی‌های خاک مقادیر R^2 بالاتر از ۰/۷ غیر معمول و مقادیر ۰/۵ و یا کمتر رایج می‌باشد (Malone, 2009).

¹ Eigen value

جدول ۳ - نتایج اعتبار سنجی رویکردهای داده کاوی در مدلسازی پیش بینی ویژگیهای خاک

مدل استنباطی خاک-زمین نما (SoLIM)		جنگل تصادفی (Random forest)		مدل
R ²	RMSE(%)	R ²	RMSE(%)	
۰/۴۷	۰/۲	۰/۶۳	۰/۱۷	کربن آلی
۰/۴۲	۸/۰۸	۰/۷۵	۷/۵	رس
۰/۴۲	۴/۶۸	۰/۶۳	۵/۷۷	آهک

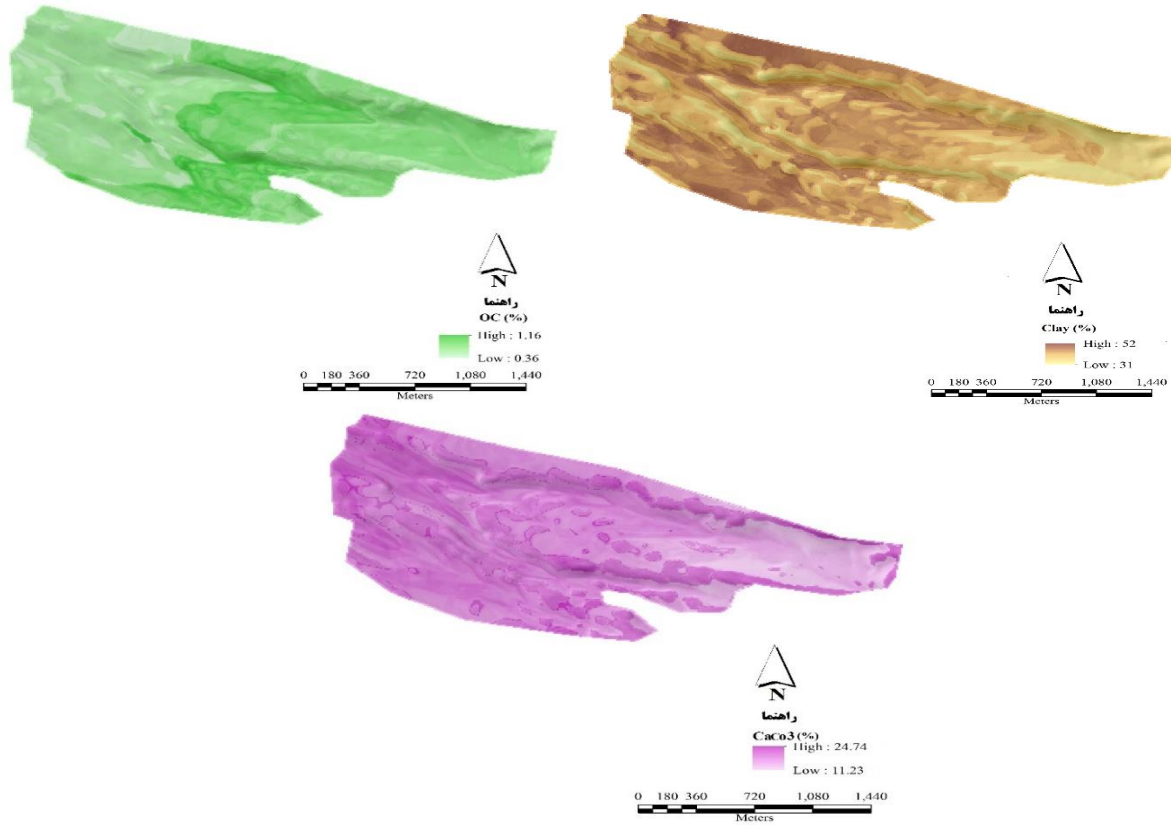
همچنین نتایج آنالیز اهمیت نسبی هم متغیرهای ژئومورفومتری بر اساس شکل ۳ نشان داد که به ترتیب ارتفاع، شاخص شدت پستی و بلندی، لندفرم و شاخص موقعیت توپوگرافی بر کربن آلی خاک سطحی و شاخص شدت پستی و بلندی، ارتفاع، شاخص موقعیت توپوگرافی و لندفرم بر درصد رس خاک سطحی و شاخص موقعیت توپوگرافی، لندفرم، ارتفاع و شاخص شدت پستی و بلندی از لحاظ عدم حضور و تاثیر آن بر افزایش %MSE پیش‌بینی مدل جنگل تصادفی، موثرترین پارامترها می‌باشند.



شکل ۳- آنالیز اهمیت نسبی پارامترهای ژئومورفومتری موثر بر پیش بینی ویژگی های خاک سطحی

نقشه‌های پیش‌بینی مکانی رویکرد جنگل تصادفی به عنوان مدل مناسب و با دقت بالاتر در شکل ۴ ارائه شده است. بر اساس نتایج حاصل، بیشترین میزان کربن آلی خاک سطحی در نواحی مرکزی مشاهده می‌شود و از دلایل آن می‌توان به کشاورزی حفاظتی به کار گرفته شده در ایستگاه کوهین و نقش آن در حفظ کربن خاک سطحی و در نواحی خارج از ایستگاه به مطابقت داشتن مقادیر بالای کربن آلی با سطوح پایدار ژئومورفیک که در نقشه لندفرم به صورت دشت جدا گردیده، اشاره نمود. الگوی مکانی درصد رس و آهک بصورت برعکس یکدیگر تغییر می‌نماید بطوریکه بیشترین میزان آهک در نواحی تیغه‌ها و موقعیت‌های شانه شیب و در این نواحی کمترین میزان رس قابل مشاهده است بر اساس مشاهدات صحرایی و مطابقت با نقشه لندفرم‌ها این نواحی دارای فرسایش پذیری سطحی شدید بصورت ورقه‌ای می‌باشد که نتیجه آن بروز آهک لایه‌های زیرین و شسته شدن رس از این لایه‌ها به سمت پایین دست شیب می‌باشد. همچنین در نواحی سطوح پایدار ژئومورفیک

(دشت‌ها در نقشه لندفرم) و داخل ایستگاه کوهین به دلیل اقدامات حفاظتی میزان رس که در بخش‌های مرکزی نقشه قابل مشاهده است، بیشتر می‌گردد.



شکل ۴- نقشه پیش بینی مکانی ویژگی‌های خاک سطحی با استفاده از جنگل تصادفی

نتیجه‌گیری

نتایج بیانگر آن است که استفاده از روش‌های جدید آنالیز چندمتغیره مانند PCA، همبستگی و همراستایی خطی متغیرهای ژئومورفومتری استخراج شده از مدل رقومی ارتفاع را می‌توان کاهش داد و موثرترین متغیرها با حداکثر واریانس را استخراج و زمان انجام محاسبات تهیه نقشه را کاهش داد. در این پژوهش تنها با استفاده از سه متغیر شاخص موقعیت توپوگرافی، شاخص شدت پستی و بلندی، ارتفاع بدست آمده از تجزیه و تحلیل PCA و نقشه رستری لندفرم، می‌توان نقشه رقومی ویژگی‌های درصد کربن آلی، رس و آهک خاک سطحی را با استفاده از مدل جنگلی تصادفی نسبت به مدل استنباطی خاک-زمین نما با دقت بالاتر تهیه نمود که در مدیریت خاک و بهره برداری از اراضی مورد استفاده واقع گردند و پهنه‌های مدیریتی را با شناسایی تغییرپذیری پراکنش این ویژگی‌ها جداسازی نمود.

منابع

- Kaufmann, M., Meier, A., & Stoffel, K. 2015. IFC-Filter: Membership function generation for inductive fuzzy classification. *Expert Systems with Applications*, 42(21), 8369-8379.
- Hengl, T., Heuvelink, G.B., Kempen, B., Leenaars, J.G., Walsh, M.G., Shepherd, K.D., Sila, A., MacMillan, R.A., de Jesus, J.M., Tamene, L., 2015. Mapping soil properties of Africa at 250 m resolution: random forests significantly improve current predictions. *PLoS One* 10, e0125814.



- Pahlavan-Rad, M.R., Akbarimoghaddam, A., 2018. Spatial variability of soil texture fractions and pH in a flood plain (case study from eastern Iran). *Catena* 160, 275–281.
- Pelegriño, M. H. P., Silva, S. H. G., Menezes, M. D. D., Silva, E. D., Owens, P. R., & Curi, N. 2016. Mapping soils in two watersheds using legacy data and extrapolation for similar surrounding areas. *Ciencia e Agrotecnologia*, 40(5), 534-546.
- Taghizadeh-Mehrjardi, R., Minasny, B., Sarmadian, F., Malone, B., 2014. Digital mapping of soil salinity in Ardakan region, central Iran. *Geoderma* 213, 15–28.
- Taghizadeh-Mehrjardi, R., Nabiollahi, K., Kerry, R., 2016. Digital mapping of soil organic carbon at multiple depths using different data mining techniques in Baneh region, Iran. *Geoderma* 266, 98–110.
- Viscarra Rossel, R. A., Webster, R., & Kidd, D. (2014). Mapping gamma radiation and its uncertainty from weathering products in a Tasmanian landscape with a proximal sensor and random forest kriging. *Earth Surface Processes and Landforms*, 39(6), 735-748.
- Wilding, L. P. (1985). Spatial variability: its documentation, accommodation and implication to soil surveys. In *Soil spatial variability. Workshop* (pp. 166-194)
- Zeraatpisheh, M., Ayoubi, S., Jafari, A., Tajik, S., & Finke, P. 2019. Digital mapping of soil properties using multiple machine learning in a semi-arid region, central Iran. *Geoderma*, 338, 445-452.
- Zeng, C., Yang, L., & Zhu, A. (2017). Construction of membership functions for soil mapping using the partial dependence of soil on environmental covariates calculated by random forest. *Soil Science Society of America Journal*, 81(2), 341-353.
- Zhu, A.X., Burt, J.E., Du, F. (2015). Digital soil mapping software: SoLIM Solutions. user manual.



16th Iranian Soil Science Congress

University of Zanjan, Iran, August 27-29, 2019



Topic for submission: Soil genesis and classification

Digital mapping of some surface soil properties using RF and Fuzzy approaches (A case study: Kuhin region, Qazvin province)

Rahmani^{1*}, A., Sarmadian², F., Mousavi¹, S.R. khamoshi¹, SE.

^{1*} Ph.D. Student, Soil Science and engineering College of Agriculture and natural resources University of Tehran, Iran

² Prof., Soil Science and engineering Department, College of Agriculture and natural resources University of Tehran, Iran

Abstract

Conventional soil mapping approaches rely on field achievement and its spatial variation affected by survey scale and expert knowledge also data mining methods is essential for preparing soil characteristics mapping. In this study 65 sample points were selected and soil organic carbon, clay and CaCO₃ had been determined. Data in two grouped such as calibration (20%) and validation (80%) were divided. PCA analysis showed that three covariates as an example TPI, TRI and DEM mostly essential variables. RF results demonstrated that SOC, Clay and CaCO₃ based on R² are 0.63, 0.75 and 0.63, RMSE 0.17, 7.5 and 5.77, respectively. SoLIM approach showed that R² value 0.47, 0.42 and 0.42, RMSE 0.2, 8.08 and 4.68, respectively were achieved. RF method can be predicted digital maps with appropriate precise for sustainable land management.

Key words: Data mining, Random forest, Soil-land inference model, fuzzy logic

*Corresponding author: a.rahmani@ut.ac.ir