تهیه نقشه رقومی خاک با استفاده از مدل جنگل‌های تصادفی در منطقه آبیک، استان قزوین

سیدعرفان خاموشی، فریدون سرمدیان[1] و علی کشاورزی

دانشجوی دکتری، دانشگاه تهران؛ khamoshierfan@ut.ac.ir

استاد گروه علوم و مهندسی خاک، دانشکده مهندسی و فناوری کشاورزی، دانشگاه تهران؛ fsarmad@ut.ac.ir

استادیار گروه علوم و مهندسی خاک، دانشکده مهندسی و فناوری کشاورزی، دانشگاه تهران؛ alikeshavarzi@ut.ac.ir

دریافت: 96/11/11 و پذیرش: 97/5/24

چکیده

نیاز روز افزون به داشتن اطلاعات دقیق از خاک و همچنین لزوم به روز رسانی آسان این اطلاعات در جهت تعیین روابط بین خاک و زمین‌نما به یکی از مباحث مهم در علوم خاک تبدیل شده است. تکنیک‌های نقشه برداری رقومی خاک به دلیل توانایی در بدست آوردن اطلاعات دقیق در مورد انواع خاک‌ها در مناطق وسیع و همچنین مقرون به صرفه بودن، به عنوان یک راه حل قابل طرح است. هدف از مطالعه حاضر، تهیه نقشه رقومی خاک در منطقه آبیک استان قزوین با استفاده از مدل جنگل‌های تصادفی است. بدین منظور، با استفاده از روش مکعب لاتین مشروط از منطقه نمونه‌برداری شد و پس از انجام آنالیزهای آزمایشگاهی، با استفاده از مدل جنگل‌های تصادفی و متغیرهای کمکی مستخرج از مدل رقومی ارتفاع با قدرت تفکیک مکانی 30 متر و تصاویر ماهواره لندست 8، نقشه خاک منطقه با مدل‌سازی تهیه شد. از مجموع 7261 هکتار اراضی مطالعه شده، بیشترین وسعت مربوط به فامیل Loamy-skeletal, mixed, superactive, thermic Typic Calcixerepts بود. نتایج حاصل نشان داد که مدل‌سازی خاک با استفاده از الگوریتم جنگل های تصادفی توانست کلاس‌های خاک منطقه را با دقت بالایی (ضریب کاپای 0/83) پیش‌بینی کند.

واژه‌های کلیدی: پیش‌بینی مکانی، مدل‌سازی، مکعب لاتین مشروط، متغیر کمکی، فامیل خاک

---

[1] نویسنده مسئول، آدرس: کرج، پردیس کشاورزی و منابع طبیعی دانشگاه تهران، گروه علوم و مهندسی خاک

مقدمه

خاک‌ها به طور فزاینده‌ای به‌عنوان عوامل اصلی خدمات اکوسیستم مانند تولید غذا و غیره، گزارش‌شده‌اند و تقاضا برای به‌روزرسانی و اطلاعات مربوط به خاک در حال افزایش است؛ اما ارتباط برقرار کردن بین این اطلاعات و مخاطبان مختلف به‌عنوان یک چالش باقی می‌ماند که دلیل آن هم استفاده از اصطلاحات فنی و روش‌های مبهم می‌باشد. همچنین، رزولوشن یا دقت مکانی نقشه‌های خاک برای اکثر نقاط جهان برای کمک به مدیریت عملی اراضی بیش از حد کم است. در حالی که سایر علوم زمین (اقلیم شناسی، زمین شناسی و غیره) کمی‌تر شده و از تحول رقومی شدن استفاده کرده‌اند. (پالم و همکاران، 2007).

ینی (1941) معادلهClORPT معرفی کرد که در این معادله، تشکیل خاک بسته به پنج فاکتور اقلیم، ارگانیسم ها، پستی و بلندی، مواد مادری و زمان می‌باشد. در سال 2003 مک براتنی و همکاران با اضافه کردن کلاس یا خصوصیات خاک و موقعیت مکانی معادله ایی جدیدی به صورت $S=f(s, c, o, r, p, a, n)$ ارائه کردند که در این معادله، خصوصیات یا کلاس خاک با یک سری متغیرهای محیطی ارتباط داده می‌شود. به این صورت که در هر موقعیت مشاهده خاک، یک سری متغیرهای محیطی وجود دارد که همبستگی بالای با خصوصیات یا کلاس‌های خاک دارند که یافتن این ارتباط و پیش بینی خصوصیات یا کلاس‌های خاک در نقاط دیگر، منجربه تهیه نقشه رقومی خاک[1] در تمام منطقه مورد مطالعه می‌گردد. نقشه رقومی خاک را می‌توان به صورت «ساماندهی و طبقه‌بندی کلاس یا خصوصیات خاک به وسیله مدل‌های ریاضی مبتنی بر آمار با کمک دانش کارشناس و متغیرهای کمکی » تعریف کرد (لاگاچری و مک براتنی، 2007).

به طور کلی، سه گام اصلی در نقشه‌برداری رقومی خاک وجود دارد، گام اول، شامل داده‌های ورود به سیستم است که با تولید نقشه‌های پایه، استخراج متغیرهای کمکی مکانی به صورت داده‌های رستری از منابع داده در دسترس (تصاویر ماهواره) شروع می‌شود. (مک براتنی و همکاران، 2003). گام دوم، استفاده از مدل‌های پیش‌بینی کننده مختلف برای پیش‌بینی خصوصیات یا کلاس خاک است. این خصوصیات با استفاده از مدل‌های آماری مبتنی بر کامپیوتر بین اندازه‌گیری‌های نقطه‌ای خاک (اطلاعات حاصل از نیمرخ-های خاک) و متغیرهای کمکی، به دست می‌آید. گام سوم، شامل استفاده از خصوصیات پیش بینی شده خاک برای برآورد خصوصیات دیر یافت خاک، مانند اندازه‌گیری مقدار آب خاک در دسترس، تراکم کربن و تثبیت فسفرمی‌شود (سنچز و همکاران، 2009).

مدل درختی طبقه‌بندی و رگرسیون، یک روش ناپارامتری الگوریتمی است که قادر به پیش بینی متغیرهای کمی یا متغیرهای طبقه‌بندی‌شده بر اساس مجموعه‌ای از متغیرهای پیش‌بینی کننده کمی و کیفی است. در این روش، مجموعه‌ای از شرطهای منطقی به صورت یک الگوریتم با ساختار درختی برای طبقه بندی یا پیش‌بینی کمی یک متغیر به کار می‌رود؛ که اصولاً در این مدل از دسته بالای داده‌های برازش داده شده استفاده می‌شود که معمولاً دارای خطا است و داده‌هایی که حاصل تجزیه و تحلیل مدل هستند با خطا مواجه هستند. برای از بین بردن این خطا از الگوریتم جنگل های تصادفی[2] استفاده می‌شود به طوری که این الگوریتم مدل را قادر می‌سازد تا برای پیش بینی متغیرهای خروجی از داده‌های برازش شده به طور تصادفی استفاده کند که این کار باعث می‌شود دقت پیش بینی مدل تا حد بسیار زیادی افزایش یابد.

جنگل‌های تصادفی در حقیقت مجموع‌هایی از درخت‌های پیش‌بینی کننده با احتمال یکسان و دارای پراکندگی یکسانی هستند. اساس این روش بر انتخاب خطای داده‌های بزرگتر به‌عنوان خطای اصلی و ایجاد همبستگی بین خطاهای دیگر داده‌ها به ترتیب است. انتخاب خطای عمومی بر اساس بزرگترین داده‌ها، پیش‌بینی دقیق و قابل استنادی به ما می‌دهد؛ اما زمانی این پیش‌بینی دقیق‌تر هم می‌شود که داده‌های ورودی به مدل بدون اعمال تغییرات باشد (بریمان، 2001). بعضی از مزایای استفاده از درخت تصمیم که مور و همکاران (1991) در مقایسه با سایر روش‌های طبقه بندی بیان کردند عبارت‌اند از:

1- وقتی با دادهای کمی و کیفی مواجه هستیم، تفسیر کردن آن‌ها راحت‌تر است. 2- بسیار راحت‌تر با دادهای گمشده یا دادهای پرت برخورد می‌کند. 3- توانایی درک و تحلیل روابط غیر خطی خصوصیات خاک. 4- هیچ فرضیه‌ای برای توزیع داده‌ها نیاز ندارد. 5- بسیار راحت به روز می‌شود.

کامرا و همکاران (2017) با استفاده از دو مدل رگرسیون لجستیک چند متغیره و جنگل‌های تصادفی اقدام به تهیه نقشه رقومی کشور قبرس کردند که نتایج

---

1. Digital Soil Mapping

2. Random Forests

آن‌ها نشان داد که مدل جنگل‌های تصادفی نسبت به مدل رگرسیون لجستیک چند متغیره از توانایی بالاتری برای پیش بینی کلاس‌های خاک منطقه داشته و همچنین این پیش بینی با خطای کمتری نسبت به مدل دیگر همراه بوده است. سرینیواس و همکاران (2016) برای تهیه نقشه رقومی کربن آلی و غیر آلی در هندوستان اقدام به نمونه‌برداری 1198 نقطه به روش طبقه‌بندی تصادفی کردند و پس از انجام آزمایشات روی نمونه‌ها نتایج را به همراه برخی داده‌های کمکی از جمله شاخص پوشش گیاهی، داده‌های اقلیمی، پوشش زمین و ... در مدل جنگل‌های تصادفی برازش دادند. مقدار 25 درصد از کل نقاط نمونه‌برداری به‌عنوان داده‌های اعتبار سنجی مدل انتخاب شد. نتایج آن‌ها نشان داد که مقدار کربن آلی، کربن غیر آلی و مقدار مجموع ذخیره کربن خاک‌ها به ترتیب، 22/73، 12/83 و 35/55 پکاگرم می‌باشد. هیونگ و همکاران (2016) در مطالعه‌ای با عنوان مرور و مقایسه تکنیک‌های یادگیری ماشینی برای اهداف طبقه‌بندی در تهیه نقشه رقومی خاک، به این نکته اشاره کردند که مدل‌های جنگل‌های تصادفی و مدل‌های درختی به دلیل سرعت در پردازش پارامتری و سهولت تفسیر داده‌های خروجی، مدل‌های مناسبی جهت استفاده در اهداف طبقه‌بندی هستند.

پهلوان راد و همکاران (2016) داده‌های حاصل از نتایج آزمایشگاهی و مطالعات میدانی حدود 85000 هکتار از اراضی استان گلستان در ایران را به‌منظور تهیه نقشه رقومی خاک منطقه تحت دو مدل جنگل‌های تصادفی و رگرسیون لجستیک چند متغیره برازش دادند که نتایج آن‌ها نشان داد مدل جنگل‌های تصادفی به طور معنی‌داری نسبت به مدل رگرسیون لجستیک چند متغیره توانایی پیش بینی سری‌های خاک منطقه را داشته است. هنگل و همکاران (2015) در بخشی از آفریقا اقدام به تهیه نقشه رقومی خصوصیات خاک با دو مدل رگرسیون چندگانه خطی و جنگل های تصادفی کردند که نتایج اعتبار سنجی مدل‌ها نشان داد مدل جنگل‌های تصادفی توانسته به طور معنی داری دقت بیشتری نسب به مدل رگرسیون خطی چندگانه از خود نشان دهد. هیونگ و همکاران (2014) اقدام به تهیه نقشه مواد مادری خاک با استفاده از متغیرهای کمکی و مدل جنگل‌های تصادفی کردند که نتایج آن‌ها نشان داد زمانی که از متغیرهای کمکی مستخرج از تصاویر ماهواره‌ای و متغیرهای کمکی مدل رقومی ارتفاع به‌عنوان داده‌های ورودی به مدل استفاده شد، ضریب کاپای[1] اعتبار سنجی مدل 89/6 درصد استخراج گردید.

با توجه به اهمیت استراتژیک نقشه‌های پایه اراضی برای هر منطقه، هدف از مطالعه حاضر تهیه نقشه رقومی خاک با استفاده از یک مدل تصمیم‌گیری با دقت مطلوب، برای به دست آوردن اطلاعات خاک منطقه جهت بررسی و اجرای برنامه‌های مدیریتی کشاورزی پایدار و توسعه پایدار اراضی می باشد.

مواد و روش‌ها

موقعیت منطقه مورد مطالعه

منطقه مورد مطالعه با مساحت 7261 هکتار در شرق استان قزوین در محدوده شهرستان آبیک و بین 36 درجه و 2 دقیقه تا 36 درجه و 5 دقیقه عرض شمالی و 50 درجه و 24 دقیقه تا 50 درجه و 29 دقیقه طول شرقی واقع شده است. میانگین بارندگی سالیانه این منطقه 278 میلیمتر و رژیم‌های رطوبتی و حرارتی خاک منطقه به ترتیب زریک و ترمیک طبقه‌بندی شدند. منطقه از لحاظ پستی و بلندی بیشتر مسطح بوده، بلندترین ارتفاع منطقه 1459 متر و کمترین ارتفاع 1175 است.

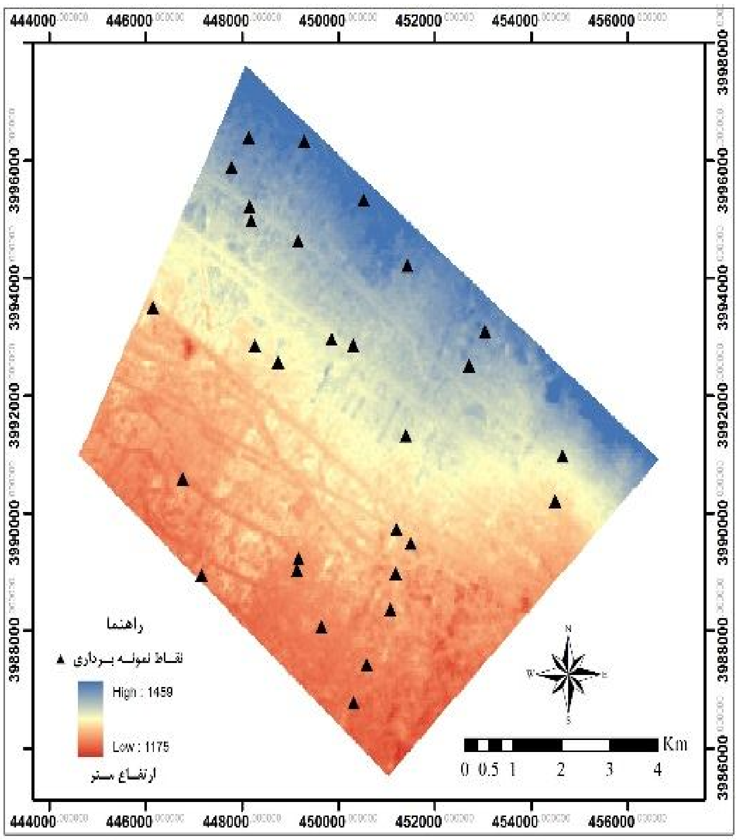
الگوی نمونه‌برداری

برای تعیین مکان حفر پروفیل‌ها از روش نمونه‌برداری مکعب لاتین مشروط[2] (میناسنی و مک براتنی، 2006)، استفاده شد. روش نمونه‌برداری مکعب لاتین مشروط در سال 2006 توسط میناسنی و مک براتنی به‌منظور استفاده در تعیین نقاط نمونه‌برداری ارائه شد. روش ارائه‌شده بر پایه روش نمونه‌برداری مربع لاتین است که در سال 1979 مک کی پیشنهاد داده شد (مک کی، 1979) اما با این تفاوت که در روش مربع لاتین شرطی تمامی نقاطی که برای نمونه‌برداری مشخص می‌شود در جهان واقعی وجود دارند. این روش با استفاده از تصاویر ماهواره لندست 8 و مدل رقومی ارتفاع مستخرج از سنجنده استر، استخراج متغیرهای کمکی با استفاده از نرم‌افزارهای آرک جی ای اس نسخه 9/3[3]، ارداس ایمیجن نسخه 2014[4] و ساگا نسخه 2/2/2[5] انجام شد. این متغیرهای کمکی شامل اطلاعات حاصل از باندهای طیفی تصویر ماهواره لندست، شاخص پوشش گیاهی[6]، شاخص پوشش گیاهی اصلاح شده خاک[7]، درجه

---

1. Kappa Index
2. Conditioned Latin Hypercube Sampling (cLHS)
3. ARC GIS 9.3
4. ERDAS IMAGINE 2014
5. SAGA 2.2.2
6. Normalized difference vegetation index
7. Soil-adjusted Vegetation Index

شیب، جهت شیب، شاخص خیسی، فاکتور طول شیب و شاخص قدرت جریان می‌باشند. پس از استخراج متغیرهای کمکی و وارد کردن آن‌ها به‌عنوان فایل رستری به محیط برنامه آرک جی ای اس، در این نرم‌افزار هرکدام از متغیرها به‌صورت نقطه‌ای به نمایش درآمدند به‌طوری که هر پیکسل به صورت جداگانه دارای طول و عرض جغرافیایی و ارزش شد. برای به دست آوردن اهمیت هرکدام از متغیرها از طریق نرم‌افزار SAS تحلیل مؤلفه‌های اصلی اجرا شد و 11 متغیر به‌عنوان مهم‌ترین متغیرهای کمکی انتخاب شدند. در نهایت متغیرهای کمکی انتخاب شده وارد مدل مکعب لاتین مشروط شد و داده‌ها به n مقدار با احتمال برابر تقسیم شد، پس از اعمال شرایط مربع لاتین (یک نقطه در هر ستون و هر سطر)، لایه‌ها با هم ترکیب می‌شوند و الگوریتم شروع به تکرار فرایند می‌کند تا زمانی که تابع تکاملی هدف به پایداری برسد. در نهایت یک ترکیب از مجموعه متغیرهایی که شرایط مربع لاتین را داشته باشند و در داده‌های دیجیتالی متغیرهای محیطی هم به عنوان یک نقطه باشند به دست می‌آیند. طبق روش مکعب لاتین مشروط 29 نقطه جهت حفر پروفیل در منطقه مشخص شد (شکل 1).



شکل 1- نقاط تعیین شده برای نمونه برداری

تجزیه‌های آزمایشگاهی

با استفاده از دستگاه موقعیت یاب جهانی[1] و نقاط مشخص شده توسط روش مکعب لاتین مشروط ، اقدام به حفر شد، سپس پروفیل‌ها با استفاده از برگه تشریح پروفیل مورد مطالعه قرار گرفتند و خصوصیات خاک نظیر رنگ، بافت، ساختمان، افق‌های ژنتیکی و مشخصه در برگه‌های تشریح پروفیل ثبت گردید. با استفاده از این اطلاعات خاک‌ها در صحرا بر اساس سیستم طبقه‌بندی آمریکایی 2014طبقه‌بندی اولیه شدند و از هرکدام از افق‌های ژنتیکی نمونه‌برداری انجام گرفت و نهایتاً 97 نمونه خاک جمع آوری شد. پس از هوا خشک کردن نمونه‌های خاک، جرم مخصوص ظاهری، بافت خاک به روش هیدرومتری (گی و بادر، 1986)، واکنش خاک، قابلیت هدایت الکتریکی، کربن آلی (والکلی و

1. Global Positioning System

بلک، 1934)، آهک، ظرفیت تبادل کاتیونی، فسفر قابل جذب و پتاسیم تبادلی بر اساس روش های استاندارد (اسپارک و همکاران، 1996) اندازه‌گیری شد.

متغیرهای کمکی

در تحقیق حاضر از مدل رقومی ارتفاع مستخرج از سنجنده استر[1] استفاده شده است. پس از آماده کردن مدل رقومی ارتفاع، پارامترهای، شیب، جهت شیب، ارتفاع، شاخص خیسی، شاخص قدرت جریان و شاخص طول شیب در محیط نرم افزار ساگا محاسبه و استخراج گردید (هنگل و روزیتر، 2003). استخراج این پارامترها با استفاده از ابزارهای طراحی شده در نرم افزار ساگا و مدل رقومی ارتفاع منطقه انجام شد. در این مطالعه از تصاویر ماهواره لندست استفاده شد. این تصاویر از 11 باند تشکیل شده و قدرت تفکیک مکانی آن 15 متر است. داده‌های کمکی مستخرج از این تصاویر شامل شاخص پوشش گیاهی، شاخص پوشش گیاهی اصلاح شده خاک، تهیه نقشه هر کدام از باندها به صورت جداگانه (تقی زاده مهرجردی و همکاران، 2014) و نسبت باندهای 2-4، 7-4، 6-5، 7-5، 2-6 و 3-6 می‌باشد (استوم و همکاران، 2010).

مدل‌سازی خاک

پس از پایان آنالیزهای آزمایشگاهی با استفاده از داده‌های به دست آمده و راهنمای تشریح صحرایی (2012) اقدام به نام‌گذاری خاک‌ها تا سطح فامیل بر اساس کلید رده‌بندی 2014[2] گردید و پس از مشخص شدن نام خاک‌ها آن‌ها کدگذاری شدند و متغیرهای کمکی برای هر کدام از پروفیل‌ها از طریق نرم افزار آرک جی ای اس استخراج شد و به‌عنوان متغیر وابسته در نظر گرفته شدند. همچنین تمام لایه‌های کمکی به فرمت رستری با اندازه پیکسلی 30 متر تهیه شدند، این اطلاعات به‌عنوان متغیرهای وابسته در مدل اسکورپن شناخته می‌شوند. تمامی اطلاعات با فرمت نهاییxlxs. تبدیل شده و برای فرایند مدل‌سازی در نرم‌افزارهای مربوطه آماده شدند. پس از آماده‌سازی داده‌ها با استفاده از نرم افزار متلب 2015[3] اقدام به اجرای مدل گردید که در این تحقیق از مدل جنگل‌های تصادفی با اجرای 500 درخت پیش بینی استفاده شد.

نتایج و بحث

الگوی نمونه‌برداری

در این تحقیق از روش مکعب لاتین مشروط برای انتخاب نقاط نمونه‌برداری استفاده گردید. پس از انجام آنالیز مولفه‌های اصلی ورودی‌های مدل شامل، ارتفاع، شیب، جهت شیب، شاخص خیسی، فاکتور طول شیب، شاخص قدرت جریان، باند 4، باند 5، باند 9، شاخص پوشش گیاهی و شاخص پوشش گیاهی اصلاح شده خاک وارد مدل شدند. نتایج آنالیز مولفه‌های اصلی در جدول 1 نشان داده شده است.

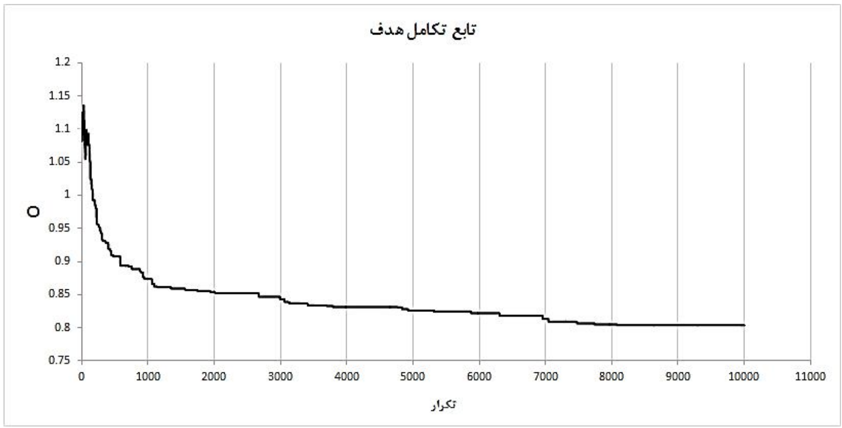
پس از اجرای مدل و تعیین نقاط نمونه‌برداری توسط مدل، تابع تکامل هدف با ده هزار تکرار (شکل2) ارائه شد؛ که مقدار این تابع با افزایش تکرار کاهش یافت به طوری که در شرایط تکرارهای کمتر از 1000 تابع از آشفتگی بیشتری برخوردار است اما با افزایش تعداد تکرارها پس از تکرارهای 9000 بار تابع تکامل به پایداری می‌رسد. مک براتنی و میناسنی، 2006 نیز با اجرای مدل مکعب لاتین مشروط در بخشی از استرالیا نشان دادند که در تکرارهای بالای 10000 بار، تابع تکامل هدف به پایداری قابل قبولی می رسد. تقی زاده مهرجردی و همکاران نیز در سال 1393 در تحقیقی در منطقه اردکان یزد جهت تعیین نقاط نمونه برداری از روش نمونه برداری مکعب لاتین مشروط استفاده کردند که آنها نشان دادند مقدار تابع تکامل هدف با افزایش تکرار کاهش می‌یابد.

---

1. Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER)
2. Key to soil taxonomy 2014
3. MATLAB 2015

جدول 1- نتایج آزمون تجزیه مؤلفه‌های اصلی

	PC1	PC2	PC3	PC4	PC5	PC6
مدل رقومی ارتفاع	0/207786	-0/228812	0/208337	0/172672	0/012328	0/011436
جهت شیب	0/105112	-0/2228	0/460606	0/365844	-0/034253	0/008632
شاخص پوشش گیاهی	-0/277343	0/248914	0/228519	0/055386	0/079729	0/002409
شاخص خیسی	-0/0079	0/155385	-0/343465	0/604556	-0/003557	-0/016944
فاکتور طول شیب	0/003643	-0/004159	0/011631	-0/043702	0/007452	0/996507
شاخص شدت جریان	0/032949	-0/186343	0/383148	-0/556726	-0/001346	-0/078692
شاخص پوشش گیاهی اصلاح شده خاک	-0/278003	0/24629	0/228138	0/055793	0/059022	0/002614
شیب	0/112566	-0/18306	0/485866	0/367488	-0/045615	0/009764
باند1	0/296416	0/156355	-0/036719	-0/059063	-0/026697	-0/002801
باند2	0/302806	0/154783	-0/036153	-0/058129	-0/051901	-0/002637
باند3	0/303907	0/194458	-0/012535	-0/055593	-0/079072	-0/002268
باند4	0/307585	0/139835	-0/026006	-0/04488	-0/083647	-0/001425
باند5	0/033519	0/627326	0/340524	0/027907	-0/010261	-0/003584
باند6	0/289028	0/17622	0/087356	0/014545	-0/016072	-0/00246
باند7	0/303591	0/110894	0/018273	-0/008052	-0/031169	-0/004428
باند8	0/295746	0/166487	-0/022581	-0/051139	-0/085925	-0/001814
باند9	0/04614	0/125395	0/033052	-0/009972	0/932695	-0/010037
باند10	0/268711	-0/221484	-0/073059	0/031427	0/202938	0/00815
باند11	0/269422	-0/221161	-0/069633	0/032303	0/224069	0/002878



شکل 2- تابع تکامل هدف

نقشه رقومی خاک

جهت تهیه نقشه رقومی خاک ابتدا با استفاده از آنالیزهای آزمایشگاهی و برگه‌های تشریح پروفیل، خاک‌های منطقه بر اساس کلید رده‌بندی تا سطح فامیل رده‌بندی شدند و در مجموع 17 فامیل خاک در 5 زیر گروه، 4 گروه بزرگ، 3 تحت رده و 2 رده شناسایی شد اسامی خاک‌های شناسایی شده در جدول 2 ارائه شده است. پس از مشخص شدن کلاس خاک‌ها، داده‌های کمکی هر خاکرخ با استفاده از نرم افزار Arc Gis 9.3 استخراج گردید. داده‌های کمکی برای تمام پیکسل‌های منطقه نیز به صورت فایل اکسل از نرم افزار Arc Gis 9.3 استخراج شد و برای استفاده در مدل‌سازی آماده شد. با استفاده از نرم افزار متلب و مدل جنگل‌های تصادفی پیش بینی کلاس‌های فامیل منطقه انجام شد که در نهایت نقشه

به دست آمده از منطقه به صورتی که در شکل 3 آمده، به دست آمد.

نتایج پیش‌بینی کلاس‌های خاک منطقه نشان داد که از مجموع 7261 هکتار اراضی، بیشترین وسعت مربوط به واحد خاک 15 با 1521 هکتار مساحت که 21 درصد کل منطقه را شامل می‌شود و بعد از آن واحد خاک 1 با 1387 هکتار مساحت که 19درصد کل منطقه را شامل می‌شود و کمترین وسعت نیز مربوط به واحد خاک 5 با حدود 3 هکتار مساحت که 0/04 درصد کل منطقه را شامل می‌شود، است. همچنین برای اعتبار سنجی مدل جنگل‌های تصادفی از 25 درصد داده‌های به عنوان داده‌های اعتبار سنجی استفاده شد که پس از اجرای مدل با استفاده از نرم افزار متلب محاسبه شد.

ضریب کاپا و دقت کلی بترتیب 0/83 و 87/5 درصد به دست آمدند. ضریب کاپای به دست آمده در رده تقریباً عالی می‌باشد (لاندیس و کوچ، 1977) که بر پایه آن می‌توان گفت مدل قادر به یادگیری داده‌های آموزشی بوده و پیش بینی مطلوبی را از کلاس‌های خاک انجام داده است. این نتایج با نتایج سایر محققان همچون کامرا و همکاران (2017)، تقی زاده مهرجردی و همکاران (2014)، پهلوان راد و همکاران (2014) و هیونگ و همکاران (2014)، به جهت اینکه مدل جنگل‌های تصادفی قادر به پیش‌بینی مطلوب کلاس‌ها خاک و خصوصیات خاک با درصد خطای پایین است، مطابقت داشت. باتوجه به اینکه تحقیق حاضر با استفاده از تعداد پروفیل‌های مشاهداتی کمتر نسبت به سایر مطالعات، همچنین عدم استفاده از داده‌های پروفیلی قدیمی[1] (پهلوان راد و همکاران، 2016 و کامرا و همکاران، 2017 و ور و همکاران، 2015) انجام گردید؛ به نظر می‌رسد استفاده از روش نمونه‌برداری هدفمند (cLHS) و استخراج متغیرهای کمکی مناسب مانند نسبت‌های باندی ماهواره لندست، مدل جنگل‌های تصادفی را قادر ساخته تا در پیش بینی کلاس‌های خاک به نحو مطلوبی عمل نماید.

با بررسی متغیرهای کمکی ورودی به مدل نتایج به دست آمده نشان داد متغیرهایی که بیشترین تأثیر را در تعیین کلاس‌های خاک دارا بودند به ترتیب شامل: باند 9 تصاویر ماهواره، شاخص‌های پوشش گیاهی و پوشش گیاهی اصلاح شده خاک و نسبت‌های باندی 5/6 و 5/7 می‌باشند و متغیرهایی که کمترین تأثیر را در تعیین کلاس‌های خاک داشته‌اند به ترتیب شامل شاخص خیسی، نقشه شیب، شاخص قدرت جریان بودند در شکل 4 نمودار سه‌بعدی مربوط به تأثیر متغیرهای کمکی در تعیین کلاس‌های خاک آمده است.

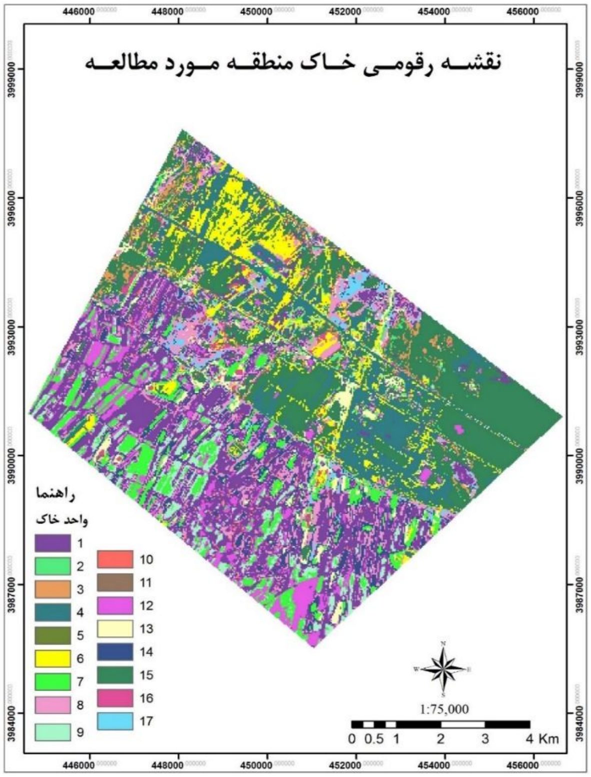
احتمالاً با توجه به نتایج در این منطقه ویژگی-های توپوگرافیکی و شاخص‌های مربوط به آن یکنواخت یکنواخت هستند بنابراین می‌توان بیان داشت که متغیرهای کمکی که مربوط به توپوگرافی باشند نقش مهمی را در تعیین کلاس‌های خاک منطقه ایفا نمی‌کنند و در مقابل متغیرهای کمکی استخراج شده از تصاویر ماهواره توانسته‌اند به خوبی از طریق طول موج‌های منعکس شده متفاوت، در منطقه ایجاد تفاوت کنند و بنابراین نقش مهمی در تعیین کلاس‌های خاک منطقه داشته باشند، که نتایج مشابهی توسط برخی پژوهشگران نیز گزارش شده است (پهلوان راد و همکاران، 2014).

---

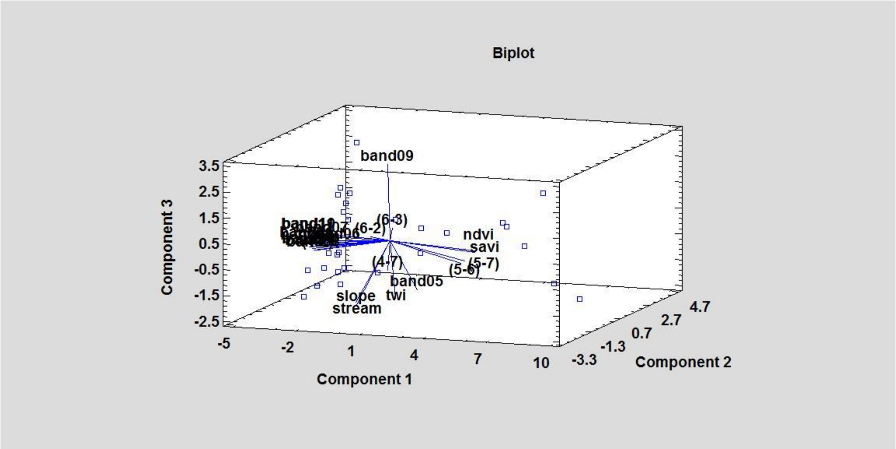
1. Legacy Data

جدول 2- رده بندی خاک های منطقه تا سطح فامیل

واحد خاک	فامیل خاک
1	Coarse-loamy over fragmental, mixed, superactive, calcareous, thermic Typic xerorthents
2	Coarse-loamy over fragmental, mixed, superactive, thermic Typic Calcixerepts
3	Coarse-loamy over loamy skeletal, mixed, superactive, thermic Typic Calcixerepts
4	Coarse-loamy over sandy, mixed, superactive, calcareous, thermic Typic Xerofluvents
5	Coarse-loamy, mixed, superactive, thermic Fluventic Haploxerepts
6	Coarse-loamy, mixed, superactive, thermic Typic Calcixerepts
7	Coarse-loamy, mixed, superactive, thermic Typic Haploxerepts
8	Fine-loamy over fragmental, mixed, superactive, thermic Typic Haploxerepts
9	Fine-loamy over sandy skeletal, mixed, superactive, thermic Typic Haploxerepts
10	Fine-loamy, mixed, superactive, thermic Fluventic Haploxerepts
11	Fine-loamy, mixed, superactive, thermic Typic Calcixerepts
12	Fine-loamy, mixed, superactive, thermic Typic Haploxerepts
13	Loamy-skeletal over fragmental, mixed, superactive, thermic Typic Calcixerepts
14	Loamy-skeletal, mixed, superactive, calcareous, thermic Typic Xerofluvents
15	Loamy-skeletal, mixed, superactive, thermic Typic Calcixerepts
16	Sandy skeletal, mixed, superactive, thermic Typic Calcixerepts
17	Sandy- skeletal, mixed, superactive, calcareous, thermic Typic Xerorthents



شکل 3- پیش‌بینی کلاس‌های خاک منطقه توسط مدل جنگل‌های تصادفی در سطح فامیل خاک



شکل 4- نمودار سه بعدی تأثیر متغیرهای کمکی در تعیین کلاس‌های خاک

نتیجه‌گیری

امروزه نقشه برداری رقومی خاک به طور گسترده برای تولید نقشه های خاک و پردازش داده‌های رقومی در سراسر جهان مورد استفاده قرار می‌گیرد. با توجه به اینکه مدل جنگل‌های تصادفی به عنوان یکی از مناسب ترین روش‌های تولید نقشه های رقومی خاک شناخته شده است؛ در این مطالعه نیز این روش قادر به پیش بینی مطلوب کلاس‌های خاک گردید. به نظر می‌رسد علی رقم استفاده از تعداد پروفیل‌های مشاهداتی کمتر نسبت به سایر مطالعات، بهره‌گیری از یک روش نمونه-برداری هدفمند (cLHS) و متغیرهای کمکی مناسب به جای استفاده از داده‌های پیشین توانسته‌اند صحت پیش بینی مدل جنگل‌های تصادفی را بهبود بخشند؛ بنابراین استفاده از این مدل منجر به تهیه رقومی خاک با ضریب کاپا 0/83 و دقت کلی 87/5 درصد گردید.

نتایج مدلسازی خاک نشان داد که منطقه دارای 17 نوع خاک مختلف در سطح فامیل می‌باشد. براساس خروجی نقشه رقومی خاک، منطقه مورد مطالعه از دو نوع خاک انتی سول و اینسپتی سول ها تشکیل شده‌اند به طوری که در شمال و شمال غربی منطقه که از نظر ارتفاع بالاتر از قسمت‌های جنوبی است، فامیل خاک Loamy-skeletal, mixed, superactive, thermic Typic Calcixerepts نوع خاک قالب منطقه است. در قسمت جنوب و جنوب غربی منطقه نیز خاک Coarse-loamy over fragmental, mixed, calcareous, superactive, thermic Typic Xerorthents خاک غالب بود. باتوجه به ارائه پراکنش مکانی دقیق از احتمال رخداد کلاس‌های خاک توسط این مدل، می‌توان بهره وری استفاده از اراضی بهبود بخشید.

فهرست منابع:

1. Breiman, L. 2001. "Random forests" Machine learning. 45:5-32.
2. Camera, C., Z. Zomeni., J.S. Noller., A.M. Zissimos., I.C. Christoforou., and A. Bruggeman. 2017. A high resolution map of soil types and physical properties for Cyprus: A digital soil mapping optimization. Geoderma 285:35-49.
3. Gee, G.W., and J.W. Bauder. 1986. Particle size analisis. p. 383-411. In Methods of Soil Analysis. Part l. Soil Science Society of America. Madison, WI.
4. Hengl, T., D.G. Rossiter., and A. Stein. 2003. Soil sampling strategies for spatial prediction by correlation with auxiliary maps. Soil Research. 41:1403-1422.
5. Hengl, T., G.B. Heuvelink., B. Kempen., J.G. Leenaars., M.G. Walsh., K.D. Shepherd., K. D., and J.E. Tondoh. 2015. Mapping soil properties of Africa at 250 m resolution: Random forests significantly improve current predictions. PloS one. 10:125-814.

6. Heung, B., C.E. Bulmer., and M.G. Schmidt. 2014. Predictive soil parent material mapping at a regional-scale: a random forest approach. Geoderma. 214:141-154.
7. Heung, B., H.C. Ho., J. Zhang., A. Knudby., C.E. Bulmer., and M.G. Schmidt. 2016. An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. Geoderma. 265:62-77.
8. Jenny, H. 1941. Factors of soil formation: A system of quantitative pedology. Courier Corporation.
9. Lagacherie, P., A.B. McBratney., M. Voltz. 2007. Digital soil mapping: an introductory perspective. Elsevier.
10. Landis, J. R., and G.G. Koch. 1977. The measurement of observer agreement for categorical data. Biometrics. 159-174.
11. McBratney, A.B., M.L. Mendonça Santos., and B. Minasny. 2003. On digital soil mapping. Geoderma. 117:3-52.
12. McKay, M.D., R.J. Beckman., and W.J. Conover. 1979. Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics. 21:239-245.
13. Minasny, B., and A.B. McBratney., 2006. A conditioned Latin hypercube method for sampling in the presence of ancillary information. Computers & Geosciences. 32:1378-1388.
14. Moore, I. D., R.B. Grayson., and A.R. Ladson. 1991. Digital terrain modelling: a review of hydrological, geomorphological, and biological applications. Hydrological processes. 5:3-30.
15. Pahlavan-Rad, M. R., F. Khormali., N. Toomanian., C.W. Brungard., F. Kiani., C.B. Komaki., and P. Bogaert. 2016. Legacy soil maps as a covariate in digital soil mapping: A case study from Northern Iran. Geoderma. 279:141-148.
16. Pahlavan-Rad, M. R., N. Toomanian., F. Khormali., C.W. Brungard., C.B. Komaki., and P. Bogaert. 2014. Updating soil survey maps using random forest and conditioned Latin hypercube sampling in the loess derived soils of northern Iran. Geoderma. 232:97-106.
17. Palm, C., P. Sanchez., S. Ahamed., and A. Awiti. 2007. Soils: A contemporary perspective. Annu. Rev. Environ. Resour. 32:99-129.
18. Sanchez, P. A., S. Ahamed., F. Carré., A.E. Hartemink., J. Hempel., J. Huising., ... and B. Minasny. 2009. Digital soil map of the world. Science. 325:680-681.
19. Sparks, D. L., and J.M. Bartels. 1996. Chemical methods (No. 5-3). Madison: Soil Science Society of America.
20. Sreenivas, K., V.K. Dadhwal., S. Kumar., G.S. Harsha., T. Mitran., G. Sujatha., ... and T. Ravisankar. 2016. Digital mapping of soil organic and inorganic carbon status in India. Geoderma. 269:160-173.
21. Stum, A. K., J.L. Boettinger., M.A. White., and R.D. Ramsey. 2010. Random forests applied as a soil spatial predictive model in arid Utah. In Digital Soil Mapping. Springer, Netherlands.
22. Taghizadeh-Mehrjardi, R., B. Minasny., F. Sarmadian., and B.P. Malone. B. 2014. Digital mapping of soil salinity in Ardakan region, central Iran. Geoderma. 213:15-28.
23. Walkley, A., and I.A. Black. 1934. An examination of the Degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. Soil science. 37:29-38.
24. Were, K., D.T. Bui., Q.B. Dick., and B.R. Singh. 2015. A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape. Ecological Indicators. 52:394-403.

Digital Soil Mapping Using Random Forests Model in Abyek, Qazvin Province

S. E. Khamoshi, F. Sarmadian[1], and A. Keshavarzi
Ph.D Student, University of Tehran; E-mail: khamoshierfan@ut.ac.ir
Professor, Dept. of Soil Science, Faculty of Agricultural Engineering and Technology, University of Tehran; E-mail: fsarmad@ut.ac.ir
Assistant Professor, Dept. of Soil Science, Faculty of Agricultural Engineering and Technology, University of Tehran; E-mail: alikeshavarzi@ut.ac.ir
Received: April, 2018 and Accepted: August, 2018

Abstract

Today, there is great demand for accurate soil information to determine the relationship between soil and landscape, and easy updating of soil maps has increasingly gained importance. Digital soil mapping techniques can be cost-effective solutions to obtain information dealing with the soil types over large areas. The objective of this study was to provide a digital soil map in the Abyek region of Qazvin province using random forest model for management purposes and sustainable land use planning. To this end, soil samples were collected based on cLHS. After the laboratory analysis, using random forest model and auxiliary variables derived from a digital elevation model with a spatial resolution of 30 m and Landsat 8 imagery, the modeling and preparation of regional soil maps were performed. Out of 7261 ha, the dominant soil was classified as Loamy-skeletal, mixed, superactive, thermic Typic Calcixerepts. The results showed that soil modeling using random forest algorithm could accurately predict (Kappa coefficient ~ 0.83) soil classes in the region.

Keywords: Spatial prediction, Modeling, Conditioned Latin Hypercube Sampling, Auxiliary Variable, Soil Family

[1]. Corresponding author: Soil Science Department, University of Tehran, Iran