

مقایسه گروه‌بندی و کیفیت مرزبندی‌های موروثی خاک با مدل‌های داده‌کاوی کمی: مطالعه موردی بخش‌هایی از استان چهارمحال و بختیاری

زهرا رسائی^۱، جهانگرد محمدی^۲، اعظم جعفری^۳

(تاریخ دریافت: ۱۳۹۸/۰۷/۰۷ تاریخ پذیرش: ۱۳۹۸/۱۰/۲۳)

چکیده

کشف ارتباط بین خاک‌ها و گروه‌بندی آن‌ها بر اساس فاکتورهای مختلف، دربرگیرنده اهمیت بسزایی در زمینه‌های مختلف از جمله مدیریت اراضی و کشاورزی پایدار می‌باشد. این امر با ترسیم مدل‌های ذهنی و بر اساس فاکتورهای محیطی در قالب نقشه‌های سنتی خاک آغاز، و با استفاده از مفهوم فاصله و شباهت به کمک مدل‌های کمی و ریاضی، ادامه یافته است. تحقیق حاضر به منظور مقایسه گروه‌بندی‌های در دسترس نقشه سنتی خاک با مدل‌های کمی کلاسیک و مدرن صورت می‌گیرد. به این منظور، داده‌های موروثی خاک منطقه شهرکرد-بروجن با کمک الگوریتم‌های مختلف از جمله خوشه‌بندی سلسله مراتبی، میانگین‌های کا، درخت طبقه‌بندی و فاصله تاکسونومیکی گروه‌بندی شدند و نتایج با کلاس‌های خاک نقشه موروثی مورد مقایسه قرار گرفتند. نتایج بررسی‌ها در رزلوشن مکانی ۹۰ متری نشان داد که گروه‌بندی‌های بدست آمده از درخت طبقه‌بندی، با کلاس‌های نقشه موروثی، بیشترین همخوانی را دارند. همچنین، الگوریتم‌های خوشه‌بندی سلسله مراتبی و میانگین‌های کا معمولی نیز به ترکیب گروهی مشابه با الگوی فیزیوگرافیک سنتی از نظر ویژگی‌های محیطی و مورفولوژیکی منجر شدند. آنالیز فاصله تاکسونومیکی با در نظر گرفتن همبستگی بین گروه‌ها و همچنین ویژگی‌های مورد بررسی، به بهترین ترکیب کلاس‌ها از نظر ویژگی‌های مختلف آن‌ها و همچنین بالاترین همبستگی درون کلاسی (۰/۵۲۲) و کمترین نسبت تغییر واریانس گروهی (۰/۹۱۵) منجر گردید. مقادیر کم آماره آنالیز تجزیه واریانس چندگانه (بین ۰/۰۰۱ در مدل سنتی تا ۰/۰۱۴ در مدل درختی)، نشان داد مدل‌های مورد بررسی بجز میانگین‌های کا، خاک‌ها را بطور مؤثری از یکدیگر تفکیک کرده‌اند. بطور کلی، استفاده از مدل‌های طبقه‌بندی عددی می‌تواند به نمایان کردن روابط کمی بین خاک‌ها منجر شود. خاک‌شناس می‌تواند در نهایت ترکیب این گروه‌بندی‌ها را با استفاده از تجربه و آگاهی خویش از منطقه مورد بررسی در راستای دستیابی به گروه‌های یکنواخت‌تر از نظر ویژگی‌های مدیریتی و تاکسونومی آن‌ها تعدیل یا تصحیح کند.

واژه‌های کلیدی: فاصله تاکسونومی، طبقه‌بندی عددی، نقشه سنتی خاک، مرزبندی واحد نقشه، واریانس چندگانه

رسائی ز.، محمدی ج.، جعفری ا. ۱۳۹۹. مقایسه گروه‌بندی و کیفیت مرزبندی‌های موروثی خاک با مدل‌های داده‌کاوی کمی: مطالعه موردی بخش‌هایی از استان چهارمحال و بختیاری. تحقیقات کاربردی خاک. جلد ۸، شماره ۴. صفحه: ۲۸-۴۳.

۱- دانشجوی دکتری گروه علوم خاک، دانشکده کشاورزی دانشگاه شهرکرد (مکاتبه کننده)

۲- استاد گروه علوم خاک، دانشکده کشاورزی دانشگاه شهرکرد

۳- استادیار گروه علوم خاک، دانشکده کشاورزی دانشگاه شهید باهنر کرمان

*پست الکترونیک: zahra.rasaei@gmail.com

مقدمه

خاک به عنوان یکی از هسته‌های اصلی دستیابی به اهداف توسعه پایدار، نقش مهمی در مدیریت مسائل زیست محیطی دارد. بنابراین، تهیه الگوی دقیق پراکنش ویژگی‌ها و کلاس‌های خاک که به خوبی جنبه‌های مختلف مدیریتی خاک را در بر بگیرد، می‌تواند در این زمینه راهگشا باشد (Bouma *et al.*, 2019). یافتن ارتباط بین خاک‌ها و گروه‌بندی آن‌ها همواره موضوع جالب توجه مسئولین برای شناسایی و مدیریت آن‌ها بوده است. همچنین، آگاهی از تنوع و تغییر خاک‌ها و الگوی کلاس بندی آن‌ها می‌تواند مدیریت بهتر کلاس‌های خاک با ویژگی‌های مشترک و مشابه را راحت‌تر و مؤثرتر سازد (Van Huyssteen *et al.*, 2013; Edokpayi *et al.*, 2017). این امر خاک‌شناسان را بر آن داشته است تا در پی یافتن راه‌های مناسب برای تفکیک و گروه‌بندی خاک‌ها بر اساس ویژگی‌های مختلف آن‌ها از جمله مشخصات آزمایشگاهی، مورفولوژیکی و ژنتیکی به علاوه موقعیت قرارگیری آن‌ها در زمین‌نما برآیند فرآیند کلاس‌بندی خاک‌ها شامل گروه‌بندی آن‌ها بر اساس میزان شباهت آن‌ها به یکدیگر می‌باشد. به این منظور، خاک‌های مشابه بر اساس ویژگی‌ها و فاکتورهای مختلف از جمله آزمایشگاهی، مورفولوژیکی، ژنتیکی و موقعیت قرارگیری در زمین‌نما، در گروه‌های مختلف قرار می‌گیرند تا بتوانند راحت‌تر مورد بررسی و مدیریت قرار بگیرند. بطورکلی، شباهت خاک‌ها به دو صورت عملکردی و تاکسونومیکی است که مشابهت عملکردی بیانگر شباهت فنوفرم^۱ و مشابهت تاکسونومیکی بیانگر شباهت ژنوفرم^۲ است (Bouma *et al.*, 2019). بدین منظور، تلاش‌ها با گروه‌بندی خاک‌ها به منظور دستیابی به خالص‌ترین کلاس‌ها از نظر ویژگی‌های مختلف خاکی و مدیریتی به کمک سیستم‌های طبقه‌بندی تاکسونومیکی و همچنین در قالب نقشه سری‌ها و واحدهای خاک آغاز شدند. در مطالعات سنتی خاک‌ها، گروه‌بندی‌ها در قالب سری‌ها و واحدهای خاک با استفاده از ویژگی‌های تاکسونومیکی و فاکتورهای زمین‌نما بر اساس نظر و قضاوت شخص خاک‌شناس و عمدتاً بدون در نظر گرفتن تغییرپذیری مکانی آن‌ها انجام می‌شود. این مطالعات بصورت کیفی

بوده و محور اصلی آن‌ها مدل‌های ذهنی و مفهومی خاک شناس می‌باشد که در نهایت، با در نظر گرفتن هر دو جنبه شباهت عملکردی و تاکسونومیکی، خاک‌ها در قالب گروه‌های مختلف در نقشه‌های خاک نمود می‌یابند. گروه بندی‌های بدست آمده از این روش که مدت‌ها به عنوان تنها راه نمایش ارتباط بین خاک‌ها مورد استفاده قرار می‌گرفت، به دلیل ماهیت کیفی و محوریت نظر کارشناس، کیفیت نامشخصی داشته و قابل ارزیابی کمی نمی‌باشند. گاهی اوقات ویژگی‌های خاک باهم هم‌پوشانی دارند و در نتیجه، در تمایز یک گروه خاک از گروه دیگر مشکل ایجاد کرده و منجر به نتیجه‌گیری نادرست می‌شود. بنابراین، این مدل‌ها و کلاس‌های بدست آمده از آن‌ها کاربرد محدودی در زمینه مدیریت خاک‌ها دارند (Edokpayi *et al.*, 2017). در این راستا، روش‌های پدومتریکی با بکارگیری مدل‌های ریاضی برای مطالعه توزیع و گروه بندی عددی خاک‌ها، مورد توجه قرار گرفته‌اند. خروجی آن‌ها قابل ارزیابی بوده و صحت و کارایی گروه‌بندی‌ها قابل بررسی می‌باشد. این مدل‌ها بر اساس اندازه‌گیری فاصله عددی بین خاک‌ها پایه‌گذاری شده و اقدام به تفکیک آن‌ها در گروه‌های با ویژگی‌های مشابه‌تر می‌کنند. بدین معنی که، میزان فاصله بین خاک‌ها با استفاده از برخی روش‌های کمی محاسبه شده، سپس خاک‌های مشابه که فاصله کمتری با یکدیگر دارند با کمک تکنیک‌های خوشه‌بندی در گروه‌ها توزیع می‌شوند (Carré & Jacobson, 2009). ساده‌ترین و معمول‌ترین شاخص فاصله، فاصله اقلیدسی است که اساس و پایه بسیاری از تکنیک‌های گروه‌بندی سنتی خاک‌ها است. بر اساس اندازه‌گیری فاصله و شباهت، الگوریتم‌های خوشه‌بندی نظارت نشده و نظارت شده مختلفی برای گروه‌بندی مشاهدات از جمله خاک ارائه شده است. تکنیک‌هایی مانند خوشه‌بندی سلسله مراتبی و میانگین‌های کا، مشاهدات را بر اساس فاصله بین آن‌ها گروه‌بندی کرده و نتایج نهایی ارائه می‌شود. این در حالی است که در مباحث طبقه‌بندی عددی نمونه‌ها و بویژه خاک‌ها، علاوه بر محاسبه فاصله و شباهت، آگاهی از قواعد حاکم بر شباهت بین آن‌ها بسیار جالب توجه می‌نماید. از این میان، می‌توان به الگوریتم‌های طبقه‌بندی مانند درختان تصمیم اشاره کرد. اخیراً، استفاده از مفهوم فاصله تاکسونومیکی

اصلی خاک‌های منطقه را تشکیل می‌دهند. ارتفاع در منطقه مطالعاتی از ۱۸۰۰ تا ۳۷۸۰ متر بالاتر از سطح دریا تغییر می‌کند. واحدهای مختلف فیزیوگرافی شامل کوه، تپه، فلات، آبرفت‌های بادبزی شکل سنگریزه‌دار (فن)، دشت‌های آبرفتی و اراضی پست در سطح منطقه وجود دارند (Mohammadi, 1986).

روش اجرا

جمع‌آوری و تهیه پایگاه داده

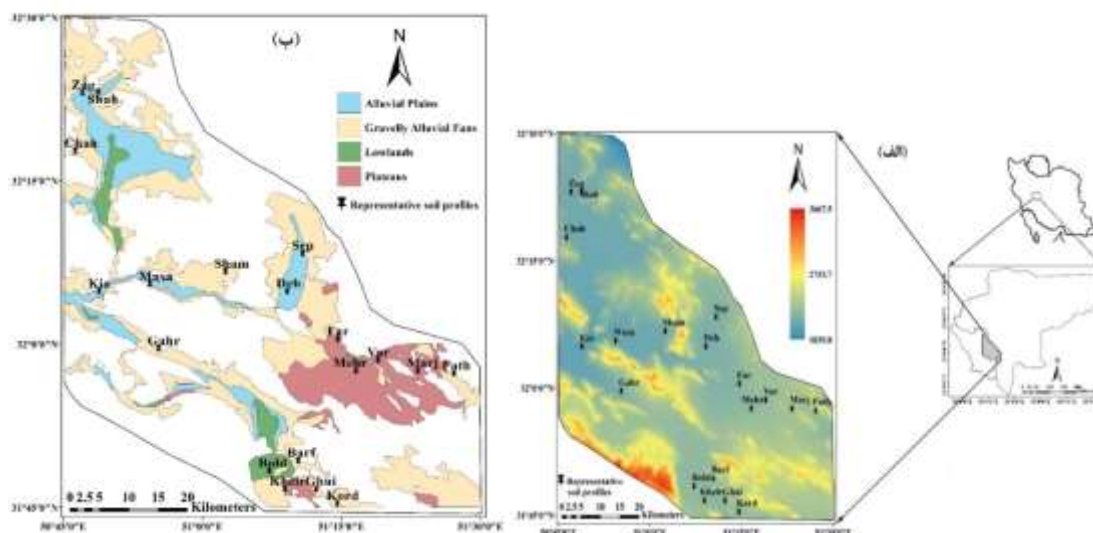
مطالعه موروثی نیمه‌تفصیلی خاک منطقه مطالعاتی در دسترس می‌باشد (Mohammadi, 1986). ویژگی‌های خاکی شامل درصد اجزای بافت خاک (سنگریزه، شن، سیلت، رس)، کربن آلی، کربنات کلسیم، رطوبت اشباع، اسیدیته، هدایت الکتریکی و عمق نیم‌رخ مربوط به نیم‌رخ‌های ۱۹ سری خاک مورد استفاده قرار گرفتند. شکل ۱ محل نقاط نیم‌رخ‌های شاهد سری‌های خاک در منطقه مطالعاتی را نشان می‌دهد. با توجه به نظر کارشناس و اطلاعات در دسترس از منطقه مورد مطالعه، مشتقات درجه اول و دوم از نقشه مدل رقومی ارتفاع (DEM) منطقه با رزولوشن مکانی ۹۰ متر در ۹۰ متر که روی پراکنش کلاس‌های خاک اثر دارند، استخراج شدند. این متغیرها شامل ارتفاع، شیب، جهت شیب، شاخص خیزی توپوگرافی، فاصله عمودی تا شبکه آبراهه، عمق دره، شاخص قدرت جریان، MRVBF، موقعیت شیب میانی و تحدب (Conrad et al., 2015) می‌باشند. همچنین، نقشه متغیرهای محیطی شامل شاخص پوشش گیاهی (Rouse Jr et al., 1974)، شاخص رس و کربنات کلسیم (Boettinger et al., 2008) و شاخص گج (Nield et al., 2007) از تصویر ماهواره‌ای لندست ۸ با رزولوشن مکانی ۳۰ متر در ۳۰ متر تهیه و با استفاده از روش نزدیک‌ترین همسایه (Resampling type: Nearest) در نرم‌افزار ArcMap 10.3.1 به رزولوشن مکانی ۹۰ متر در ۹۰ متر تغییر داده شدند. برای تمامی نقشه‌ها سیستم مختصات مرجع متریک (WGS 1984, UTM39N) مشابه با متغیرهای محیطی بدست آمده از نقشه DEM تعریف شد. با روی هم اندازی موقعیت جغرافیایی پروفیل‌ها و این نقشه‌ها، مقادیر این متغیرهای محیطی در موقعیت جغرافیایی نیم‌رخ‌های خاک استخراج شدند. در نهایت، یک ماتریس با ابعاد ۱۹ (سری خاک) در ۲۴ (مجموع ویژگی‌های خاکی و متغیرهای محیطی) به دست آمد.

مبتنی بر محاسبه فاصله ماهالونوبیس در زمینه یافتن ارتباط بین خاک‌ها گسترش زیادی یافته است (Michéli et al., 2016) که به‌عنوان معیاری برای طبقه‌بندی آن‌ها در نظر گرفته می‌شود. بررسی‌ها بیانگر کاربرد گسترده این تکنیک در مطالعات خاک می‌باشد (به عنوان مثال، Hughes et al., 2017; Edokpayi et al., 2017; Palepu & Muley, 2017). علی‌رغم مشکلات و محدودیت‌هایی که در رابطه با روش‌های سنتی کلاس‌بندی خاک‌ها ذکر شده، نادیده گرفتن و به عبارت دیگر بلااستفاده قرار دادن آن‌ها، هدررفت علمی و حتی سرمایه ملی تلقی می‌شود. مطالعات انجام شده نشان می‌دهد که این داده‌ها کارایی خوبی در مطالعات جدید و نقشه‌برداری رقومی خاک دارند. بنابراین، تحقیقات و مطالعات در راستای بررسی کیفیت و کارایی آن‌ها رو به افزایش می‌باشد (Yang et al., 2011). این مطالعه با هدف گروه‌بندی عددی و کمی خاک‌ها با استفاده از تکنیک‌های مختلف طبقه‌بندی عددی کلاسیک و مدرن و مقایسه نتایج با گروه‌بندی و مرزبندی واحدهای خاک نقشه سنتی با استفاده از داده‌های محیطی در رزولوشن مکانی ۹۰ متری طراحی و اجرا شده است. بدین منظور، گروه‌بندی‌های خاک به دست آمده از نقشه موروثی منطقه شهرکرد-بروجن (استان چهارمحال و بختیاری) با گروه‌بندی‌های بدست آمده از تکنیک‌های خوشه‌بندی سلسله مراتبی، میانگین‌های کا، فاصله تاکسونومیکی و درختان تصمیم مقایسه و نتایج بصورت توصیفی تفسیر می‌شوند.

مواد و روش‌ها

منطقه مطالعاتی

منطقه مطالعاتی با وسعتی حدود ۳۶۵۰ کیلومتر مربع، در استان چهارمحال و بختیاری بین عرض‌های جغرافیایی ۳۱° و ۴۵' تا ۳۲° و ۳۰' شمالی و طول‌های جغرافیایی ۵۰° و ۴۵' تا ۵۱° و ۳۰' شرقی واقع شده است (شکل ۱). میانگین دمای سالانه ۱۱ درجه سانتی‌گراد و میانگین بارندگی سالانه ۳۲۰ میلی‌متر می‌باشد. رژیم حرارتی در کل منطقه مزیک می‌باشد. رژیم رطوبتی عمدتاً زیریک و در اراضی پست و مناطق با آثار ماندابی، اکوتیک می‌باشد. سنگ‌های رسوبی (آهکی، سنگریزه‌ای، شیل و مارن) و رسوبات آبرفتی مواد مادری



شکل ۱- موقعیت منطقه مطالعاتی در مرز استان‌های استان اصفهان و چهارمحال و بختیاری به همراه نقشه رقومی ارتفاع (الف) و واحدهای فیزیوگرافی منطقه (ب). محل پروفیل‌های شاهد مطالعه موروثی مورد بررسی مربوط به هریک از سری‌های خاک نیز بر روی نقشه‌ها مشخص شده‌اند.

Figure 1. Location of the study area coupled with the location of the representative soil profiles of the legacy soil map corresponding to each soil series on the digital elevation model (DEM) (a) and the physiographic units map of the area (b).

نحوه تغییر هریک از ویژگی‌های خاکی و محیطی در سری‌های خاک، طرح شماتیک تغییر میزان مقادیر استاندارد آن‌ها با استفاده از بسته "ggplot2" نرم‌افزار R تهیه گردید.

مدل‌های نظارت نشده طبقه‌بندی عددی خوشه‌بندی سلسله مراتبی

الگوریتم خوشه‌بندی سلسله مراتبی^۱ ساده‌ترین و پایه‌ای‌ترین الگوریتم در فرآیند داده‌کاوی نمونه‌ها از جمله خاک می‌باشد که برای دستیابی به یک دید کلی از نحوه تغییر نمونه‌ها و روابط بین آن‌ها بر اساس فاصله اقلیدسی مورد استفاده قرار می‌گیرد (Edokpayi et al., 2017; Zirnea et al., 2013). این آنالیز بصورت نظارت نشده و بدون نیاز به تعریف تعداد اولیه کلاس‌ها، مشاهدات را طی مراحل مختلف با یکدیگر ترکیب کرده و کلاس‌های بزرگ‌تری ایجاد می‌کند. این مدل سعی در یافتن خوشه‌های همگن مشاهدات با حداکثر شباهت درون کلاسی و حداکثر تفاوت با سایر کلاس‌ها دارد. بدین صورت که یک مشاهده با سایر مشاهدات درون آن کلاسی که به آن تعلق گرفته است، کمترین فاصله و با مشاهدات بیرون از کلاس، حداکثر تفاوت را از نظر ویژگی‌های

از آنجایی که این ویژگی‌ها واحدهای متفاوتی دارند، به منظور از بین بردن اثر مقیاس، ویژگی‌های مورد بررسی با استفاده از فرمول ۱ استاندارد شدند (Everitt et al., 2011).

$$\frac{x_i - \text{center}(x)}{\text{scale}(x)} \quad (1)$$

که x_i : متغیرهای مورد مطالعه، $\text{center}(x)$: میانگین مقادیر متغیرها و $\text{scale}(x)$: واریانس متغیرها می‌باشند.

تمامی محاسبات در محیط نرم‌افزار R (Development Core Team, 2017) انجام شدند. از آنجایی که نقشه موروثی خاک مورد بررسی بر اساس روش سنتی فیزیوگرافی که در آن مرز واحدها عمدتاً بر مبنای واحدهای فیزیوگرافی ترسیم شده‌اند (Mahler, 1970)، ارتباط ویژگی‌های هریک از سری‌های خاک نقشه موروثی با واحدهای فیزیوگرافی بررسی گردید. به این منظور، میزان هریک از ویژگی‌ها در واحدهای فیزیوگرافی متوسط‌گیری شدند. سپس، نحوه تغییرات آن‌ها تا عمق ۱۵۰ سانتی‌متری از سطح خاک در هریک از واحدهای فیزیوگرافی با استفاده از بسته "aqp" نرم‌افزار R ترسیم گردید. همچنین، با هدف آگاهی از

این روش توسط محققان زیادی در خاک‌شناسی استفاده شده است. به‌عنوان نمونه می‌توان به زرینا و همکاران (Kumar & Zirnea *et al.*, 2013)، کومار و کاناتسن (Kannathasan, 2013) و ادوکپی و همکاران (Edokpayi *et al.*, 2017) اشاره کرد. به منظور خوشه‌بندی داده‌های موروثی خاک‌ها با استفاده از این الگوریتم، ابتدا ماتریس فاصله بین خاک‌ها با توجه به ویژگی‌های خاکی و محیطی آن‌ها با استفاده از مفهوم فاصله اقلیدسی محاسبه گردید. سپس، روش اتصال کامل جهت دسترسی به بهترین کلاس‌بندی با حداکثر همگنی (Kaufman & Rousseeuw, 2005) با استفاده از تابع *hclust* نرم‌افزار R مورد استفاده قرار گرفت.

مدل‌های نظارت شده طبقه‌بندی عددی

میانگین‌های کامعمولی

آنالیز میانگین‌های کامعمولی^۱ یکی از معمول‌ترین روش‌های گروه‌بندی بر پایه تابع هدف می‌باشد که در کلاس‌بندی مشاهدات مورد استفاده قرار می‌گیرد. این آنالیز برخلاف روش خوشه‌بندی سلسله‌مراتبی، بر اساس تعداد کلاس‌های اولیه از پیش تعریف شده مشاهدات را بصورت تصادفی به گروه‌های مختلف وارد و با توجه به میانگین مقادیر مشاهدات درون هر کلاس، مرکز کلاس را تعیین می‌کند. مدل در مراحل متوالی مشاهدات را با توجه به فاصله بین آن‌ها و مرکز گروه‌ها به گروه‌های مشابه‌تر منتقل و بر اساس مشاهدات موجود در هر کلاس، مرکز جدیدی برای کلاس‌ها تعیین می‌کند. این فرآیند تا جایی تکرار می‌شود که مرکز گروه‌ها ثابت بماند و حداقل مجموع مربعات خطای فاصله اقلیدسی بین مشاهدات و مراکز کلاسی کاهش یابد. مراکز گروه‌ها با کاهش عدم شباهت تابع هدف (J) بصورت زیر محاسبه می‌شوند (Hartigan & Wong, 1979):

$$J = SSE = \sum_{k=1}^n \sum_{i=1}^c x_{ik} d_{ik}^2 \quad (3)$$

که J: مجموع مربعات خطا برای تمامی مشاهدات در مجموعه داده، n: تعداد مشاهدات، c: تعداد کلاس‌ها، d: ماتریس تشابه $c \times n$ ، x_k : نقطه مشاهداتی مربوط به کلاس i ام می‌باشد.

مختلف داشته باشد. برای تعیین شباهت بین مشاهدات از ماتریس فاصله استفاده می‌شود. روش‌های مختلفی برای تعیین فاصله وجود دارد که ساده‌ترین آن‌ها فاصله اقلیدسی می‌باشد (فرمول ۲).

$$d_{euc} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2)$$

که d_{euc} : فاصله اقلیدسی بین دو مشاهده $x = x_1, x_2, \dots, x_m$ و $y = y_1, y_2, \dots, y_m$ می‌باشد. که در آن مجموع تفاضل بین مقادیر یک متغیر در دو مشاهده x و y محاسبه می‌گردد.

به عنوان مثال، چهار مشاهده X, Y, Z و W را که هر یک معرف یک کلاس خاک می‌باشند را در نظر بگیرید. اگر بخواهیم فاصله بین این کلاس‌های خاک را با توجه به مقادیر دو متغیر درصد سنگریزه و کربن آلی ($X: 12, Y: 0.7, Z: 27, 0.2, W: 10, 0.3, X: 18, 0.5$) محاسبه کنیم بصورت زیر اقدام می‌شود؛

$$d_{x,y} = [(12-27)^2 + (0.7-0.2)^2]^{1/2}; d_{x,y} = 15; d_{x,z} = [(12-10)^2 + (0.7-0.3)^2]^{1/2}; d_{x,z} = 2/0.4; d_{x,w} = [(12-18)^2 + (0.7-0.5)^2]^{1/2}; d_{x,w} = 6; d_{y,z} = [(27-10)^2 + (0.2-0.3)^2]^{1/2}; d_{y,z} = 17; d_{y,w} = [(27-18)^2 + (0.2-0.5)^2]^{1/2}; d_{y,w} = 9; d_{z,w} = [(10-18)^2 + (0.3-0.5)^2]^{1/2}; d_{z,w} = 8$$

خروجی این مدل بصورت یک دندروگرام می‌باشد که ساختاری درخت مانند دارد. در پایین‌ترین سطح دندروگرام، در ابتدا هر مشاهده (در اینجا کلاس‌های خاک) به عنوان یک خوشه در نظر گرفته می‌شود. سپس در مراحل بعدی، نزدیک‌ترین دو خوشه (با کمترین فاصله) با یکدیگر ترکیب شده و خوشه‌های بزرگ‌تر را ایجاد می‌کنند. با ادامه این روند، در نهایت مجموع تمامی مشاهدات به عنوان یک خوشه بزرگ نهایی در نظر گرفته می‌شوند (Mohammadi, 2013). با در نظر گرفتن مثال بالا، ابتدا کلاس‌های X و Z که کمترین فاصله را با یکدیگر دارند، با یکدیگر ترکیب می‌شوند، سپس کلاس‌های W و Y که در فاصله بیشتری از این دو کلاس قرار دارند، به ترتیب به آن‌ها اضافه می‌گردند. بدین ترتیب، کلاس‌های اولیه خاک به سه گروه XZ, W و Y تفکیک می‌شوند.

1. k-means

و تا جایی ادامه پیدا می‌کند که تقسیم بیشتری ممکن نباشد (Mohammadi, 2013). از جمله بررسی‌ها و مطالعات داده‌کاوی که از مدل‌های درخت مدار استفاده کرده‌اند می‌توان به پالپو و مولی (Palepu & Muley, 2017) و ریبریو و همکاران (Ribeiro et al., 2014) اشاره کرد. در مطالعه حاضر، کلاس‌های اولیه بدست آمده از نقشه سنتی موروثی خاک برای مدل تعریف گردیدند. سپس، سری‌های خاک با استفاده از بسته C5.0 در نرم افزار R انجام شدند.

فاصله تاکسونومیکی

از آنجایی که ویژگی‌های خاک با یکدیگر همبستگی دارند، در تعیین فاصله و یا شباهت بین گروه‌های خاک، باید الگوریتمی استفاده شود که به این مسئله توجه دارد (Michéli et al., 2016). برای این منظور، فاصله ماهالونوبیس (فاصله تاکسونومیکی) بین خاک‌ها که با در نظر گرفتن کو-واریانس بین خاک‌ها آن‌ها را گروه‌بندی می‌کند، مورد استفاده قرار می‌گیرد. فاصله تاکسونومیکی بر مبنای محاسبه فاصله ماهالونوبیس بصورت زیر محاسبه می‌گردد:

$$d_{ij} = \sqrt{(x_i - x_j)^t S^{-1} (x_i - x_j)} \quad (4)$$

که d_{ij} عنصر ماتریس فاصله D با ابعاد $c \times c$ ، c تعداد گروه‌های خاک، S ماتریس کوواریانس می‌باشد. مقدار d_{ij} عبارت است از فاصله تاکسونومیکی بین گروه خاک i و گروه j و x ویژگی‌های مختلف مرتبط با هر خاک می‌باشد. با در نظر گرفتن مثال ارائه شده در بخش‌های قبل، با در نظر گرفتن کو-واریانس بین دو ویژگی مورد بررسی (۰/۵۲۷-)، فاصله بین خاک‌های مورد مطالعه بصورت زیر بدست می‌آیند:

$$d_{X,Y} = 2/43; d_{X,Z} = 2/3; d_{X,W} = 0/97; d_{Y,Z} = 2/38; \\ d_{Y,W} = 1/46; d_{Z,W} = 2/01$$

بنابراین با در نظر گرفتن مقادیر فاصله محاسبه شده، ابتدا گروه XW تشکیل شده و سپس خاک‌های Y و Z به ترتیب با فواصل بیشتر با آن‌ها ترکیب می‌شوند.

این معیار در مطالعات زیادی از جمله میشل و همکاران (Hughes et al., 2017) و هاقس و همکاران (Michéli et al., 2016) مورد استفاده قرار گرفته است. در این مطالعه، ابتدا ابعاد ماتریس داده‌ها با استفاده از آنالیز

به عنوان نمونه، با در نظر گرفتن مثال ذکر شده در قسمت قبل، فرض کنید بخواهیم مشاهدات را به دو گروه اولیه اختصاص دهیم. در اینصورت، ابتدا دو گروه فرضی ایجاد شده و میانگین هریک از ویژگی‌ها در گروه‌ها محاسبه می‌گردد. به عنوان مثال، $(0/5, 11) XZ$ و $(0/8, 21/5) YW$. در مرحله بعد فاصله اقلیدسی بین هریک از کلاس‌ها با مراکز گروه‌ها محاسبه می‌گردد:

$$d_{X(XZ)} = 1/02; d_{Y(XZ)} = 16; d_{Z(XZ)} = 1/02; d_{W(XZ)} = 7 \\ d_{X(YW)} = 10/5; d_{Y(YW)} = 4/5; d_{Z(YW)} = 12/5; d_{W(YW)} = 4/5$$

با توجه به نتایج، فاصله کلاس‌های Y و W به گروه YW یکسان می‌باشد. اگر در مرحله بعد مدل، گروه‌های $(0/5, 13/3) XZW$ و $(27, 0/2) Y$ را در نظر بگیرد، فاصله کلاس W تا کلاس XZW برابر با $4/7$ و تا کلاس Y برابر با 9 می‌باشد، بنابراین مدل، خاک W را به دلیل داشتن فاصله کمتر، به گروه XZW اختصاص داده و نتیجه نهایی را بصورت دو گروه XZW و Y ارائه می‌دهد.

انتخاب تعداد مناسب کلاس‌های اولیه تعریف شده برای مدل، از جمله مهم‌ترین مراحل در اجرای این مدل می‌باشد. یکی از معمول‌ترین روش‌ها، انتخاب یک محدوده از تعداد کلاس‌ها و اجرای مدل با هریک از آن‌ها می‌باشد. در نهایت، مدلی انتخاب می‌شود که به کمترین مربع خطا منجر شود. مطالعات زیادی از جمله کومار و کاناتسن (Kumar & Kannathasan, 2013) و هیل و همکاران (Heil et al., 2019)، گویای کارایی این مدل در خاک‌شناسی می‌باشد. در مطالعه حاضر، خاک‌ها با استفاده از الگوریتم میانگین‌های کا بسته نرم‌افزاری cclust نرم‌افزار R گروه‌بندی شدند.

درخت‌های طبقه‌بندی

مدل‌های درخت محور به‌عنوان یک الگوریتم داده‌کاوی می‌باشند. یکی از مهم‌ترین آن‌ها الگوریتم C5.0 می‌باشد (Quinlan, 1986). درخت‌های طبقه‌بندی با مدل‌سازی‌های گردشی و متوالی مجموعه‌ای از n واحد آماری (متغیرهای وابسته) را بر اساس مجموعه متغیرهای مستقل به تعدادی گروه تقسیم می‌کنند. تقسیم و گروه‌بندی متغیرها بر اساس یکسری قاعده تقسیمی و تصمیم بصورت اگر مقدمه، آن‌گاه نتیجه صورت می‌گیرد

در نقشه موروثی خاک، خاک‌ها بر اساس قرارگیری بر روی چهار واحد فیزیوگرافی فلات، آبریزه‌های سنگریزه‌ای (فن)، دشت و اراضی پست به ترتیب به چهار، نه، پنج و یک سری تفکیک شده‌اند. بیش از ۷۰ درصد از خاک‌های منطقه درصد سنگریزه بالا و یا لایه‌های سخت عمقی دارند که بر روی واحدهای فیزیوگرافی فلات و فن قرار گرفته‌اند (شکل ۲، جدول ۱).

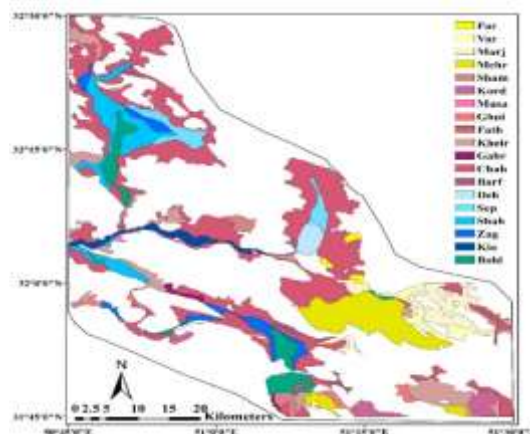
تجزیه به مؤلفه‌های اصلی (PCA) کاهش داده شد. سپس، کلاس‌های خاک نقشه موروثی برای مدل تعریف شدند. در مرحله بعد، فاصله تاکسونومی بین سری‌های خاک با استفاده از مفهوم فاصله ماهالانویس بسته HDMA نرم افزار R محاسبه و در نهایت، خاک‌ها گروه‌بندی گردیدند.

نتایج و بحث

جدول ۱- مشخصات سری‌های خاک نقشه موروثی مورد مطالعه

Table 1. General characteristics of studied legacy soil series

Physiographic unit	Soil series	Gravel (%)	Limiting layer	Clay accumulation and coatings	Slikenside and cracks	Gleyic condition	Soil Taxonomy (Soil Survey Staff, 2014)
Plateaus	Far	15-35	Calcareous (50cm)	-	-	-	Petrocalcic Calcixerepts
	Var	15-35	Calcareous (50cm)	-	-	-	Typic Xerorthents
	Marj	15-35	Lithic (50cm)	-	-	-	Petrocalcic Calcixerepts
	Mehr	15-35	Calcareous (100cm)	-	-	-	Petrocalcic Calcixerepts
	Sham	35-50	75% gravel (50 cm)	-	-	-	Lithic Xerorthents
Gravelly alluvial fans	Kord	35-50	Calcareous (100cm)	-	-	-	Petrocalcic Calcixerepts
	Musa	35-75	-	-	-	-	Typic Xerorthents
	Ghui	3-15	-	-	-	+	Typic Haploxererts
	Fath	3-15	-	-	-	-	Typic Xerorthents
	Kheir	3-15	-	-	-	-	Typic Calcixerepts
	Gahr	15-35	-	-	-	-	Typic Haploxerepts
	Chah	15-35	-	-	-	-	Typic Calcixerepts
	Barf	35-50	-	-	-	-	Typic Calcixerolls
	Deh	-	-	-	-	-	Typic Endoaquepts
	Sep	-	-	-	-	-	Typic Calcixerepts
Alluvial plains	Shah	-	-	+	-	-	Calcic Haploxeralfs
	Zag	-	-	+	-	-	Calcic Haploxeralfs
	Kia	-	-	+	+	+	Vertic Endoaqualfs
Low lands	Bold	-	-	-	-	+	Typic Endoaquepts



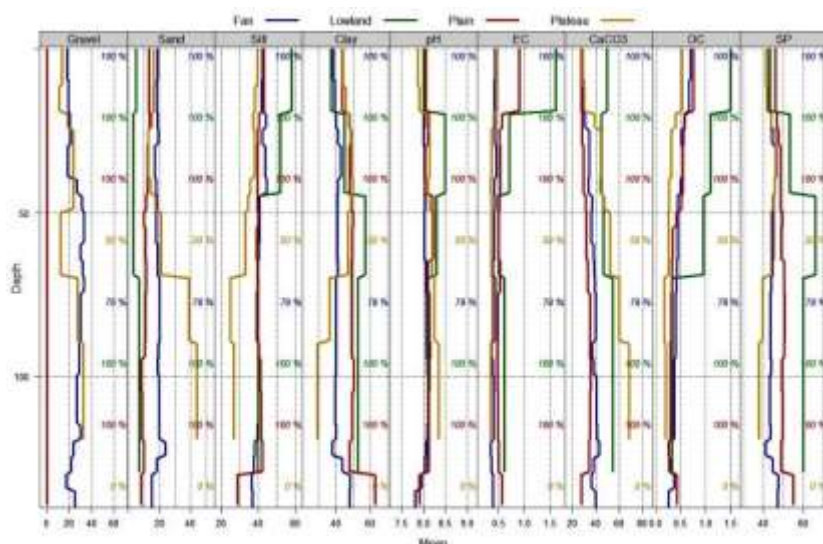
شکل ۲- نقشه موروثی خاک منطقه مورد مطالعه.

Figure 2. The legacy soil map of the study area.

آن‌ها، قابل توجه می‌باشند. بطوریکه تغییرات بافت مرتبط با جنس مواد مادری و شرایط خاک‌سازی آن‌ها می‌باشد. همچنین، روند تغییر هدایت الکتریکی و رطوبت اشباع با میزان رس در خاک مطابق می‌باشد (Karaca *et al.*, 2018).

نتایج نشان می‌دهد که ویژگی‌های خاکی هماهنگ با تغییر توپوگرافی تغییر می‌کند (Ovalles & Collins, 1986) و اینکه جداسازی خاک‌ها با توجه به ویژگی‌های فیزیکی و شیمیایی مورد بررسی آن‌ها، بخوبی با تغییر ویژگی‌های واحدهای فیزیوگرافی که بر روی آن‌ها قرار گرفته‌اند، همخوانی دارد.

شکل ۳، تغییر میزان ویژگی‌های خاک‌های پروفیل‌های شاهد سری‌های خاک موروثی در واحدهای فیزیوگرافی مختلف را نشان می‌دهد. اختلاف میزان درصد سنگریزه و اجزای بافت خاک بین واحدهای فیزیوگرافی فن و فلات با واحدهای دشت و اراضی پست به خوبی نشان داده شده است. در اراضی بالا دست (واحدهای فلات و فن)، خاک‌ها بیشترین میزان سنگریزه و شن را دارند که نتایج قابل انتظار می‌باشند. با حرکت به سمت میانه دشت و اراضی کم ارتفاع (واحد اراضی پست)، میزان سیلت سطحی، رس عمقی، اسیدیته و هدایت الکتریکی سطحی، کربن آلی و رطوبت اشباع افزایش می‌یابند. این یافته‌ها با توجه به موقعیت قرارگیری این خاک‌ها و ارتباط بین ویژگی‌های

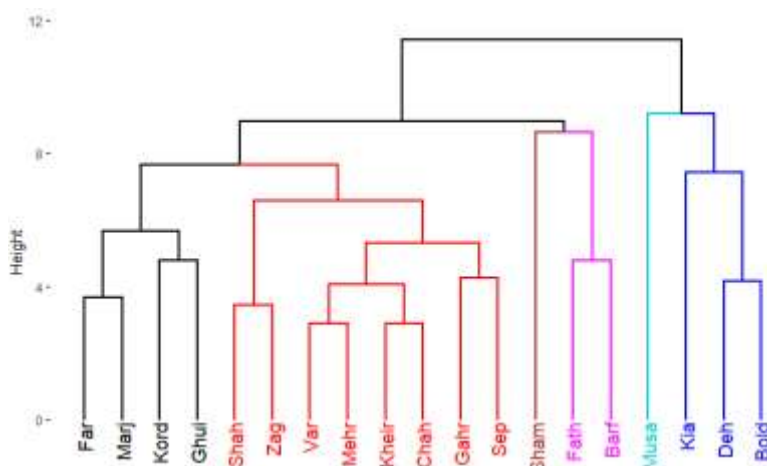


شکل ۳- تغییر میزان ویژگی‌های خاک در واحدهای فیزیوگرافی مختلف در عمق ۰ تا ۱۵۰ سانتی‌متری از سطح خاک. متوسط میزان ویژگی‌های خاک در هر یک از واحدهای فیزیوگرافی به رنگ‌های مختلف نشان داده شده است.

Figure 3. Changes of soil properties of legacy soil series in different physiographic units at a depth of 0 to 150 cm from the soil surface. The average of each soil characteristic corresponding to each physiographic unit is shown in different colors.

همچنین، می‌توان دریافت که خاک Musa از واحد آبرفت‌های سنگریزه‌دار (Gravelly alluvial fans) به دلیل دارا بودن بیشترین میزان سنگریزه، کربن آلی (ویژگی خاکی) و عمق دره (ویژگی محیطی)، وجه تمایز زیادی با سایر خاک‌های این واحد نشان می‌دهد. با یک نگاه کلی می‌توان متوجه شد که سه خاک Sham, Far, Marj و Fath به ترتیب کم عمق‌ترین خاک‌ها (عمق کمتر یا مساوی ۵۰ سانتی‌متر) در منطقه می‌باشند.

شکل ۴ تغییر مقادیر استاندارد ویژگی‌های فیزیوشیمیایی و محیطی خاک‌ها در سری‌های مختلف را نشان می‌دهد. مقادیر ویژگی‌ها بین ۲- تا ۳- متغیر می‌باشد که نشان می‌دهد در هر یک از سری‌های خاک، کدام ویژگی حداکثر و کدامیک کمترین می‌باشد. با در دست داشتن این اطلاعات، مقایسه کلی بین سری‌های خاک امکان‌پذیر می‌باشد. به عنوان مثال، بیشترین میزان اسیدیته و هدایت الکتریکی در سری Bold واقع در واحد فیزیوگرافی اراضی پست (Lowlands) دیده می‌شود.



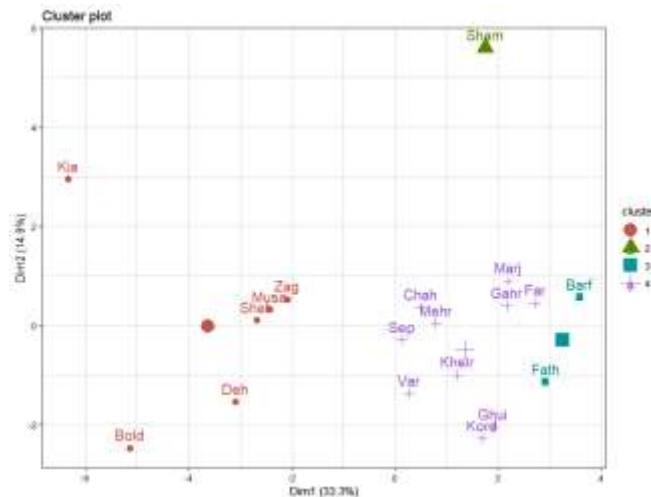
شکل ۵- دندروگرام خوشه‌بندی سلسله مراتبی سری‌های خاک. گروه‌های بدست آمده با رنگ‌های مختلف نشان داده شده‌اند.
Figure 5. The dendrogram arising from heretical clustering analysis of legacy soils. Obtained groups are differentiated in different colors.

امر می‌تواند تأییدی بر تفاوت این خاک‌ها نسبت به سایر گروه‌ها باشد. در گروه ۱ بدست آمده از این آنالیز، خاک‌هایی از واحدهای دشت و اراضی پست که ویژگی‌های نزدیک به هم دارند، در یک گروه قرار گرفته‌اند. همچنین، خاک Musa با وجود تفاوت مورفولوژیکی، میزان سنگریزه، موقعیت توپوگرافی و مدیریت، به گروه اضافه شده است. این امر می‌تواند به دلیل درصد بالای کربن آلی این خاک باشد. در این آنالیز، بجز گروه‌های دو و سه که بصورت استثنا (خاک‌های با شرایط خاص) جدا شده‌اند، خاک‌ها بطور کلی به دو دسته خاک‌های پایین دست (گروه ۱ بجز خاک Musa) و بالا دست (گروه چهار) تقسیم شده‌اند. اگرچه این طبقه‌بندی ممکن است از نظر تغییرات توپوگرافی در منطقه قابل توجه واقع شود، اما مقایسه این نتایج با جدول ۱ نشان می‌دهد که خاک‌های موجود در هر گروه، تفاوت‌های تاکسونومی و مدیریتی با یکدیگر دارند. همانند آنچه در مورد آنالیز قبل گفته شد، خاک‌ها بر اساس تغییر توپوگرافی و فیزیوگرافی منطقه تفکیک شده‌اند، با این تفاوت که در آنالیز میانگین‌های خاک‌های از گروه‌های مختلف، با یکدیگر ادغام شده‌اند.

بنابر آنچه از نتایج بر می‌آید، این مدل در برخی موارد خاک‌های با ویژگی‌های متفاوت از نظر مورفولوژی و موقعیت قرارگیری در زمین‌نما را با یکدیگر در یک گروه قرار داده است. اما از طرف دیگر، تعدادی خاک را علی‌رغم شباهت آن‌ها با سایر گروه‌ها، جدا کرده است. بطور کلی، مدل خاک‌های با شرایط خاص مانند سنگریزه زیاد، وجود لایه محدود کننده عمقی و همچنین شرایط ماندابی از سایر خاک‌ها، با موفقیت تفکیک کرده است. نتایج این مدل نقش بارز ویژگی‌های محیطی در تفکیک خاک‌ها را نشان می‌دهد و الگویی مشابه با الگوی فیزیوگرافی نقشه موروثی را دنبال کرده است. اگرچه در مواردی به دلیل ترکیب خاک‌های متفاوت از نظر مورفولوژی، موقعیت قرارگیری در زمین‌نما و مدیریت به‌عنوان یک گروه، نتایج معقولی در مقایسه با نقشه موروثی خاک ارائه نکرده است.

میانگین‌های کا معمولی

کلاس‌بندی‌های بدست آمده از آنالیز میانگین‌های کا معمولی به چهار کلاس نهایی با کمترین مربع خطا در شکل ۶ نشان داده شده‌اند. همانند آنالیز خوشه‌بندی سلسله مراتبی، خاک‌های Sham و همچنین Fath و Barf در این گروه‌بندی نیز از سایر خاک‌ها متمایز شده‌اند. این



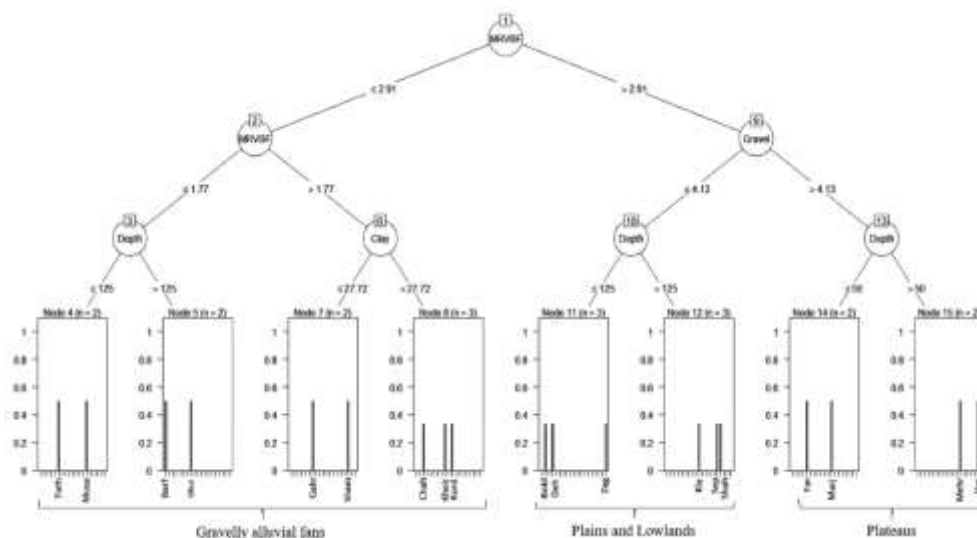
شکل ۶- گروه‌بندی سری‌های خاک بر اساس آنالیز میانگین‌های کا معمولی در دو مؤلفه اول PCA

Figure 6. The results of the k-means analysis in two first compounds of PCA.

خلوص گروه‌بندی از نظر ویژگی‌های مدیریتی و تاکسونومیکی همچنان باقی مانده است. به‌عنوان مثال، گروه‌بندی خاک‌های گره شماره ۵ با یکدیگر و همچنین در ترکیب با گره ۴، نیازمند بازنگری بیشتری می‌باشد. همچنین، خاک‌ها در گره‌های ۱۱ و ۱۲ به دلیل تفاوت در عمق و در مرحله بالاتر به دلیل سنگریزه، تفکیک شده‌اند. در حالی که انتظار می‌رفت چنین مدل نظارت شده‌ای با در دست داشتن اطلاعات جانبی بیشتر از جمله اطلاعات محیطی در رابطه با مشکل بالا بودن سطح آب زیرزمینی (به‌عنوان مثال، شاخص خیسی توپوگرافی)، نسبت به شخص خاک‌شناس (نتایج نقشه موروثی)، در تفکیک این خاک‌ها بهتر عمل می‌کرد. اگرچه خاک‌شناس به این مقوله بخوبی توجه داشته و در تفکیک این گونه خاک‌ها بهتر عمل کرده است.

درخت طبقه‌بندی

در این آنالیز، خاک‌ها در هشت گره نهایی قرار داده شده‌اند (شکل ۷). بر اساس این نتایج، MRVBF مهم‌ترین فاکتور متمایز کننده خاک‌های مورد بررسی می‌باشد و درصد رس، عمق خاک و درصد سنگریزه، در درجه‌های بعدی اهمیت قرار دارند. این آنالیز خاک‌ها را بر اساس همان الگوی سنتی فیزیوگرافی مدار تفکیک کرده است. بر خلاف آنالیزهای خوشه‌بندی میانگین‌های کا که خاک‌هایی از واحدهای مختلف با ویژگی‌های متفاوت را در گروه‌های با ویژگی‌های غیر مشابه قرار داده بود، در اینجا گروه‌بندی‌ها از این نظر خلوص بالاتری دارند. به‌عنوان مثال، خاک Musa اکنون از خاک‌های با شرایط متفاوت مثل Bold جدا شده و در کنار خاک سنگریزه‌ای Fath قرار گرفته است. با این وجود، مشکل



شکل ۷- گروه‌بندی خاک‌ها در آنالیز درخت طبقه‌بندی C.5

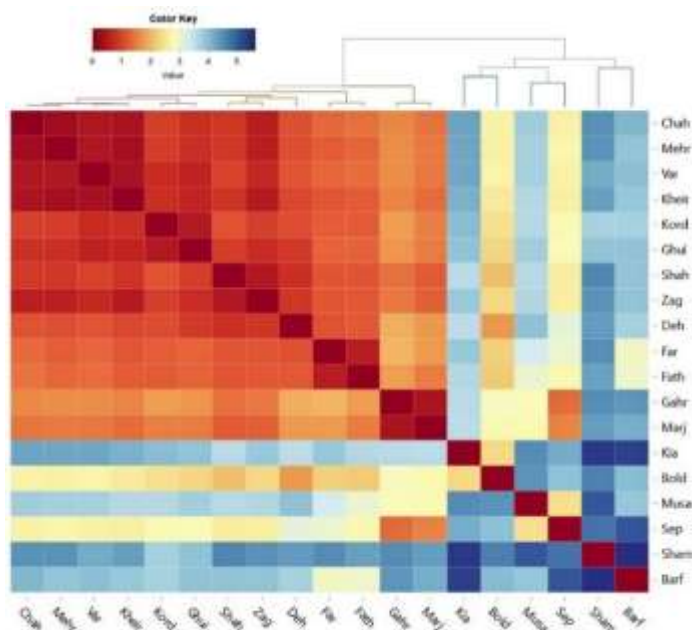
Figure 7. Grouping legacy soils from the C.5 tree.

فاصله تاکسونومیکی

کمترین فاصله را دارند و می‌توانند به‌عنوان یک گروه در نظر گرفته شوند. در آنالیزهای میانگین‌های کا، خاک Musa با شرایط و ویژگی‌های کاملاً متفاوت، با این خاک‌ها در یک گروه قرار گرفته بود. در حالیکه، این آنالیز بخوبی به تفاوت و فاصله خاک Musa با آن‌ها اشاره کرده است. از طرف دیگر، در آنالیز خوشه‌بندی سلسله مراتبی، این سه خاک با خاک‌هایی مثل Var و Mehr در یک گروه قرار داشتند. در این آنالیز اگرچه به فاصله کم این سه خاک با دو خاک دیگر اشاره دارد که به گونه‌ای با نتایج آنالیز خوشه‌بندی تطابق دارد، اما همان‌طور که در نتایج آنالیز درخت طبقه‌بندی نیز ذکر شد، این دو دسته خاک با یکدیگر تفاوت دارند که این امر بخوبی در آنالیز فاصله تاکسونومیکی با تفکیک این خاک‌ها به گروه‌های جداگانه مورد توجه قرار گرفته است. بنابراین، می‌توان نتیجه گرفت که در مقایسه با آنالیزهای کلاسیک و حتی مدل درخت طبقه‌بندی، این مدل با توجه به ویژگی‌های خاکی و محیطی مورد استفاده، خاک‌ها را بهتر و معقول‌تر تفکیک کرده است. اگرچه همچنان خاک‌های با مدیریت متفاوت در یک گروه قرار دارند (به‌عنوان مثال، ترکیب خاک Ghui با خاک Kord). در مقایسه با نقشه موروثی خاک که خاک‌ها را بر اساس تغییر واحدهای فیزیوگرافی و بطور سختگیرانه در ۱۹ سری متفاوت قرار داده بود، این آنالیز خاک‌های با شباهت بیشتر (به‌عنوان مثال Marj و Gahr) را با هم ترکیب کرده است.

نتایج این مدل نشان می‌دهد که شش خاک شامل Sham، Barf، Kia، Bold، Musa و Sep، به ترتیب بیشترین اختلاف و فاصله را با سایر خاک‌ها و همچنین با یکدیگر دارند (شکل ۸). در مقایسه با آنالیز میانگین‌های کا معمولی و خوشه‌بندی سلسله مراتبی که خاک‌های Sham، Musa و Barf را از سایر خاک‌ها متمایز کرده بودند، این آنالیز علاوه بر تأیید متمایز بودن این خاک‌ها از سایرین، به تفاوت خاک‌های Kia، Bold و Sep نیز اشاره دارد. مقایسه نتایج بدست آمده در این مرحله با توزیع ویژگی‌های خاک‌ها (شکل ۴) نشان می‌دهد که در این شش خاک، حداقل یکی از ویژگی‌های خاکی و یا محیطی آن‌ها حداقل یا حداکثر می‌باشد که باعث تمایز این خاک‌ها با سایر خاک‌های مشابه با آن‌ها (از نظر ویژگی‌های مورفولوژیکی، موقعیت قرارگیری بر روی واحدهای فیزیوگرافی) شده است.

بر اساس نتایج بدست آمده از این مدل، سایر خاک‌ها بطور کلی فاصله کمتری با یکدیگر دارند که مدل بر اساس تغییر ویژگی‌های خاکی و محیطی در دسترس، اقدام به تفکیک این خاک‌ها به ۶ گروه کرده است. گروه‌بندی‌های بدست آمده از این روش، در مقایسه با نتایج آنالیزهای قبل قابل قبول‌تر می‌باشند. به‌عنوان مثال، سه خاک Shah، Deh و Kia از قسمت پایین دست منطقه را در نظر بگیرید. این سه خاک با خاک‌های Bold و Kia



شکل ۸- نقشه موفقیت مدل فاصله تاکسونومیکی در یافتن فاصله بین خاک‌ها و گروه‌بندی‌های مربوطه.
Figure 8. Heatmap of Taxonomic distance between soils in different obtained groups.

مورد استفاده، مدلی که کمترین نسبت واریانس درون گروهی به واریانس بین گروهی (نسبت تغییر واریانس گروهی) و همبستگی درون کلاسی منجر شود، گروه‌بندی‌های بهتری ارائه داده و کارکرد بهتری در تفکیک خاک‌ها و بارز کردن تفاوت بین آن‌ها داشته است.

با توجه به این نکته که هدف از گروه‌بندی نمونه‌ها (در اینجا، خاک) بدست آوردن گروه‌هایی با بیشترین خلوص (کمترین واریانس درون گروهی) و بیشترین تفاوت (واریانس) بین گروه‌ها می‌باشد، این دو مشخصه برای ارزیابی گروه‌بندی‌های به دست آمده از آنالیزهای مختلف مورد بررسی قرار گرفتند (جدول ۲). بر اساس مفاهیم

جدول ۲- مقایسه آماری گروه‌بندی‌های بدست آمده از روش‌های مختلف مورد بررسی

Table 2. Statistical comparison of obtained soil groups from different studied methods

Analysis	Between group variance (BSS)	Within group variance (WSS)	WSS to BSS ratio	Relative variance	Within class correlation	Wilks' Lambda test	
						F	p-value
Legacy map	9.158	15.420	1.684	0.627	0.373	1.757	0.001
Clustering	11.611	14.658	1.262	0.558	0.442	7.641	0.002
K-means	12.051	16.301	1.353	0.575	0.425	2.226	0.239
C5 tree	10.608	16.094	1.517	0.603	0.397	12.812	0.014
Taxonomic distance	14.045	12.855	0.915	0.478	0.522	4.105	0.003

بررسی کیفیت گروه‌بندی‌ها می‌تواند مورد استفاده قرار گیرد. این آنالیز اخیراً در مطالعات خاک‌شناسی متفاوتی از جمله تصمیم‌گیری برای تفکیک اراضی با کارایی‌های مختلف مدیریتی (Uribeetxebarria *et al.*, 2018a)، یافتن الگوی آلودگی فلزات سنگین در کاربری‌های متفاوت (Qishlaqi *et al.*, 2009) و بررسی و یافتن شباهت بین واحدهای نقشه سنتی خاک (Toomanian & Esfandiarpour Boroujeni, 2017) مورد استفاده قرار گرفته است. به این منظور، آزمون لاندای ویلک برای بررسی میزان کارایی تفکیک گروه‌های خاک بر اساس مجموعه تمام ویژگی‌های خاکی و محیطی مورد استفاده در سطح احتمال ۹۵ درصد مورد استفاده قرار گرفت (جدول ۲). نتایج نشان داد بجز آزمون میانگین‌های کا معمولی، سایر مدل‌ها خاک‌ها را بطور مؤثری از یکدیگر تفکیک کرده‌اند. این آنالیزها بر کیفیت خوب نقشه موروثی خاک در نمایش و تمایز خاک‌ها از نظر مجموعه ویژگی‌ها نیز تأکید می‌کند. این درحالی است که نتایج آزمون هاتلینگ اجرا شده در بررسی خلوص و همگنی واحدهای نقشه خاک سنتی تهیه شده به روش ژئوپدولوژی در بخش‌هایی از منطقه بروجن استان چهارمحال و بختیاری، بیانگر کیفیت پایین واحدهای نقشه‌ها از این دید بودند (Esfandiarpour Boroujeni, 2017).

بیشترین نسبت تغییر واریانس گروهی و کمترین همبستگی درون کلاسی، مربوط به نقشه موروثی می‌باشد که مؤید کارایی کمتر این مدل کیفی نسبت به سایر مدل‌های کمی مورد استفاده است. پایین بودن کیفیت نقشه‌های سنتی خاک از نظر جداسازی صحیح واحدهای خاک، به مراتب در مقالات متفاوت عنوان شده است (Kempen *et al.*, 2012; Boettinger *et al.*, 2010). بازگلیا فیلهو و همکاران (Bazaglia Filho *et al.*, 2013) در مقایسه کارایی نقشه موروثی منطقه‌ای در برزیل با نقشه بدست آمده از روش کمی میانگین‌های کا، دریافتند که کارایی نقشه موروثی از نظر کیفیت جداسازی واحدهای خاک با رفتن به سطوح پایین طبقه‌بندی، کاهش می‌یابد. همانطور که قبلاً نیز اشاره شد، مدل درخت طبقه‌بندی به نتایج مشابه با نقشه موروثی دست یافته است. این امر در جدول ۲ با کارایی نزدیک این دو مدل به یکدیگر تأیید شده است. همچنین، نتایج نشان می‌دهند که مدل فاصله تاکسونومیکی به بهترین ترکیب گروه‌ها با بالاترین همبستگی درون کلاسی و کمترین نسبت تغییر واریانس گروهی منجر شده است. از آنجایی که گروه‌بندی خاک‌ها بر اساس ترکیبی از ویژگی‌های (فاکتورهای) مختلف انجام می‌شود، آنالیز تجزیه واریانس چندگانه که به این مقوله توجه کرده و تفاوت بین گروه‌های خاک را با توجه به ویژگی‌های مختلف بطور همزمان بررسی کرده و نشان می‌دهد، در

نتیجه‌گیری کلی

ممکن است دو خاک با ویژگی‌ها و کلاس تاکسونومی یکسان، به دلیل قرارگیری در دو واحد مجاور، از یکدیگر جدا شده باشند. این مدل که بر اساس قضاوت و تجربه شخص خاک‌شناس (و نه بر اساس روابط کمی بین خاک و عوامل موثر در تشکیل آن و یا در نظر گرفتن اصل پیوستگی مکانی خاک‌ها) برپا شده است، کیفی بوده (Kempen *et al.*, 2012; Zinck *et al.*, 2016) مورد توجه محققین قرار دارد. اما نتایج این مطالعه نشان داد مدل‌سازی کیفی موروثی با مدل‌های کمی، رقابت می‌کند. بنابراین، ترکیب این دو مدل می‌تواند به نتایج بهتری منجر شود (Ashtekar & Owens, 2013).

مقایسه مرزبندی و تفکیک خاک‌های نقشه موروثی با مدل‌های طبقه‌بندی عددی در رزلوشن مکانی ۹۰ متری نشان داد که گروه‌بندی‌های موروثی با مدل درخت طبقه‌بندی بیشترین همخوانی را دارند. اما مدل فاصله تاکسونومیکی به نتایج قابل قبول‌تری منجر شده است. مدل‌های عددی خاک واحدهای فیزیوگرافی مختلف را با یکدیگر ترکیب کرده‌اند. در حالیکه در مدل سنتی، خاک‌ها بدون در نظر گرفتن ویژگی تغییرپذیری مکانی و پیوستگی آن‌ها، با مرزهای دقیق از یکدیگر جدا شده‌اند.

References

- Ashtekar J.M., and Owens P.R. 2013. Remembering knowledge: an expert knowledge based approach to digital soil mapping. *Soil Horizons*, 54(5): 1-6.
- Bazaglia Filho O., Rizzo R., Lepsch I.F., do Prado H., Gomes F.H., Mazza J.A., and Demattê J.A.M. 2013. Comparison between detailed digital and conventional soil maps of an area with complex geology. *Revista Brasileira de Ciência do Solo*, 37: 1136–1148.
- Boettinger J.L., Howell D.W., Moore A.C., Hartemink A.E., and Kienast-Brown S. 2010. Digital Soil Mapping: Bridging Research, Environmental Application, and Operation. *Springer Science and Business Media*, 433p.
- Boettinger, J.L., Ramsey, R.D., Bodily, J.M., Cole, N.J., Kienast-Brown, S., Nield, S.J., Saunders, A.M. and Stum, A.K., 2008. Landsat spectral data for digital soil mapping. *In: Hartemink A.E., McBratney A.B., Mendona-Santos M. (Eds.), Digital Soil Mapping with Limited Data*. Springer, Dordrecht, pp. 193-202.
- Bouma J., Montanarella L., and Evanylo G. 2019. The challenge for the soil science community to contribute to the implementation of the UN Sustainable Development Goals. *Soil Use and Management*, 35(4): 538-546.
- Carré F., and Jacobson M. 2009. Numerical classification of soil profile data using distance metrics. *Geoderma*, 148: 336-345.
- Conrad O., Bechtel B., Dietrich H., Fischer E., Gerlitz L., Wehberg J., Wichmann V., and Böhner J., 2015. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. *Geoscientific Model Development*, 8: 1991–2007.
- Edokpayi A.A., Agho C.A., Adeb S.A., and Okpamen S.U. 2017. Comparison of the different hierarchical clustering techniques for the classification of soils under oil palm in Nigeria. *International Journal of Basic Science and Technology*, 3 (1): 37-46.
- Everitt B.S., Landau S., Leese M., and Stahl M. 2011. *Cluster Analysis* (5th Ed.). Chichester, West Sussex, U.K., Wiley. 346p.
- Hartigan J.A., and Wong M.A. 1979. Algorithm AS 136: A K-means clustering algorithm. *Applied Statistics*, 28: 100–108.
- Heil J., Haring V., Marschner B., and Stumpe B. 2019. Advantages of fuzzy k-means over k-means clustering in the classification of diffuse reflectance soil spectra: A case study with West African soils. *Geoderma*, 337: 11-21.
- Hughes P., McBratney A.B., Huang J., Minasny B., Micheli E., and Hempel J. 2017. Comparisons between USDA Soil Taxonomy and the Australian Soil Classification System I: Data harmonization, calculation of taxonomic distance and inter-taxa variation. *Geoderma*, 307:198-209.
- Karaca S., Gülser F., and Selçuk R. 2018. Relationships between soil properties, topography and land use in the Van Lake Basin, Turkey. *Eurasian Journal of Soil Science*, 7 (2): 115-120.
- Kaufman L., and Rousseeuw P.J. 2005. *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley, Hoboken, New York. 342p.

- Kempen B., Brus D.J., Stoorvogel J.J., Heuvelink G.B.M., and de Vries F. 2012. Efficiency comparison of conventional and digital soil mapping for updating soil maps. *Soil Science Society of America Journal*, 76: 2097–2115.
- Kumar D.A., and Kannathan N. 2013. A study and characterization of chemical properties of soil surface data using K-means algorithm. *In: International Conference on Pattern Recognition, Informatics and Mobile Engineering*, pp. 264-270.
- Mahler P.J. 1970. Manual of multipurpose land classification. Joint project with Food and Agriculture Organization (FAO-Undp). Publication No. 212. Ministry of Agriculture, Soil Institute of Iran.
- Michéli E., Láng V., Owens P.R., McBratney A.B., and Hempel J. 2016. Testing the pedometric evaluation of taxonomic units on soil taxonomy-a step in advancing towards a universal soil classification system. *Geoderma*, 264: 340-349.
- Mohammadi J. 2013. *Pedomining, Vol. 3, Postmodern Statistics*. Pelk Press. Tehran, Iran. 594p. (In Persian)
- Mohammadi M. 1986. Semi-detailed soil studies report Chaharmahal-Va-Bakhtiari province (Shahrekord and Borujen areas). Iranian Soil and Water Research Institute. No. 696. Tehran, Iran. (In Persian)
- Nield S.J., Boettinger J.L., and Ramsey R.D. 2007. Digitally mapping gypsic and natric soil areas using Landsat ETM data. *Soil Science Society of America Journal*, 71: 245-252.
- Ovalles F.A., and Collins M.E. 1986. Soil-landscape relationships and soil variability in north central Florida. *Soil Science Society of America Journal*, 50(2): 401-408.
- Palepu R.B., and Muley R.R. 2017. An Analysis of Agricultural Soils by using Data Mining Techniques. *International Journal of Engineering Science and Computing*, 7(10): 15167-15177.
- Qishlaqi A., Moore F., and Forghani G. 2009. Characterization of metal pollution in soils under two landuse patterns in the Angouran region, NW Iran; a study based on multivariate data analysis. *Journal of Hazardous Materials*, 172: 374–384.
- Quinlan J.R. 1986. Induction of Decision Trees. *Machine Learning*, 1(1): 81–106.
- R Development Core Team. 2017. R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria.
- Ribeiro M.V., Cunha L.M.S., Camargo H.A., and Rodrigues L.H.A. 2014. Applying a fuzzy decision tree approach to soil classification. *In: International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*. pp. 87-96.
- Rouse Jr J.W., Haas R.H., Schell J.A., and Deering D.W. 1974. Monitoring Vegetation Systems in the Great Plains with ERTS. *NASA Special Publication*, 351: 309-317.
- Soil Survey Staff. 2014. *Keys to Soil Taxonomy (12th Ed.)*. USDA-Natural Resources Conservation Service, Washington DC. 372p.
- Toomanian N., and Esfandiarpour Boroujeni I. 2017. Outcomes of applying a geopedologic approach to soil survey in Iran. *Desert*, 22-2: 239-247.
- Uribeetxebarria A., Arnó J., Escolà A., and Martínez-Casasnovas J.A. 2018a. Apparent electrical conductivity and multivariate analysis of soil properties to assess soil constraints in orchards affected by previous parcelling. *Geoderma*, 319: 185–193.
- Van Huyssteen C.W., Turner D.P., and Le Roux P.A.L. 2013. Principles of soil classification and the future of the South African system. *South African Journal of Plant and Soil*, 30(1): 23-32.
- Yang L., Jiao Y., Fahmy S., Zhu A-X., Hann S., Burt J.E., and Qi F. 2011. Updating Conventional Soil Maps through Digital Soil Mapping. *Soil Science Society of America Journal*, 75: 1044–1053.
- Zinck J.A., Metternicht G., Bocco G., and Del Valle H.F. (Eds.). 2016. *Geopedology: An Integration of Geomorphology and Pedology for Soil and Landscape Studies (1st Ed.)*. Springer, New York, 550p.
- Zirnea S., Lazar I., Saha Foudjo B-U., Vasilache T., and Lazar G. 2013. Cluster Analysis Based of Geochemical Properties of Phosphogypsum Dump Located Near Bacau City in Romania. *APCBEE Procedia*, 5: 317-322.

Comparison of Grouping and the Quality of Legacy Soil Map Boundaries with Numerical Data Mining Models: A Case Study of Some Regions of Chaharmahal-va-Bakhtiari Province

Zahra Rasaei^{1*}, Jahangard Mohammadi², Azam Jafari³

(Received: September 2019 Accepted: January 2020)

Abstract

Investigation of the the relationship between soils and grouping them based on different factors, plays an important role in different fields and aspects such as land management and sustainable agriculture. This launched by creating subjective or mental models and using environmental factors in the format of traditional soil maps. It afterward continued by relying on distance and similarity measurement through quantitative or mathematical models. This study aims to compare soil groups in soil maps with classical and modern models. For this purpose, legacy soil data of a map of Shahrekord-Borujen in Chaharmahal-va-Bakhtiari province is classified by using various algorithms: Heretical clustering, k-means, classification tree, and taxonomic distance. Soil groups gained from these numerical methods were then compared with soil groups in the legacy maps. Results of investigations in 90m spatial resolution showed that the classes from the decision tree were more in line with the legacy soil classes. The hierarchical clustering and k-means also resulted in group compositions similar to those of the legacy maps in terms of soil environmental and morphological characteristics that follow the traditional photographic units. The taxonomic distance led to the best combination of soil classes in the term of their traits with the highest within-class correlation (0.522) and the least within-group variance to between-group variance ratio. Low p -values in multivariate analysis of variance (MANOVA) between 0.001 in the traditional model up to 0.014 in the tree model showed that the models used in this study have effectively separated soils, except for K-means. Overall, findings show that using numerical classification models can discover the quantitative relations between soils. The surveyor can afterward modify the composition of soil classes considering his experience and knowledge of the study area in order to achieve more homogenous soil groups in terms of their management and taxonomic characteristics.

Keywords: Boundary of map unit, MANOVA, Traditional soil map, Numerical classification, Taxonomic distance

Rasaei Z., Mohammadi J. and Jafari A. 2021. Comparison of grouping and the quality of legacy soil map boundaries with numerical data mining models: a case study of some regions of Chaharmahal-va-Bakhtiari province. *Applied Soil Research*, 8(4):28-43.

¹ Ph.D. Student, Department of Soil Science, Faculty of Agriculture, Shahrekord University

² Professor, Department of Soil Science, Faculty of Agriculture, Shahrekord University

³ Assistant Professor, Department of Soil Science, Faculty of Agriculture, Shahid Bahonar University of Kerman

* Corresponding Author Email: zahra.rasaei@gmail.com