

مقایسه روش‌های مختلف نمونه‌برداری جهت پهنه‌بندی رقومی خاک در منطقه اردکان

روح‌الله تقی‌زاده مهرجردی^۱، فریدون سرمیدیان^{۲*}، مهدی تازه^۱، محمود امید^۲، نورابیر تومانیان^۳، محمد جواد روستا^۴ و محمد حسن رحیمیان^۵

^۱ استادیار، دانشکده کشاورزی و منابع طبیعی، دانشگاه اردکان، استاد، دانشکده مهندسی و فناوری کشاورزی، پردیس کشاورزی و منابع طبیعی دانشگاه تهران، ^۲ استادیار، مرکز تحقیقات کشاورزی و منابع طبیعی اصفهان و ^۴ استادیار، مرکز ملی تحقیقات شوری و ^۵ مربی، مرکز ملی تحقیقات شوری

تاریخ پذیرش: ۱۳۹۳/۰۹/۰۱

تاریخ دریافت: ۱۳۹۳/۰۲/۲۲

چکیده

اخیرا محققین از روش نقشه‌برداری رقومی برای غلبه بر مشکلات نقشه‌برداری سنتی خاک بهره گرفته‌اند. با توجه به تغییرات مکانی زیاد خاک، نمونه‌برداری مهم‌ترین مرحله در مطالعات نقشه‌برداری رقومی خاک محسوب می‌شود. هدف از پژوهش حاضر، مقایسه سه روش مختلف نمونه‌برداری شامل مربع لاتین، فازی-میانگین k و تصادفی به‌منظور دستیابی به یک الگوی فضایی مناسب برای نمونه‌برداری خاک در منطقه اردکان به وسعت ۷۲۰ کیلومتر مربع بود. متغیرهای محیطی که در این مطالعه استفاده شدند، عبارت است از مشتقات اولیه و ثانویه مدل رقومی ارتفاع، تصاویر ETM^+ ماهواره لندست و نقشه سطوح ژئومورفولوژی. نتایج بر اساس معیارهای آماری (میانگین و انحراف معیار) نشان داد که بهترین روش نمونه‌برداری، روش مربع لاتین می‌باشد. به‌عنوان مثال در مورد پارامتر کمکی شاخص خیزی مقدار میانگین داده‌های انتخاب شده به‌وسیله این روش و کل فضای نمونه‌برداری برابر با ۱۸/۱۹ می‌باشد. به‌طور مشابه مقدار میانگین شاخص همواری دره با درجه تفکیک بالا در روش مربع لاتین (۳/۶۴) بیشترین شباهت را با جامعه نمونه‌برداری (۳/۶۳) دارد. همچنین، توزیع فراوانی متغیرهای کمکی در محل‌های منتخب (نمونه) در روش مربع لاتین بسیار به توزیع فراوانی متغیرهای کمکی در کل منطقه (جامعه آماری) شبیه بود. همچنین، نتایج نشان داد که همیشه توزیع جغرافیایی مناسب (فازی-میانگین k) نشان‌دهنده الگوی نمونه‌برداری مناسب نمی‌باشد. بنابراین، روش مربع لاتین به‌عنوان بهترین روش نمونه‌برداری در منطقه مورد مطالعه معرفی می‌شود و پیشنهاد می‌شود که محققین در مطالعات آتی نقشه‌برداری خاک از روش مربع لاتین استفاده شود.

واژه‌های کلیدی: توزیع فراوانی، فازی-میانگین k ، مربع لاتین، نقشه‌برداری رقومی، نمونه‌برداری تصادفی

مقدمه

وجود آمده در پردازش اطلاعات، تکامل یافته است و با استفاده از مدل‌های مختلف، اقدام به ساده‌سازی پیچیدگی‌های موجود در سامانه طبیعی خاک می‌نماید (Minasny و همکاران، ۲۰۱۰). کاربرد موفقیت-

نقشه‌برداری رقومی خاک، بیان‌گر مجموعه‌ای از محاسبات رایانه‌ای برای پیش‌بینی پراکنش خاک‌ها در سیمای سرزمین می‌باشد که همگام با پیشرفت‌های به

بعضی روش‌ها هم از اطلاعات کمی استفاده نمی‌کنند. Lark (۲۰۰۰) مجموعه‌های فازی فضاهای شبکه‌ای را، در صورت وجود عدم قطعیت برای واریوگرام‌ها معرفی کرد. در سال‌های اخیر، روش مربع لاتین به‌عنوان یک روش نمونه‌برداری در جاهایی که نمونه قبلی خاک وجود ندارد توسط Minasny و McBratney (۲۰۰۷) پیشنهاد شده است. مربع لاتین یک طرح نمونه‌برداری مونتوکارلوی محدود شده است. این روش، یک طرح تصادفی طبقه‌بندی شده‌ای است که باعث نمونه‌برداری موثری به کمک توزیع چند متغیر می‌شود. Minasny و McBratney (۲۰۰۷) اقدام به مقایسه چندین روش نمونه‌برداری کردند و کارایی بالای نمونه‌برداری مربع لاتین را متذکر شدند. به‌طور مشابه Roudier و همکاران (۲۰۱۲) نیز کارایی بالای مربع لاتین را تایید کردند.

امروزه نقشه‌برداری رقومی خاک در سطح دنیا توانسته جایگاه مناسبی پیدا کند و تا حدودی جانشینی مناسب برای نقشه‌برداری سنتی باشد. اگر چه روش‌های مختلفی برای نقشه‌برداری رقومی وجود دارد و در مکان‌های مختلف جغرافیایی مورد ارزیابی قرار گرفته‌اند، اما ادامه این ارزیابی‌ها مخصوصاً در مناطقی که توزیع مکانی خاک‌ها پیچیده است، ضروری به نظر می‌رسد. به‌طوری‌که بعضی از روش‌ها در شرایط خاصی کارایی چندانی نداشته و باید مورد شناسایی بیشتر قرار بگیرند (Scull و همکاران، ۲۰۰۵). به‌عنوان مثال Minasny و McBratney (۲۰۰۷) بیان داشتند که در مناطقی با تغییرات توپوگرافی زیاد، نمونه‌برداری شبکه‌ای کارایی ندارد. همچنین، باید تاثیر متغیرهای محیطی که به‌عنوان داده‌های پیش‌بینی کننده در نقشه‌برداری رقومی خاک استفاده می‌شود، مورد بررسی قرار گیرد، چون با تغییر شرایط محیطی تاثیر هر کدام از پارامترها نیز دستخوش تغییر قرار می‌گیرد (به‌عنوان مثال اقلیم مرطوب در مقایسه با اقلیم بیابانی). لذا هدف از تحقیق حاضر مقایسه سه روش نمونه‌برداری (مربع لاتین، فازی-میانگین k و تصادفی) به‌منظور دستیابی به مناسب‌ترین روش نمونه‌برداری به‌عنوان اولین گام در تهیه نقشه رقومی خاک در منطقه بود.

آمیز نقشه‌برداری رقومی خاک به‌میزان داده‌های موجود از منطقه مطالعاتی و داده‌های محیطی به‌کار برده شده در مدل بستگی دارد. بنابراین در برخی از مناطق، با وجود داده‌های محدود، صحت نقشه‌های به‌دست آمده به‌عنوان چالشی برای این‌گونه مطالعات به‌حساب می‌آید (Scull و همکاران، ۲۰۰۳).

در نقشه‌برداری رقومی خاک، مکان‌های نمونه-برداری خاک برای ایجاد روابط کمی بین خاک و متغیرهای محیطی تأثیر بنیادین دارد. مشاهدات خاک شاخص‌هایی از تأثیر فاکتورهای خاک‌سازی بر خاک می‌باشند، کیفیت نقشه حاصله به موقعیت مکانی نمونه خاک بستگی بسیار خواهد داشت. در حقیقت صحت و دقت نقشه‌برداری رقومی خاک به تعداد و موقعیت مکانی نمونه بستگی دارد. از آن‌جایی‌که تمامی روش‌های پدومتری (نقشه‌برداری رقومی) سعی در حل نمودن معادله Jenny دارند، لذا مشاهدات صحرائی را با استفاده از رابطه‌ای ریاضی به متغیرهای تخمین زنده ارتباط می‌دهند.

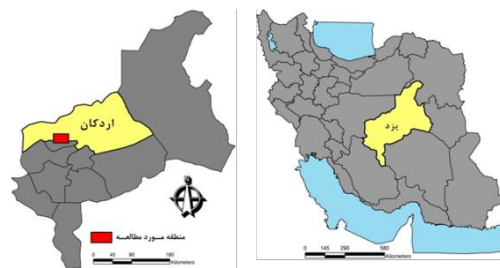
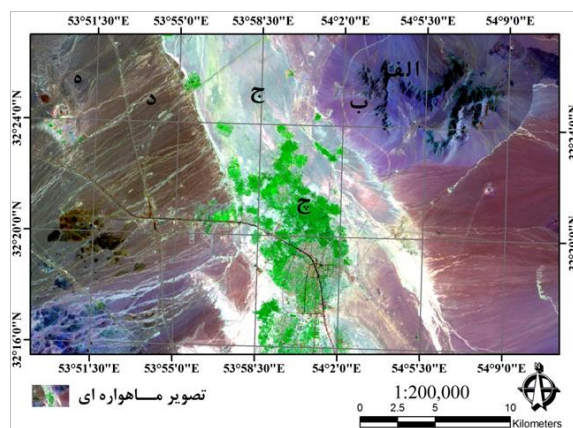
با توجه به این‌که، هرکدام از متغیرهای تخمین-زنده یک فضای مجازی را در محدوده مورد مطالعه تشکیل می‌دهند و کمیت یا طبقه خاک‌های مورد تحقیق تابعی از تغییرات تک‌تک آن‌ها می‌باشند، لذا باید شبکه نمونه‌برداری توانائی دخالت دادن انواع تغییرات موجود در فضاهای مجازی وارد شده به مدل را داشته باشد (Carre و همکاران، ۲۰۰۷).

از سال ۱۹۹۰ به بعد برای بهبود نحوه نمونه-برداری، روش‌های آماری متفاوتی در نقشه‌برداری رقومی خاک توسعه یافته که در بعضی از آن‌ها از اطلاعات کمی استفاده شده است. Heuvelink و همکاران (۲۰۰۷) نمونه‌برداری با حداقل کردن واریانس کریجینگ جهانی را طراحی کردند. Simbahan و Dobermann (۲۰۰۶) سه معیار متفاوت بهینه‌سازی را مقایسه کردند: حداقل کردن متوسط کوتاه‌ترین فواصل، یک توزیع یکنواخت از جفت نقاط برای تخمین واریوگرام و ترکیبی از هر دو. همه این روش‌ها شامل الگوریتم‌های شبیه‌سازی شده است. Hengl و همکاران (۲۰۰۳) نمونه‌برداری را در طول اجزاء اصلی متغیرهای کمی پیشنهاد کردند.

مواد و روش‌ها

منطقه مورد پژوهش: منطقه مورد مطالعه در ایران مرکزی واقع و وسعتی برابر با ۷۲۰۰۰ هکتار از اراضی دشت یزد-اردکان را شامل می‌شود. این منطقه بین طول‌های جغرافیایی ۵۰' ۵۳° تا ۱۰' ۵۴° شرقی و عرض‌های جغرافیایی ۱۶' ۳۲° تا ۲۷' ۳۲° شمالی قرار دارد و دارای مشکلات شوری منابع آب و خاک است. دامنه تغییرات شوری خاک در منطقه مورد مطالعه از یک تا ۱۵۸ دسی‌زیمنس بر متر متغیر می‌باشد. منطقه مورد نظر از جنوب به شهر اردکان و از شرق به رشته کوه هرشت و از شمال به چاله سیاه‌کوه و از غرب به رشته کوه‌های شیرکوه منتهی می‌شود (شکل ۱).

میانگین تبخیر و تعرق پتانسیل، دما و بارندگی سالیانه منطقه مورد مطالعه به ترتیب برابر ۳۴۸۳ میلی‌متر، ۱۸/۵ درجه سانتی‌گراد و ۷۵ میلی‌متر است. بارندگی بسیار کم بوده و محدود به بارش‌های زمستانه می‌باشد. رژیم رطوبتی و حرارتی خاک به ترتیب اریدیک و ترمیک طبقه‌بندی می‌شود. منطقه مورد مطالعه از نظر توپوگرافی بیشتر مسطح بوده و حداکثر و حداقل ارتفاع منطقه برابر با ۱۹۴۴ و ۹۴۴ متر است. اجزاء واحد اراضی منطقه با روند کلی شرق به غرب شامل کوه، مخروط‌افکنه‌های جدید، سطوح رسی و تپه‌های گچی می‌باشد.



شکل ۱- موقعیت منطقه مورد پژوهش را بر روی تصویر ماهواره‌ای ETM⁺ در استان یزد (الف: کوه، ب: مخروط افکنه‌های قدیمی، ج: اراضی یف کرده شور، د: باهادا، ه: تپه‌های گچی، چ: اراضی باغی)

بالای پشته با درجه تفکیک بالا^۹ و شیب حوضه^{۱۰} و تصاویر ماهواره‌ای (باند‌های تصویر ETM⁺، شاخص گیاهی نرمال شده^{۱۱}، شاخص رسی^{۱۲}، شاخص کربنات^{۱۳}، شاخص گچ^{۱۴}، شاخص شوری^{۱۵} و شاخص روشنایی^{۱۶}) بهره گرفته شد (Zinck و Metternicht، ۲۰۰۳؛ Andronikov و Dorbrov'skiy، ۱۹۹۱؛ Hengl و همکاران، ۲۰۰۴). لازم به ذکر می-

جمع‌آوری داده‌ها: در تحقیق حاضر برای تهیه داده-های کمکی، از منابع مختلف از جمله نقشه‌های طبقه-بندی (ژئومورفولوژی، زمین‌شناسی و پوشش گیاهی)، پارامترهای اراضی (قبیل شیب^۱، ارتفاع^۲، ارتفاع بالای شبکه زهکشی^۳، مساحت حوضه‌های اصلاح شده^۴، موقعیت میانی شیب^۵، عمق دره^۶، شاخص خیسی^۷، شاخص همواری دره با درجه تفکیک بالا^۸، شاخص

⁹ Multi resolution Ridge Top Flatness Index (MrRTF)

¹⁰ Catchment slope

¹¹ Normalized Difference Vegetation Index (B4 - B3) / (B4 + B3)

¹² Clay Index (B5 / B7)

¹³ Carbonate Index (B3 / B2)

¹⁴ Gypsum Index (B5 - B4) / (B5 + B4)

¹⁵ Salinity Index (B3 - B4) / (B2 + B4)

¹⁶ Brightness Index ((B3)² + (B4)²)0.5

¹ Slope

² Elevation

³ Altitude Above Channel Network

⁴ Modified Catchments Area

⁵ Mid-slop Position

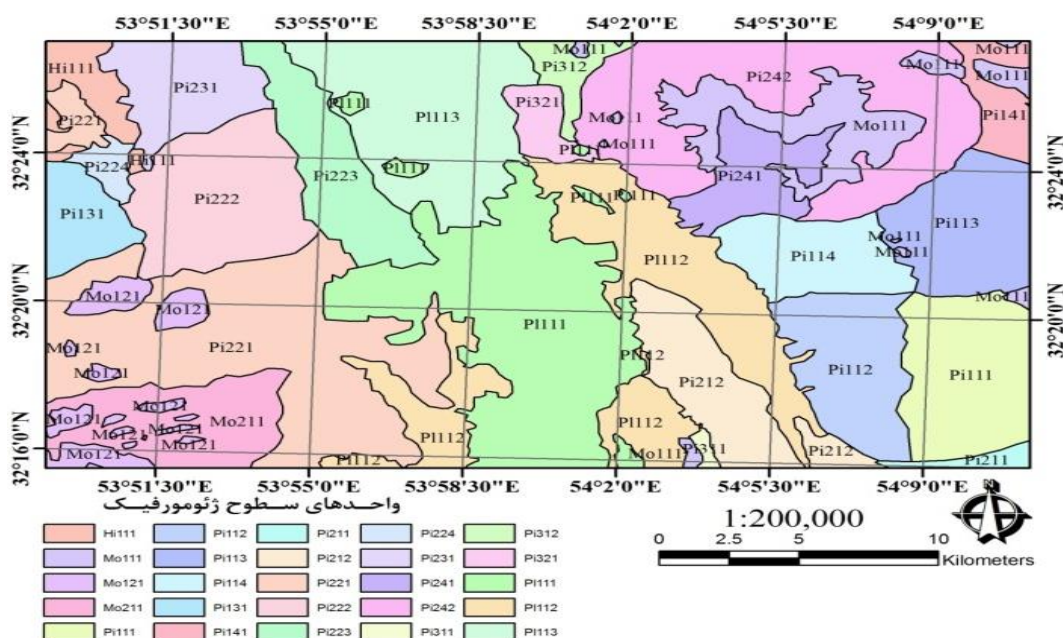
⁶ Valley Depth

⁷ Wetness Index

⁸ Multi resolution Index of Valley Bottom Flatness (MrVBF)

باشد که در این پژوهش، به علت این‌که مدل رقومی ارتفاع سازمان نقشه‌برداری کشور با تفکیک مکانی ۱۰ متر دارای نویزهای بسیار زیادی بود، از تبدیل موجک گسسته دو بعدی در جهت حذف این اطلاعات زائد و بهبود نقشه نهایی استفاده شد (Webster و Lark، ۲۰۰۴). نقشه ژئومورفولوژی بر اساس دستور کار Zinck (۱۹۹۰) تهیه شد (شکل ۲ و جدول ۱). همچنین، بعضی از این لایه‌های اطلاعاتی مانند نقشه زمین‌شناسی و پوشش گیاهی، تنها در محیط GIS رقومی شده و از آن‌ها استفاده شد.

باشد که در این پژوهش، به علت این‌که مدل رقومی ارتفاع سازمان نقشه‌برداری کشور با تفکیک مکانی ۱۰ متر دارای نویزهای بسیار زیادی بود، از تبدیل موجک گسسته دو بعدی در جهت حذف این اطلاعات زائد و بهبود نقشه نهایی استفاده شد (Webster و Lark، ۲۰۰۴).



شکل ۲- نقشه واحدهای ژئومورفولوژی

از بین ۸۴۲۷۲۶ پیکسل (۳۰ متری) از نرم‌افزار JMP استفاده شد.

طرح فازی- میانگین k: الگوریتم میانگین k یکی از روش‌های خوشه‌بندی^۴ ساده و سریع است. این الگوریتم دارای یک پارامتر به نام k است که تعداد خوشه‌هایی که باید به دست آید را مشخص می‌کند. به‌طور معمول، مرکز خوشه‌های اولیه به‌صورت تصادفی از میان نمونه‌های اولیه گزینش می‌شوند. به همین دلیل، مرکز خوشه‌های اولیه در دو خوشه‌بندی مستقل میانگین k می‌توانند متفاوت باشند. این موضوع موجب می‌شود که خوشه‌های به‌جا مانده از دو اجزای مختلف میانگین k با هم متفاوت باشند. در الگوریتم میانگین k می‌توان از معیارهای فاصله گوناگون بهره گرفت و خوبی یا بدی بکارگیری آن معیار بستگی به نوع داده‌هایی دارد که باید خوشه‌بندی شوند. در الگوریتم فازی- میانگین k

طراحی الگوی نمونه‌برداری

هایپر کیوب^۱: یک روش نمونه‌برداری تقریباً مدرن می‌باشد، که روشی طبقه‌بندی شده^۲ یا لایه‌ای است. در این روش متغیرهای کمکی به دسته‌های یکنواخت تقسیم‌بندی می‌شوند. روش کار هایپر کیوب به این صورت است که کاربر تصمیم می‌گیرد تا تعداد نمونه مشخصی را استفاده کند، این روش، داده‌های متغیرهای محیطی را به همان تعداد نمونه‌های خواسته به‌صورت خوشه تقسیم می‌کند و یک نمونه تصادفی از داده‌های ورودی متغیرهای محیطی هر خوشه را انتخاب می‌کند و سعی می‌کند تا پوشش کاملی از هر متغیر را ایجاد کند (Minasny و McBratney، ۲۰۰۶).

طرح تصادفی: در نمونه‌برداری تصادفی، هر پیکسل در کل منطقه مورد مطالعه شانس یکسانی برای برگزیده شدن دارد. جهت انتخاب ۱۵۰ نمونه تصادفی

^۳ K-Means

^۴ Clustering

^۱ Latin hypercube method

^۲ Stratified random sampling

ارتباط بین داده‌های نقطه‌ای و شاخص‌های هر خوشه به صورت فازی می‌باشد. در حقیقت به صورت درجه عضویتی که نشان‌دهنده درجه وابستگی داده نقطه‌ای را به مرکز خوشه نشان می‌دهد.

جدول ۱- راهنمای کدهای واحدهای ژئومورفولوژیکی مربوط به شکل

زمین‌نما	شکل زمین	لیتولوژی	سطح ژئومرفیک	علامت
کوه	رخنمون بریده شده	سنگ آهک و دولومیت	سطح سنگی	Mo111
		آندزیت خاکستری تا سبز به همراه آهک	سطح سنگی	Mo121
		آهک و شیل تیره	سطوح فرسوده	Mo211
تپه	رخنمون فرسایش یافته	گچ و ماسه سنگ	سیستم زهکشی شاخه درختی با ارتفاع نسبتاً زیاد	Hi111
دامنه	فن بادبزی شکل	آبرفت‌های سنگ‌های آهکی	فن فعال، قسمت بالائی	Pi111
			فن فعال، قسمت پائینی با شوری زیاد	Pi112
			فن فعال، قسمت بالائی، شبکه متراکم زهکش، سنگ-فرش بیایانی	Pi113
			فن فعال، قسمت پائینی، شبکه متراکم زهکش، سنگفرش بیایانی	Pi114
			فن فعال، قسمت پائینی، شبکه متراکم زهکش، دارای شیب نسبتاً زیاد	Pi131
			فن فعال، قسمت پائینی، دارای اراضی سنگلاخی	Pi141
	فن‌های متصل	آبرفت‌های سنگ‌های آهکی خاکستری با ماسه سنگ قرمز	قسمت بالائی، شیب زیاد، شبکه زهکشی متراکم	Pi211
			قسمت پائینی با شیب نسبتاً زیاد	Pi212
			قسمت بالائی با ذرات درشت و شوری نسبتاً زیاد	Pi221
			قسمت بالائی با شبکه زهکشی موازی	Pi222
			قسمت پائینی با آبرفت‌های جدید	Pi223
			باهادای کشت شده و شور	Pi224
			قسمت بالائی با ذرات درشت و شوری نسبتاً زیاد	Pi231
		آبرفت‌های تپه‌های گچی	ذرات درشت و شوری نسبتاً زیاد	Pi241
		آبرفت‌های سنگ‌های آهکی و دولومیتی	ذرات درشت با شبکه زهکشی متراکم	Pi242
	فن‌های متصل قدیمی	آبرفت‌های شیل و ماسه سنگ و گچ	پستی و بلندی کمتر و تقریباً مسطح با ارتفاع کم	Pi311
			پستی و بلندی بیشتر با آبراه‌های عمیق	Pi312
		آبرفت‌های گچی	پستی و بلندی بیشتر با آبراه‌های عمیق و ذرات درشت	Pi321
پلایا	گودی اردکان	آبرفت‌های دانه‌ریز و دانه‌درشت	سطوح رسی هموار، کشت شده	PI111
			سطوح رسی با شبکه زهکشی متراکم و شور	PI112
			سطوح رسی پف کرده با شوری بسیار زیاد و مرطوب	PI113

طبقه استفاده شد. در این روش فاصله داخل خوشه‌ها کمینه می‌شود. تابع هدف به صورت زیر می‌باشد:

$$J(M, C) = \sum_{i=1}^n \sum_{j=1}^k m_{ij}^0 d_{ij}^2 \quad (1)$$

در این تحقیق از نرم‌افزار فازی^۱ (Minasny و McBratney, ۲۰۰۲) برای خوشه‌بندی داده‌ها به k

^۱ FuzzMe

رتبه‌بندی کردن ورودی‌ها از نرم‌افزار اکسل استفاده شد. در مطالعه حاضر ده متغیری که بیشترین وزن را به خود اختصاص داده بودند، به‌عنوان ورودی در نظر گرفته شدند.

$$E_{ij} = -k \sum (p_{ij} \times \ln p_{ij}) \quad (5)$$

که در آن، E_{ij} آنترپپی، P_{ij} در آیه‌های ماتریس نرمال شده و k ضریب می‌باشد.

مقایسه روش‌های نمونه‌برداری: برای بررسی عملکرد مدل‌های مختلف در برآورد محل نمونه‌ها، توزیع هیستوگرامی و پارامترهای آماری داده‌های کمکی با فضای کل نمونه‌ها مورد مقایسه قرار می‌گیرد (McBratney و Minasny، ۲۰۰۶).

نتایج و بحث

ورودی‌های این مدل‌ها بر اساس روش آنترپپی شانون انتخاب شد (جدول ۲). با توجه به این جدول، ۱۰ پارامتر بر اساس روش آنترپپی شانون که دارای وزن بیشتری هستند، از قبیل نقشه ژئومورفولوژی، ارتفاع، شاخص خیزی، شاخص همواری دره با درجه تفکیک بالا، جهت حوضه، تجزیه به مولفه‌های اول و دوم تصویر ETM^+ ، شیب، موقعیت شیب میانی و انحنای شیب وارد مدل مربع لاتین و فازی-میانگین k انتخاب شد.

که در آن، \emptyset جزء فازی معادله است که درجه فازی شدن را نشان می‌دهد و d ماتریس فاصله می‌باشد و به‌صورت زیر محاسبه می‌شود:

$$d_{ij}^2 = (X_i - C_j)^T A (X_i - C_j) \quad (2)$$

که در آن، A ماتریس نرمال شده فاصله است که در اینجا از فاصله ماهالونوبیس^۱ استفاده شد. کمینه‌سازی تابع هدف J راه حلی را برای ماتریس توابع عضویت M و مرکز خوشه C به‌صورت زیر فراهم می‌کند.

$$m_{ij} = \frac{d_{ij}^{2/(\emptyset-1)}}{\sum_{j=1}^k d_{ij}^{2/(\emptyset-1)}} \quad i = 1, 2, \dots, n; j = 1, \dots, k \quad (3)$$

$$C_j = \frac{\sum_{i=1}^n m_{ij}^\emptyset X_i}{\sum_{i=1}^n m_{ij}^\emptyset} \quad j = 1, \dots, k \quad (4)$$

به‌طور خلاصه، برای تعیین بهترین تعداد طبقه (خوشه) از شاخص‌های آنترپپی خوشه‌بندی اصلاح شده^۲ و شاخص خوشه‌بندی فازی^۳ استفاده شد (McBratney و Minasny، ۲۰۰۲). سپس، در هر خوشه (بر حسب مساحت) تعدادی نقاط نمونه‌برداری (جمعاً ۱۵۰ نقطه) به‌صورت تصادفی انتخاب شد.

برای انتخاب ورودی‌های مدل‌های مربع لاتین و فازی-میانگین k از آنترپپی شانون مطابق با رابطه (۱) استفاده شد (Minasny و همکاران، ۲۰۱۰). ایده روش فوق، این است که هر چه پراکندگی در مقادیر یک شاخص، بیشتر باشد، آن شاخص نسبت به دیگر شاخص‌ها از اهمیت بیشتری برخوردار است. برای

جدول ۲- وزن‌های پارامترهای ورودی انتخاب شده برای مدل‌های مربع لاتین و فازی-میانگین k

پارامتر	ژئومورفولوژی	ارتفاع	شاخص خیزی	MrVBF	جهت	مولفه اول	مولفه دوم	شیب (درجه)	شیب میانی	انحنای شیب	وزن
	۰/۲۸	۰/۱۱	۰/۱۷	۰/۲۵	۰/۰۸	۰/۰۹	۰/۰۶	۰/۰۰۱	۰/۰۰۱	۰/۰۰۱	

تشکیل می‌دهد، پیدا کند. مقدار تابع هدف^۱ با افزایش تکرار کاهش می‌یابد و در شرایط تکرارهای کمتر از ۱۰۰۰ قدری آشفتگی از خود نشان می‌دهد. برای این تحقیق از تعداد تکرار ده‌هزار تا برای یافتن جواب بهینه استفاده شد. محلهایی که به‌وسیله مربع لاتین انتخاب شده در شکل ۷ ارائه شده است. باید توجه شود که مناطق شهری از نقشه اولیه حذف شدند.

همان‌طور که در جدول ۲ ملاحظه می‌شود، مجموع اوزان یک نمی‌شود که به‌دلیل این است که پارامترهای دیگر مقدار وزن کمتری را به خود اختصاص دادند و در نتیجه وارد مدل نشدند. پس از مشخص شدن پارامترهای ورودی، مدل مربع لاتین اجرا شد. تکامل معادله هدف با ده‌هزار تکرار در شکل ۳ ارائه شده است. این الگوریتم شروع به نمونه‌برداری تصادفی کرده تا بتواند نقاطی که مربع لاتین را

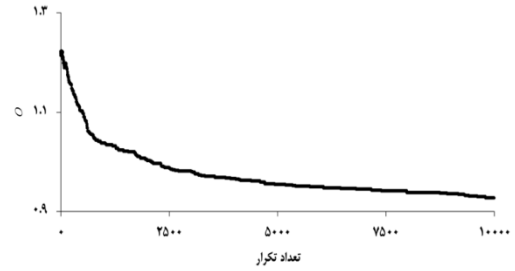
¹ Objective Function

¹ Mahalanobis Distance

² Modified Partition Entropies (MPE)

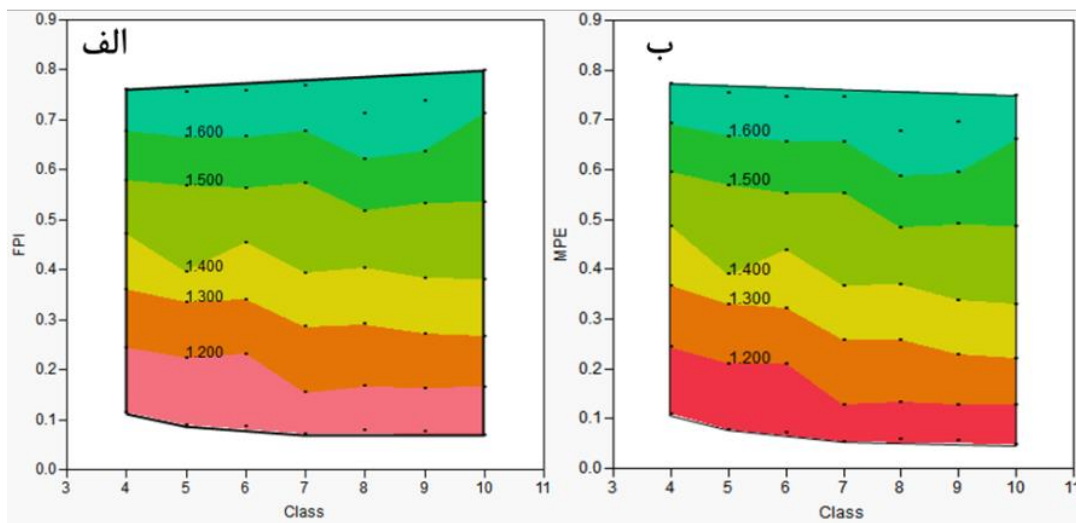
³ Fuzzy Partitioning Index (FPI)

باید تعداد طبقه بهینه شود که این انتخاب بر اساس شاخص‌های آنروپی خوشه‌بندی اصلاح شده^۱ و شاخص خوشه‌بندی فازی^۲ انجام پذیرفت (شکل‌های ۴ الف و ب). برای ارزیابی بهترین تعداد طبقه، به مقدار مثلاً شاخص خوشه‌بندی اصلاح شده دقت کرده و طبقه‌ای که دارای کمترین درجه فازی شدن و شاخص مورد نظر باشد انتخاب می‌شود که در این تحقیق تعداد هفت طبقه بهینه می‌باشد (شکل ۵). سپس در هر طبقه نسبت به وسعتی که در منطقه داشت تعدادی محل نمونه‌برداری انتخاب شد (شکل ۷).

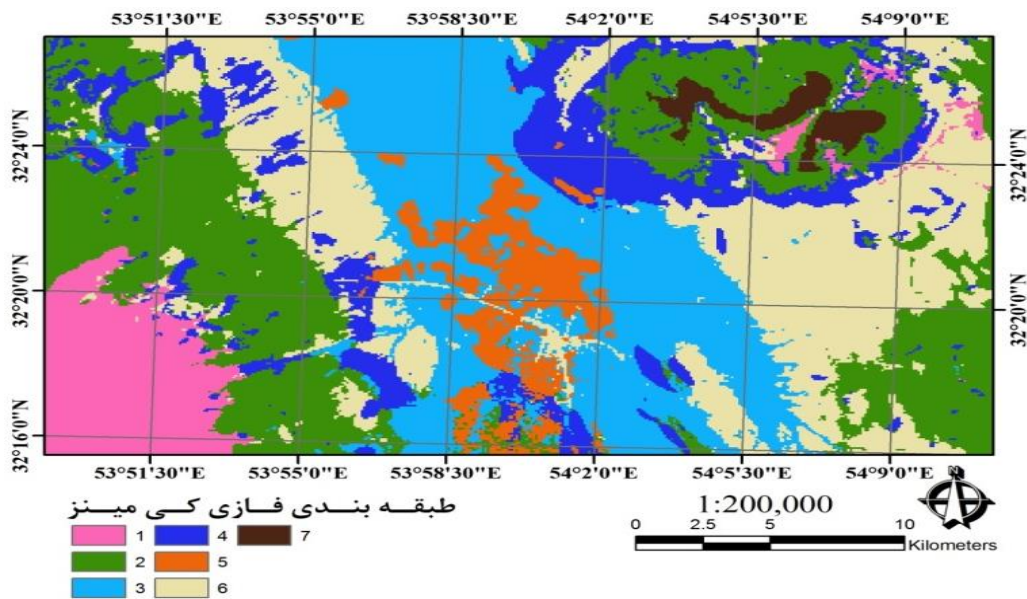


شکل ۳- تکامل معادله هدف با تعداد تکرار ده هزار

جهت نمونه‌برداری با استفاده از فازی-میانگین k ، از ورودی‌های جدول ۲ استفاده شد. در این طرح ابتدا



شکل ۴- تغییرات شاخص آنروپی خوشه‌بندی اصلاح شده و شاخص خوشه‌بندی فازی با درجات فازی شدن متفاوت



شکل ۵- نقشه طبقه‌بندی متغیرهای خاک بر اساس روش فازی-میانگین k

¹ Modified Partition Entropies (MPE)

² Fuzzy Partitioning Index (FPI)

Hunter Valley McBratney (۲۰۰۶) نیز در منطقه Hunter Valley به وسعت ۱۴ کیلومتر مربع در کشور استرالیا اقدام به مقایسه چهار روش نمونه‌برداری مربع لاتین، میانگین k ، مولفه‌های اصلی و تصادفی نمودند. ایشان از لایه رقومی ارتفاع، تصویر ماهواره‌ای ETM^+ و واحدهای یکنواخت خاک/لنداسکیپ به‌عنوان ورودی‌های مدل مربع لاتین استفاده کردند. نتایج ایشان نشان داد که روش مربع لاتین به خوبی هیستوگرام متغیرهای کمکی در منطقه را در بر گرفته است و نمونه‌برداری میانگین k تنها توزیع جغرافیایی مناسبی دارند.

همچنین، Brungard و Boettinger (۲۰۱۰) نیز شباهت توزیع فراوانی روش مربع لاتین به متغیرهای جامعه را دلیل برتری نسبت به سایر روش‌های نمونه‌برداری بیان کرده‌اند. Thomas و همکاران (۲۰۱۲) نیز جهت مشخص کردن محل‌های نمونه‌برداری در منطقه‌ای به وسعت ۲۳۰۰۰ هکتار در کشور فیلیپین، اقدام به مقایسه روش‌های تصادفی، شبکه‌ای و مربع لاتین کردند. نتایج ایشان حاکی از برتری مربع لاتین داشت.

Roudier و همکاران (۲۰۱۲) نیز به نتایج مشابه-ای در مورد مربع لاتین دست یافتند. علاوه بر این در روش نمونه‌برداری فازی-میانگین k امکان استفاده از متغیرهای مقوله (مثل ژئومورفولوژی) نمی‌باشد و تنها از پارامترهای کمی باید استفاده کرد، که این مسئله شاید دلیل دیگر عدم توزیع مناسب داده‌های انتخاب شده در این روش باشد. چون پارامتر ژئومورفولوژی به‌عنوان یکی از اصلی‌ترین متغیرها به‌وسیله روش آنتروپی وزن‌دهی شد و سپس برای مدل‌سازی انتخاب شد.

اجرای روش‌های نقشه‌برداری رقومی و رسیدن به نتایج دقیق، نیاز به اطلاعات ورودی دقیق با تراکم مناسب از منطقه مورد مطالعه دارد. اطلاعات ورودی شامل اطلاعات نقطه‌ای (خاک‌رخ‌ها)، چندوجهی (نقشه‌های چندوجهی از جمله زمین‌شناسی، ژئومورفولوژی و کاربری اراضی) و سلولی (یا پیکسلی شامل تصاویر ماهواره‌ای) می‌باشد که هر متغیر تخمین‌زننده به‌عنوان نماینده‌ای از فاکتور خاکسازي مربوطه وارد مدل می‌شود.

برای بررسی و انتخاب روش نمونه‌برداری مناسب، توزیع هیستوگرامی متغیرهای استفاده شده مدل‌ها (جدول ۳) در محل‌های منتخب شده رسم شد که نتایج در شکل ۶ ارائه شده است. همچنین، خصوصیات آماری این مکان‌ها به‌همراه خصوصیات آماری متغیرهای کمکی برای کل محدوده مورد مطالعه در جدول ۳ دیده می‌شود. به‌نظر می‌رسد که توزیع متغیرهای انتخاب شده به‌وسیله مربع لاتین بیشترین تبعیت را از توزیع جامعه دارند. به‌عنوان مثال در مورد پارامتر کمکی شاخص خیزی مقدار میانگین داده‌های انتخاب شده به‌وسیله این روش و کل فضای نمونه‌برداری برابر با ۱۸/۱۹ می‌باشد.

به‌طور مشابه مقدار میانگین شاخص $MrVBF$ در روش مربع لاتین (۳/۶۴) بیشترین شباهت را با جامعه نمونه‌برداری (۳/۶۳) دارد. همچنین، با دقت در مقادیر ضریب همبستگی متوجه شباهت ضرایب متغیرهای کمکی در محل‌های منتخب با فضای نمونه‌برداری می‌شویم. به‌عنوان مثال در مورد ارتباط بین پارامتر کمکی مولفه اول تصویر ماهواره‌ای و جهت شیب مقادیر کاملاً مشابه می‌باشد (۰/۰۸).

در حقیقت الگوی نمونه‌برداری مربع لاتین طوری نقاط نمونه را از جامعه انتخاب کرده است که توزیع نمونه‌گیری در هیستوگرام یکنواخت باشد، اما در روش‌های دیگر (تصادفی و فازی-میانگین k) شاهد نمونه‌برداری بیشتر یا کمتر در بعضی از قسمت‌های هیستوگرام‌ها و اریب هستیم.

درک و تلقی عمومی از نمونه‌برداری خوب آن است که می‌بایستی پراکنش جغرافیایی خوبی از نقاط نمونه‌برداری وجود داشته باشد. اما روش فازی-میانگین k با اینکه توزیع جغرافیایی مناسبی را در منطقه مورد مطالعه نشان می‌دهد اما نتوانسته پوشش مناسبی بر روی متغیرهای کمکی ایجاد کند.

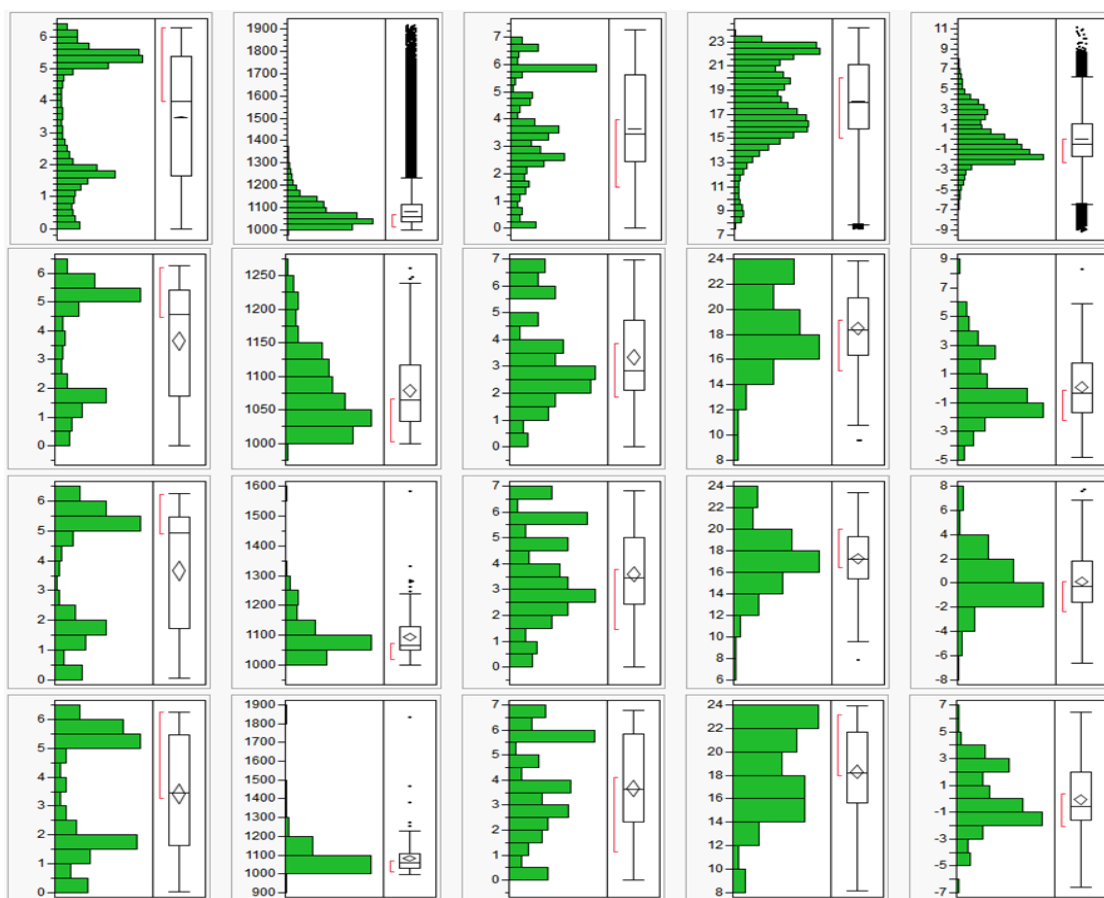
توانایی بالای مربع لاتین در انتخاب محل‌های نمونه‌برداری مناسب، ما را در مدل‌سازی مناسب و در نهایت تهیه نقشه رقوم مناسب شوری و طبقه خاک کمک خواهد کرد. این نتایج نشان داد که همیشه نمونه‌برداری با توزیع جغرافیایی مناسب تمام خصوصیات منطقه را مشخص نمی‌کند. Minasny و

جدول ۳- توزیع آماری و ضرایب همبستگی داده‌های اصلی و کمکی برای متغیرهای کمکی

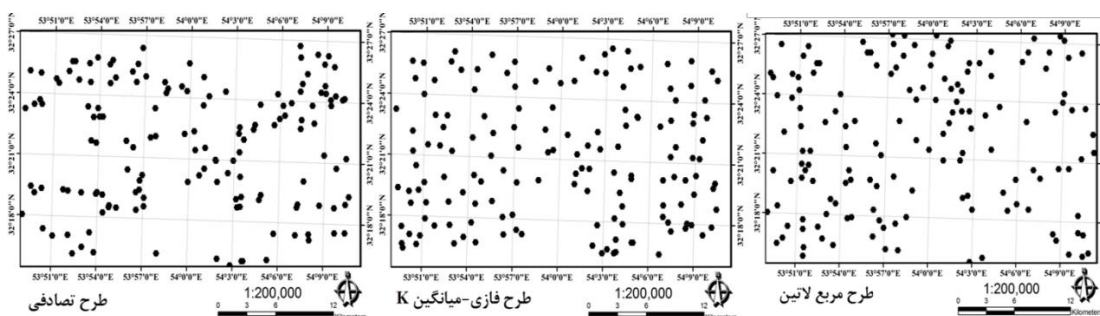
حداکثر	حداقل	انحراف معیار	میانگین	چهارک سوم	میانه	چهارک اول	تعداد داده		جهت شیب
۶/۲۸	۰/۰۰	۲/۰۳	۳/۴۸	۵/۳۸	۳/۹۸	۱/۶۴	۸۴۲۷۶۶	داده اصلی	
۶/۲۶	۰/۰۰	۱/۹۸	۳/۴۴	۵/۴۱	۴/۱۲	۱/۷۳	۱۵۰	مربع لاتین	
۶/۲۶	۰/۰۵	۲/۰۷	۳/۴۲	۵/۴۷	۳/۴۵	۱/۶۲	۱۵۰	تصادفی فاز-ی	
۶/۲۴	۰/۰۷	۲/۰۷	۳/۶۶	۵/۴۶	۴/۹۴	۱/۷۳	۱۵۰	میانگین k	
									ارتفاع
۱۹۱۵/۱۹	۹۹۷/۰۶	۷۵/۵۱	۱۰۸۱/۲۲	۱۱۱۱/۵۷	۱۰۵۹/۵۶	۱۰۳۲/۶۱	۸۴۲۷۶۶	داده اصلی	
۱۲۶۰/۶۲	۹۹۹/۸۸	۶۹/۸۲	۱۰۷۹/۰۸	۱۱۱۶/۴۱	۱۰۶۵/۴۵	۱۰۳۲/۷۴	۱۵۰	مربع لاتین	
۱۸۳۳/۳۳	۹۹۴/۹۷	۹۱/۷۲	۱۰۸۲/۷۵	۱۱۰۹/۷۵	۱۰۶۵/۸۰	۱۰۳۰/۰۳	۱۵۰	تصادفی فاز-ی	
۱۵۸۲/۶۳	۱۰۰۰/۲۲	۸۷/۱۳	۱۰۹۴/۷۶	۱۱۲۸/۵۴	۱۰۶۵/۵۰	۱۰۵۱/۵۲	۱۵۰	میانگین k	
									MrVBF
۷/۲۹	۰/۰۰	۱/۸۷	۳/۶۳	۵/۶۲	۳/۴۷	۲/۴۴	۸۴۲۷۶۶	داده اصلی	
۶/۹۵	۰/۰۰	۱/۸۰	۳/۶۴	۴/۷۰	۲/۸۳	۲/۱۱	۱۵۰	مربع لاتین	
۶/۷۹	۰/۰۰	۱/۹۹	۳/۶۸	۵/۸۴	۳/۶۲	۲/۳۳	۱۵۰	تصادفی فاز-ی	
۶/۸۳	۰/۰۰	۱/۸۰	۳/۶۰	۵/۰۱	۳/۴۴	۲/۴۳	۱۵۰	میانگین k	
									شاخص خیسی
۲۴/۱۴	۷/۴۷	۳/۳۸	۱۸/۰۹	۲۱/۰۸	۱۷/۹۹	۱۵/۷۸	۸۴۲۷۶۶	داده اصلی	
۲۳/۹۵	۹/۵۲	۳/۰۷	۱۸/۱۹	۲۰/۹۰	۱۸/۳۹	۱۶/۳۸	۱۵۰	مربع لاتین	
۲۳/۴۲	۸/۱۶	۳/۶۱	۱۸/۲۹	۲۱/۶۹	۱۸/۲۲	۱۵/۶۲	۱۵۰	تصادفی فاز-ی	
۲۳/۴۳	۷/۸۴	۲/۹۸	۱۷/۲۸	۱۹/۳۱	۱۷/۲۴	۱۵/۳۶	۱۵۰	میانگین k	
									مولفه اول
۱۱/۰۷	-۹/۱۹	۲/۳۲	۰/۰۰	۱/۵۱	-۰/۴۹	-۱/۶۶	۸۴۲۷۶۶	داده اصلی	
۸/۲۵	-۴/۸۲	۲/۳۵	۰/۰۵	۱/۷۵	-۰/۳۶	-۱/۷۰	۱۵۰	مربع لاتین	
۶/۴۷	-۶/۶۰	۲/۳۳	-۰/۰۶	۲/۰۲	-۰/۵۶	-۱/۵۸	۱۵۰	تصادفی فاز-ی	
۷/۶۸	-۶/۶۱	۲/۲۸	۰/۱۴	۱/۸۵	-۰/۳۶	-۱/۵۷	۱۵۰	میانگین k	

ادامه جدول ۳- توزیع آماری و ضرایب همبستگی داده‌های اصلی و کمکی برای متغیرهای کمکی

MrVBF	شخص	مولفه	ارتفاع- اول	MrVBF	ارتفاع- شخص	MrVBF	- شخص	-MrVBF	
ارتفاع- جهت	جهت	شیب	جهت	شیب	شیب	شیب	شیب	شیب	
۰/۰۸	-۰/۱۷	-۰/۱۵	-۰/۰۸	-۰/۶۴	-۰/۷۷	-۰/۴۳	۰/۸۶	۰/۵۱	داده اصلی
۰/۱۱	-۰/۱۶	-۰/۱۹	-۰/۰۸	-۰/۶۲	-۰/۷۷	-۰/۴۴	۰/۸۷	۰/۴۸	مربع لاتین
۰/۰۰	-۰/۰۷	-۰/۰۶	-۰/۰۱	-۰/۶۰	-۰/۷۴	-۰/۴۵	۰/۸۳	۰/۵۴	تصادفی
۰/۱۸	-۰/۱۹	-۰/۲۰	۰/۰۰	-۰/۶۲	-۰/۷۳	-۰/۴۴	۰/۸۱	۰/۴۶	فازی میانگین k



شکل ۶- هیستوگرام روش‌های مختلف از بالا به پایین: داده‌های اصلی، نمونه‌برداری مربع لاتین، نمونه‌برداری فازی-میانگین k، نمونه‌برداری تصادفی، متغیرهای مختلف از چپ به راست: جهت شیب حوضه، ارتفاع، شاخص خیزی، شاخص همواری دره با درجه تفکیک بالا و مولفه درجه اول تصویر ماهواره‌ای



شکل ۷- پراکنش مکانی نمونه‌های انتخاب شده به وسیله سه روش نمونه‌برداری

کل متغیرهای کمکی شبیه می‌باشد. همچنین، همیشه توزیع جغرافیایی مناسب نشان‌دهنده الگوی نمونه‌برداری مناسب نمی‌باشد. لذا پیشنهاد می‌شود که در مطالعات آتی نقشه‌برداری خاک از روش مربع لاتین استفاده شود.

در تحقیق حاضر اقدام به مقایسه سه روش نمونه‌برداری (مربع لاتین، فازی-میانگین k و تصادفی) جهت معرفی بهترین روش شد. نتایج نشان داد که روش مربع لاتین به خوبی قادر به تعیین توزیع محل-های نمونه‌برداری در منطقه مورد مطالعه می‌باشد، به طوری که فراوانی محل‌های منتخب بسیار به فراوانی

منابع مورد استفاده

1. Andronikov, V.L. and G.V. Dorbrolv'skiy. 1991. Theory and methods for the use of remote sensing in the study of soils. *Mapping Sciences and Remote Sensing*, 28: 92-101.
2. Brungard, C.W. and J.L. Boettinger. 2010. Conditioned Latin hypercube sampling: optimal sample size for digital soil mapping of arid rangelands in Utah, USA. Springer, Netherlands, 67-75.
3. Carre, F., A.B. Mc Bratney and B. Minasny. 2007. Estimation and potential improvement of the quality of legacy soil samples for digital soil mapping. *Geoderma*, 141: 1-14.
4. Hengl, T., D.G. Rossiter and A. Stein. 2003. Soil sampling strategies for spatial prediction by correlation with auxiliary maps. *Geoderma*, 120: 75-93.
5. Hengl, T., G.B.M. Huvelink and A. Stein. 2004. A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma*, 120: 75-93.
6. Heuvelink, G., D. Brus, J. de Gruijter. 2007. Optimisation of sample configurations for digital soil mapping with universal kriging. In: Lagacherie, P., A.B. McBratney and M. Voltz, (Eds.), *Digital Soil Mapping: An Initial Perspective*. Developments in Soil Science 31, Elsevier, Amsterdam.
7. Lark, R.M. 2000. Designing sampling grids from imprecise information on soil variability, an approach based on fuzzy kriging variance. *Geoderma*, 98: 35-59.
8. Lark, R.M. and R. Webster. 2004. Analyzing soil variation in two dimensions with the discrete wavelet transform. *European Journal of Soil Science*, 55: 777-797.
9. Metternicht, G. And J.A. Zinck. 2003. Remote sensing of soil salinity: potentials and constraints. *Remote Sensing of Environment*, 85: 1-20.
10. Minasny, B. and A.B. McBratney. 2002. FuzME version 3.0, Australian Centre for Precision Agriculture. McMillan Building A05, The University of Sydney, NSW 2006. <http://www.usyd.edu.au/su/agric/acpa>.
11. Minasny, B. and A.B. McBratney. 2006. A conditioned Latin hypercube method for sampling in the presence of ancillary information. *Computer and Geosciences*, 32: 1378-1388.
12. Minasny, B., A.B. McBratney and A.E. Hartemink. 2010. Global pedodiversity, taxonomic distance, and the World Reference Base. *Geoderma*, 155: 132-139.
13. Minasny, B. and A.B. McBratney. 2007. Incorporating taxonomic distance into spatial prediction and digital mapping of soil classes. *Geoderma*, 142: 285-293.
14. Roudier, P., A.E. Hewitt and D.E. Beaudette. 2012. A conditioned Latin hypercube sampling algorithm incorporating operational constraints. *Digital Soil Assessments and Beyond*. Taylor and Francis Group, 227-231.
15. Scull, P., J. Franklin and O.A. Chadwick. 2005. The application of classification of tree analysis to soil type prediction in a desert landscape. *Ecological Modelling*, 181: 1-15.
16. Scull, P., J. Franklin and O.A. Chadwick and D. McArthur. 2003. Predictive soil mapping: a review. *Progress in Physical Geography*, 27: 171-197.
17. Simbahan, G.G. and A. Dobermann. 2006. Sampling optimization based on secondary information and its utilization in soil carbon mapping. *Geoderma*, 133: 345-362.
18. Thomas, M., N.P. Odgers, A. Ringrose-Voase, G. Grealish, M. Glover and T. Dowling. 2012. Soil survey design for management-scale digital soil mapping in a mountainous southern Philippine catchment. *Digital Soil Assessments and Beyond*. Taylor and Francis Group, 233-238.
19. Zinck, J.A. 1990. *Physiography and Soils*. Lecture-notes for soil students. Soil Science Division. Soil survey courses subject matter: K6 ITC, Enschede, The Netherlands.

Comparison of different sampling methods for digital soil mapping in Ardakan region

Rohallah Taghizadeh Mehrjardi¹, Fereydon Sarmadian^{*2}, Mehdi Tazeh¹, Mahmood Omid², Norair Toomanian³, Mohammad Javad Rousta⁴ and Mohammad Hassan Rahimian⁵

¹ Assistant Professor, Faculty of Agriculture and Natural Resources, University of Ardakan, Iran,

² Professor, University College of Agriculture and Natural Resources, University of Tehran, ³ Assistant

Professor, Agricultural and Natural Resources Research Center, Isfahan, Iran, ⁴ Assistant Professor, National Salinity Center, Iran and ⁵ Scientific Board, National Salinity Center, Iran

Received: 12 May 2014

Accepted: 22 November 2014

Abstract

Recently, researchers are increasingly employed Digital Soil Mapping (DSM) techniques to overcome traditional soil mapping difficulties. Apparently, due to the large heterogeneity of soil environments, sampling may be the most important step in digital soil mapping studies. Therefore, in this research, we employed three different sampling strategies including Latin hypercube, Fuzzy-K-Means and random sampling to achieve the best spatial distribution of soil samples in an area around 720 km² located in Ardakan region, Yazd province, Iran. Auxiliary data that used in this study, were including terrain attributes, Landsat 7 ETM⁺ images and a geomorphologic surfaces map. Based on statistical criteria (i.e. mean and standard deviation), results showed that Latin hypercube is the best sampling method. For instance, in the selected points, the mean of wetness index is 18.19 which is the same as the mean of all area. Similarly, the mean of Multi-resolution Valley Bottom Flatness (MrRVF) in the points selected by Latin hypercube strategy is very similar to all area. Moreover, histogram of auxiliary data in selected points (samples) was more similar to histogram of auxiliary data in all area. Also, the results indicated that a good geographical coverage (Fuzzy-K-Means) does not adequately represent the distribution of the variables. Therefore, Latin hypercube is the best strategy to determine sample locations in our study area and hence, it is recommended that researchers apply Latin hypercube method in future digital soil mapping studies.

Key words: Digital mapping, Frequency distribution, Fuzzy-K-Means, Latin hypercube, Random sampling

* Corresponding author: fsarmad@ut.ac.ir