استفاده از مدل مکعب لاتین مشروط (cLHS) برای طراحی نقاط مطالعاتی در پدولوژی

سید عرفان خاموشی[۱]، فریدون سرمدیان*[۲]، علی کشاورزی[۳]
به ترتیب دانشجوی سابق کارشناسی ارشد، استاد و استادیار علوم و مهندسی خاک دانشگاه تهران

چکیده

جهت حفظ حاصلخیزی و افزایش بهره‌وری خاک در کشاورزی پایدار، آگاهی از ویژگی‌های مختلف آن ضروری است. در این تحقیق مدل مکعب لاتین مشروط برای تعیین نقاط نمونه‌برداری در مطالعات خاک‌شناسی موردبررسی قرار گرفت. منطقه موردمطالعه واقع در استان قزوین، شهرستان آبیک می‌باشد. رژیم رطوبتی و حرارتی منطقه به ترتیب زریک و ترمیک می‌باشد. با استفاده از تصاویر ماهواره لندست ۸ و تصاویر سنجنده استر مجموعاً تعداد ۱۹ متغیر کمکی استخراج گردید و سپس آنالیز مؤلفه‌های اصلی به منظور انتخاب مؤثرین متغیرها انجام گردید. متغیرهای انتخابی وارد مدل گردیده و درنهایت تعداد ۵۰ نقطه به‌عنوان نقاط مناسب برای نمونه‌برداری در منطقه موردمطالعه انتخاب گردید. همچنین دقت روش با رسم تابع تکامل هدف مورد آزمون قرار گرفت که خطای پایین حاصل از آن نشان‌دهنده کارایی مدل جهت طراحی نقاط نمونه‌برداری در پدولوژی می‌باشد.

واژه‌های کلیدی: متغیر کمکی، تابع تکامل هدف، آنالیز مؤلفه‌های اصلی.

مقدمه

در مقیاس زندگی انسانی خاک یک منبع تجدید ناپذیر و محیطی پویا بوده که با مدیریت و کاربری ناصحیح اراضی، به‌راحتی مستعد تخریب می‌گردد. ازاین‌رو گسترش تخریب این منبع محدود می‌تواند امنیت غذایی جهان را شدیداً در معرض خطر قرار دهد و تهدیدی برای سلامت محیط زیست باشد. به منظور بهره برداری بهینه از خاک لزوم یک برنامه‌ریزی صحیح و طولانی مدت در زمینه استفاده از اراضی بر اساس استعداد و نوع نیازهای جامعه امروز انسانی مشهود است (McKenzie et al, 2008). در این راستا اطلاعات منابع اراضی بخصوص نقشه‌های خاک یکی از بنیادی‌ترین نوع داده‌ها هست. روش‌های تخمین مکانی خاک با تأکید بر منابع داده‌های قابل دسترس جهت به تصویر کشیدن تغییرات مکانی خاک‌ها با هزینه کم‌تر و در مدت زمان کوتاه‌تر به گونه‌ای که نقشه‌ای قابل قبول از خاک ارائه شود، ارائه شده است (McBratney et al, 2003). نقشه‌برداری رقومی خاک در واقع توسعه یک مدل عددی یا آماری از رابطه بین متغیرهای محیطی و خصوصیات خاک است که برای داده‌های جغرافیایی زیادی به منظور تولید نقشه رقومی به کار می‌رود (Minasny et al, 2010). مکان‌های نمونه‌برداری، در نقشه‌برداری رقومی 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 (تقی زاده مهرجردی و همکاران، ۱۳۹۳). هدف از نمونه برداری بدست آوردن اطلاعاتی است که قادر به برآورد برخی پارامترهای آماری یا مکانی در یک منطقه خاص برای برنامه‌های مدیریتی است. نمونه برداری به واسطه برخی شرایط از جمله هزینه ها و منابع در دسترس محدود شده است بنابراین نیاز به داشتن یک استراتژی در نمونه برداری امری اجتناب ناپذیر است (Minasny and McBratney, 2006a).

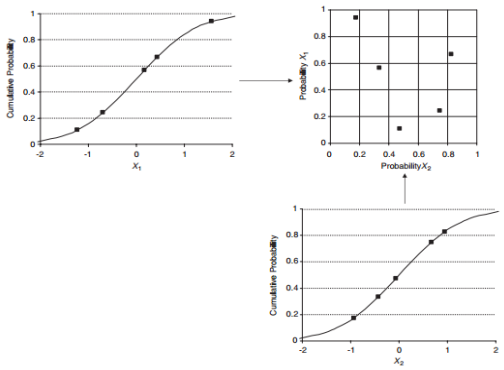
محققانی بحث در مورد دو رویکرد اساسی در نمونه برداری خاک را آغاز کردند که شامل ۱- نمونه برداری مبتنی بر تئوری: که از مطالعات کلاسیک پیروی می‌کند و بر پایه تئوری احتمالات بنا شده و با یک تخمین بی طرفانه از داده‌های آماری نقاط نمونه برداری را مشخص می‌کند. ۲- نمونه برداری مبتنی بر مدل: که براساس اصول زمین آماری است و زمانی که

* نویسنده مسئول fsarmad@ut.ac.ir

متغیرهای مکانی و پیش بینی آنها مد نظر است، عملکرد بهتری نسبت به نمونه برداری مبتنی بر تئوری دارد (Brus and Gruijter, 1997).

در بسیاری از مطالعات، نمونه برداری از یک منطقه می‌تواند در حضور داده‌های فرعی یا کمکی باشد. برای مثال در یک منطقه می‌توان از اندازه‌گیری‌های ژئوفیزیکی، تصاویر سنجش از دور، نقشه‌های پوشش گیاهی، نقشه‌های ژئولوژی و مدل رقومی ارتفاع به عنوان داده‌های فرعی استفاده کرد (Minasny and McBratney, 2006a). این داده‌ها باید طوری باشد که در تمام منطقه مورد نظر قابل دسترس باشد. داده‌های فرعی اغلب برای تخمین مکانی بسیار مفید هستند مخصوصاً زمانی که دسترسی به داده‌های اصلی توسط عواملی محدود شده باشد (Goovaerts, 1997). هرکدام از این داده‌های فرعی یک فضایی از ویژگی‌های منطقه را به خود اختصاص می‌دهد بنابراین بهتر است بجای نمونه برداری از نواحی مختلف ژئوگرافیگی از نواحی که این داده‌ها به وجود آورده‌اند استفاده کنیم، چالش اصلی برای انجام این کار طراحی یک روش نمونه برداری است که بتواند تمام این نواحی را در نظر بگیرد (Minasny and McBratney, 2006a). پیش بینی خصوصیات خاک و کلاس خاک در نقشه برداری رقومی خاک بر اساس رابطه بین داده‌های بدست آمده از مطالعات میدانی، داده‌های کمکی و متغیرهای محیطی استوار است (McBratney et al, 2003)؛ اما نمونه برداری در حضور داده‌های کمکی زمانی مفید واقع می‌شود که هرکدام از این متغیرها قادر باشد محدوده قابل توجهی از منطقه را پوشش دهند و در قالب مکعب‌هایی پوشش را انجام دهند. یک راه که این امر را امکان پذیر می‌کند استفاده از طرح مکعب لاتین است.

نمونه برداری مکعب لاتین یک طرح تصادفی طبقه بندی شده است که موجب نمونه برداری مؤثری به کمک توزیع چند متغیره می‌شود. در این روش کل جمعیت بر اساس یک سری از خصوصیات به زیرگروه‌هایی تقسیم می‌شوند که اعضای این زیرگروه‌ها نسبت به کل جمعیت به هم شبیه تر می‌باشند. سپس از داخل هرکدام از این زیرگروه‌ها به صورت تصادفی نمونه انتخاب می‌شود (Iman and Conover, 1980). یک شبکه مربع حاوی موقعیت‌های نمونه، شبکه لاتین است اگر و تنها اگر فقط یک واحد نمونه برداری در هر ردیف و هر ستون وجود داشته باشد. یک مکعب لاتین تعمیم این مفهوم به تعداد اختیاری از ابعاد است، به‌طوری که هر واحد نمونه برداری فقط در یک صفحه آن قرارگرفته باشد. مکعب لاتین شامل نمونه برداری n مقدار از توزیع تشریح شده هر متغیر هست. توزیع تجمعی هر متغیر به n فاصله با احتمال مساوی تقسیم می‌شود و یک مقدار از هر فاصله به‌طور تصادفی انتخاب می‌گردد. سپس n مقدار بدست آمده برای هر متغیر با متغیر دیگر جفت می‌شود که باعث ایجاد پوشش کاملی از هر متغیر می‌شود (شکل ۱) (Minasny and McBratney, 2006b).



شکل ۱-توزیع تجمعی هر متغیر به n فاصله با احتمال مساوی (x= متغیر)

مشکلی که در روش نمونه برداری مکعب لاتین وجود دارد این است که نمونه‌هایی که با این روش انتخاب می‌شوند ممکن است در دنیای واقعی وجود نداشته باشند برای مثال با در نظر گرفتن داده‌های ارتفاع و شیب یک منطقه و رسم نمودار

احتمال تجمعی آنها و سپس استفاده از طرح نمونه برداری مربع لاتین برای تعیین نقاط مطالعاتی نشان می¬دهد که تنها برخی از نقاط به درستی روی احتمال تجمعی متغیرها قرار دارد و برخی دیگر در حقیقت در دنیای واقعی وجود ندارد و این در حالی است که با اضافه شدن متغیرهای بیشتر این مشکل بیشتر به چشم می‌خورد (Minasny and McBratney, 2006a).

روش مکعب لاتین مشروط بر پایه روش نمونه برداری مکعب لاتین است؛ اما با این تفاوت که در روش مکعب لاتین مشروط تمامی نقاطی که برای نمونه برداری مشخص می‌شود در جهان واقعی وجود دارند. مشکل عدم وجود نقاط در جهان واقعی که در روش نمونه برداری مکعب لاتین وجود داشت با استفاده از ایجاد تکرار در محاسبات برطرف شد. از سال ۲۰۰۷ تا کنون محققان بسیاری جهت تعیین نقاط نمونه برداری جهت تهیه نقشه رقومی کلاس‌ها یا خصوصیات خاک با استفاده از متغیرهای کمکی از روش مکعب لاتین مشروط استفاده کرده‌اند (Brungard et al, 2015 and Minasny and McBratney, 2007). متغیرهای کمکی شامل نقشه‌ها و تصاویر ماهواره‌های مختلف از جمله تصاویر ماهواره لندست و مدل رقومی ارتفاع می باشند (Cambule et al, 2013 and Taghizadeh-Mehrjardi et al, 2015).

هدف از این تحقیق نیز معرفی یک روش نمونه برداری هدفمند جهت صرفه‌جویی در هزینه‌ها و زمان در مطالعات پدولوژی می‌باشد.

مواد و روش‌ها

منطقه مورد مطالعه با مساحت ۷۲/۶۱ کیلومتر مربع در شرق استان قزوین در محدوده شهرستان آبیک و بین ۳۶ درجه و ۲ دقیقه تا ۳۶ درجه و ۵ دقیقه عرض شمالی و ۵۰ درجه و ۲۴ دقیقه تا ۵۰ درجه و ۲۹ دقیقه طول شرقی واقع شده است.

با استفاده از تصاویر ماهواره لندست ۸ و مدل رقومی ارتفاع مستخرج از سنجنده استر، استخراج متغیرهای کمکی با استفاده از نرم‌افزارهایArc Gis، ERDAS IMAGING و SAGA انجام شد. این متغیرهای کمکی شامل باندهای طیفی تصویر ماهواره لندست، شاخص پوشش گیاهی، شاخص پوشش گیاهی اصلاح شده خاک، نقشه درجه شیب، نقشه جهت شیب، شاخص خیسی، فاکتور طول شیب و شاخص قدرت جریان شدند. پس از استخراج متغیرهای کمکی و وارد کردن آن‌ها به‌عنوان فایل رستری به محیط برنامه Arc Gis، 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. سپس از طریق افزونه Xtools متغیرهای کمکی به فایل اکسل انتقال داده شدند. برای به دست آوردن اهمیت هرکدام از متغیرها از طریق نرم‌افزار SAS دستور تحلیل مؤلفه‌های اصلی اجرا شد و ۱۱ متغیر به‌عنوان مهم‌ترین متغیرهای کمکی انتخاب شدند. در نهایت متغیرهای کمکی انتخاب شده وارد مدل مکعب لاتین مشروط شد و الگوریتم داده‌ها را به n مقدار با احتمال برابر تقسیم کرد، 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.

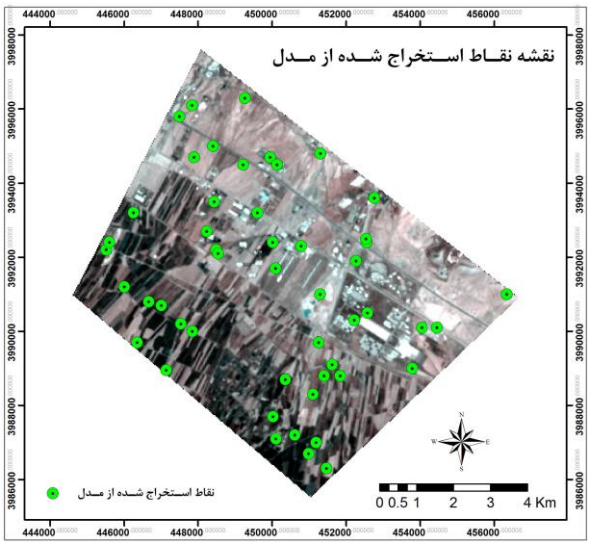
نتایج و بحث

ورودی‌های مدل، براساس استفاده از آنالیز مؤلفه‌های اصلی تعیین گردید و با توجه به جدول ۱ متغیرهای ارتفاع، شیب، جهت شیب، شاخص خیسی، فاکتور طول شیب، شاخص قدرت جریان، باند ۴، باند ۵، باند ۹، شاخص پوشش گیاهی و شاخص پوشش گیاهی اصلاح شده خاک وارد مدل شدند.

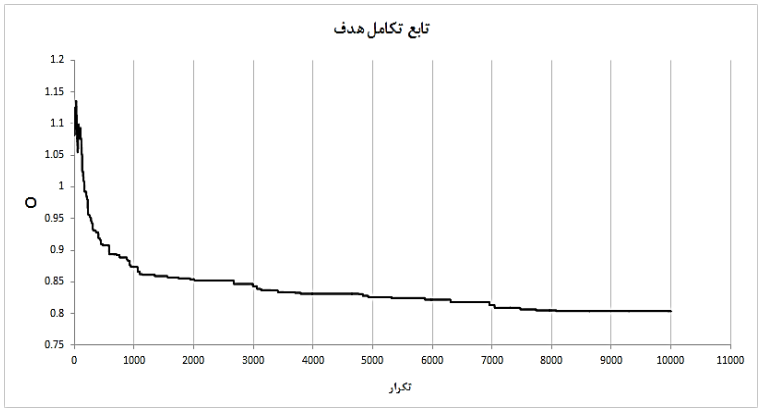
جدول ۱- نتایج آزمون تجزیه مؤلفه‌های اصلی

	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6
DEM	۰/۲۰۷۷۸۶	-۰/۲۲۸۸۱۲	۰/۲۰۸۳۳۷	۰/۱۷۲۶۷۲	۰/۰۱۲۳۲۸	۰/۰۱۱۴۳۶
Aspect	۰/۱۰۵۱۱۲	-۰/۲۲۲۸	۰/۴۶۰۶۰۶	۰/۳۶۵۸۴۴	-۰/۰۳۴۲۵۳	۰/۰۰۸۶۳۲
NDVI	-۰/۲۷۷۳۴۳	۰/۲۴۸۹۱۴	۰/۲۲۸۵۱۹	۰/۰۵۵۳۸۶	۰/۰۷۹۷۲۹	۰/۰۰۲۴۰۹
TWI	-۰/۰۰۷۹	۰/۱۵۵۳۸۵	-۰/۳۴۳۴۶۵	۰/۶۰۴۵۵۶	-۰/۰۰۳۵۵۷	-۰/۰۱۶۹۴۴
LS	۰/۰۰۳۶۴۳	-۰/۰۰۴۱۵۹	۰/۰۱۱۶۳۱	-۰/۰۴۳۷۰۲	۰/۰۰۷۴۵۲	۰/۹۹۶۵۰۷
Stream power	۰/۰۳۲۹۴۹	-۰/۱۸۶۳۴۳	۰/۳۸۳۱۴۸	-۰/۵۵۶۷۲۶	-۰/۰۰۱۳۴۶	-۰/۰۷۸۶۹۲
SAVI	-۰/۲۷۸۰۰۳	۰/۲۴۶۲۹	۰/۲۲۸۱۳۸	۰/۰۵۵۷۹۳	۰/۰۵۹۰۲۲	۰/۰۰۲۶۱۴
Slope	۰/۱۱۲۵۶۶	-۰/۱۸۳۰۶	۰/۴۸۵۸۶۶	۰/۳۶۷۴۸۸	-۰/۰۴۵۶۱۵	۰/۰۰۹۷۶۴
Band01	۰/۲۹۶۴۱۶	۰/۱۵۶۳۵۵	-۰/۰۳۶۷۱۹	-۰/۰۵۹۰۶۳	-۰/۰۲۶۶۹۷	-۰/۰۰۲۸۰۱
Band02	۰/۳۰۲۸۰۶	۰/۱۵۴۷۸۳	-۰/۰۳۶۱۵۳	-۰/۰۵۸۱۲۹	-۰/۰۵۱۹۰۱	-۰/۰۰۲۶۳۷
Band03	۰/۳۰۳۹۰۷	۰/۱۹۴۴۵۸	-۰/۰۱۲۵۳۵	-۰/۰۵۵۵۹۳	-۰/۰۷۹۰۷۲	-۰/۰۰۲۲۶۸
Band04	۰/۳۰۷۵۸۵	۰/۱۳۹۸۳۵	-۰/۰۲۶۰۰۶	-۰/۰۴۴۸۸	-۰/۰۸۳۶۴۷	-۰/۰۰۱۴۲۵
Band05	۰/۰۳۳۵۱۹	۰/۶۲۷۳۲۶	۰/۳۴۰۵۲۴	۰/۰۲۷۹۰۷	-۰/۰۱۰۲۶۱	-۰/۰۰۳۵۸۴
Band06	۰/۲۸۹۰۲۸	۰/۱۷۶۲۲	۰/۰۸۷۳۵۶	۰/۰۱۴۵۴۵	-۰/۰۱۶۰۷۲	-۰/۰۰۲۴۶
Band07	۰/۳۰۳۵۹۱	۰/۱۱۰۸۹۴	۰/۰۱۸۲۷۳	-۰/۰۰۸۰۵۲	-۰/۰۳۱۱۶۹	-۰/۰۰۴۴۲۸
Band08	۰/۲۹۵۷۴۶	۰/۱۶۶۴۸۷	-۰/۰۲۲۵۸۱	-۰/۰۵۱۱۳۹	-۰/۰۸۵۹۲۵	-۰/۰۰۱۸۱۴
Band09	۰/۰۴۶۱۴	۰/۱۲۵۳۹۵	۰/۰۳۳۰۵۲	-۰/۰۰۹۹۷۲	۰/۹۳۲۶۹۵	-۰/۰۱۰۰۳۷
Band10	۰/۲۶۸۷۱۱	-۰/۲۲۱۴۸۴	-۰/۰۷۳۰۵۹	۰/۰۳۱۴۲۷	۰/۲۰۲۹۳۸	۰/۰۰۸۱۵
Band11	۰/۲۶۹۴۲۲	-۰/۲۲۱۱۶۱	-۰/۰۶۹۶۳۳	۰/۰۳۲۳۰۳	۰/۲۲۴۰۶۹	۰/۰۰۲۸۷۸

نهایتاً تعداد ۵۰ نقطه جهت مطالعه توسط روش مکعب لاتین مشروط معرفی شد که در شکل ۲، نقاط نمونه برداری نشان داده شده است. همچنین تابع تکامل هدف نیز برای نتایج خروجی از روش رسم که مقدار این تابع با افزایش تکرار کاهش یافت به طوری که در شرایط تکرارهای کمتر از ۱۰۰۰ تابع از آشفتگی بیشتری برخوردار است اما با افزایش تعداد تکرارها پس از تکرارهای ۸۰۰۰ بار تابع تکامل به پایداری می‌رسد که نتایج به دست آمده مشابه نتایج برخی محققان بود (Minasny and McBratney, 2006a و تقی زاده مهرجردی و همکاران، ۱۳۹۳) شکل۳.



شکل۲- نقاط استخراج شده از روش مکعب لاتین مشروط



شکل ۳- تابع تکامل هدف

به طور کلی مزایای روش نمونه برداری مکعب لاتین مشروط عبارت‌اند از: ۱- نقاط خروجی از الگوریتم در دنیای واقعی وجود دارند.۲- داده‌های متغیرهای پیوسته را می‌تواند به‌خوبی با دادهای طبقه بندی شده یکپارچه شوند. ۳-نمونه برداری بر پایه توزیع تجربی داده‌های اصلی است بنابراین این یک روش ناپارامتری است. ۴- در صورت نیاز می‌تواند پراکنش مناسب نقاط را نشان دهد. ۵- محدودیت‌های اضافی توانایی اعمال در تابع نهایی هدف را دارند. مکعب لاتین مشروط، به عنوان فرآیندی برای تعیین نقاط نمونه برداری قادر است علاوه بر استفاده از داده‌های پیوسته از داده‌های کلاس بندی شده نیز در کنار داده‌های پیوسته برای رسیدن به یک نتیجه مطلوب استفاده کند و می‌تواند زمانی که در برابر طیف وسیعی از شرایط محیطی قرار دارد، به عنوان روش نمونه برداری مطلوبی عمل کند. می‌توان گفت این روش در حقیقت از خود محیط برای تعیین نقاط نمونه برداری را استفاده می‌شود، به عبارت دیگر اگر کارشناس در مورد محدوده مورد مطالعه اطلاعات کافی نداشته باشد باعث خطا در طرح نمونه برداری نمی‌شود.

منابع

تقی زاده مهرجردی، ر.، سرمدیان، ف.، تازه، م.، امید، م.، تومانیان، ن.، روستا، م ج. و رحیمیان، م ح. ۱۳۹۳. مقایسه روش‌های مختلف نمونه برداری جهت پهنه بندی رقومی خاک در منطقه اردکان، مجله مهندسی و مدیریت آبخیز، دوره ۶، شماره ۴، صفحه‌های ۳۵۳ تا ۳۶۳.

Brungard, C.W., Boettinger, J.L., Duniway, M.C., Wills, S.A. and Edwards, T.C., 2015. Machine learning for predicting soil classes in three semi-arid landscapes. Geoderma, 239: 68-83.

Brus, D.J. and De Gruijter, J.J., 1997. Random sampling or geostatistical modelling? Choosing between design-based and model-based sampling strategies for soil (with discussion). Geoderma, 80(1-2): 1-44.

Cambule, A.H., Rossiter, D.G. and Stoorvogel, J.J., 2013. A methodology for digital soil mapping in poorly-accessible areas. Geoderma, 192: 341-353.

Goovaerts, P., 1997. Geostatistics for natural resources evaluation. Oxford University Press on Demand.

Iman, R.L. and Conover, W.J., 1980. Small sample sensitivity analysis techniques for computer models. with an application to risk assessment. Communications in statistics-theory and methods, 9(17): 1749-1842.

McBratney, A.B., Santos, M.M. and Minasny, B., 2003. On digital soil mapping. Geoderma, 117(1): 3-52.

McKenzie, N.J., Grundy, M.J., Webster, R. and Ringrose-Voase, A.J. eds., 2008. Guidelines for surveying soil and land resources. CSIRO PUBLISHING.

Minasny, B. and McBratney, A.B., 2006a. A conditioned Latin hypercube method for sampling in the presence of ancillary information. Computers & Geosciences, 32(9): 1378-1388.

Minasny, B. and McBratney, A.B., 2006b. Latin hypercube sampling as a tool for digital soil mapping. Developments in soil science, 31: 153-606.

Minasny, B. and McBratney, A.B., 2007. Spatial prediction of soil properties using EBLUP with the Matérn covariance function. Geoderma, 140(4): 324-336.

Minasny, B., McBratney, A.B. and Hartemink, A.E., 2010. Global pedodiversity, taxonomic distance, and the World Reference Base. Geoderma, 155(3): 132-139.

Taghizadeh-Mehrjardi, R., Nabiollahi, K., Minasny, B. and Triantafilis, J., 2015. Comparing data mining classifiers to predict spatial distribution of USDA-family soil groups in Baneh region, Iran. Geoderma, 253: 67-77.

The use of conditioned Latin Hypercube Sampling (cLHS), to the design of study parts in Pedology.

S. E. Khamoshi, F. Sarmadian* and A. Keshavarzi

Former graduate student, Professor and Assistant Professor respectively, Department of Soil Science and Engineering, Tehran University.
Email*: fsarmad@ut.ac.ir

Abstract

To maintain soil fertility and increase productivity in sustainable agriculture, it is necessary to be aware of the different features. In this study, conditioned Latin Hypercube Sampling model to determine if soil sampling points in the study were examined. The study area in Qazvin province, the city is Abyek. Moisture and Temperature regimes are Xeric and Thermic area is. Using Landsat 8 satellite images and images of ASTER total of 19 auxiliary variables were extracted principal component analysis was performed in order to select variables among the most effective. The selected variables into the model, and finally 50 points as the right to take samples in the study area were selected. The accuracy of the method was tested by plotting the Evolution of objective function that reflects its low error performance is a model for the design of sampling points in Pedology.

Keywords: auxiliary variables, Evolution of objective function**,** principal component analysis